DYNAMIC WIFI FINGERPRINTING INDOOR POSITIONING SYSTEM

Omar Costilla Reyes

Thesis Prepared for the Degree of

MASTER OF SCIENCE

UNIVERSITY OF NORTH TEXAS

August 2014

APPROVED:

Kamesh Namuduri, Major Professor
Krishna Kavi, Committee Member
Robert Akl, Committee Member
Shengli Fu, Chair of the Department of
Electrical Engineering
Costas Tsatsoulis, Dean of the College
of Engineering
Mark Wardell, Dean of the Toulouse
Graduate School

Costilla Reyes, Omar. Dynamic WIFI Fingerprinting Indoor Positioning System. Master of Science (Electrical Engineering), August 2014, 92 pp., 10 tables, 90 figures, bibliography, 31 titles.

A technique is proposed to improve the accuracy of indoor positioning systems based on WIFI radio-frequency signals by using dynamic access points and fingerprints (DAFs). Moreover, an indoor position system that relies solely in DAFs is proposed. The walking pattern of indoor users is classified as dynamic or static for indoor positioning purposes. I demonstrate that the performance of a conventional indoor positioning system that uses static fingerprints can be enhanced by considering dynamic fingerprints and access points. The accuracy of the system is evaluated using four positioning algorithms and two random access point selection strategies. The system facilitates the location of people where there is no wireless local area network (WLAN) infrastructure deployed or where the WLAN infrastructure has been drastically affected, for example by natural disasters. The system can be used for search and rescue operations and for expanding the coverage of an indoor positioning system.

ACKNOWLEDGEMENTS

I would like to express my gratitude to my major professor Dr. Kamesh Namuduri for his guidance and support for the completion of this thesis.

I would like to thank my committee members Dr. Krishna Kavi and Dr. Robert Akl for their invaluable contributions in the development of this work, and CONACYT for providing financial support during my graduate studies.

I have no words to express my gratitude to my family and friends for their unconditional support in this journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

Global positioning systems (GPS) use satellites that orbit the earth to calculate outdoor location. Since its inception in 1973, GPS has been widely used, commercially-speaking, for many applications ranging from military to commercial. All industrialized and technologically developed societies depend on GPS.

Since GPS requires a direct line of sight between the satellites and a mobile device to correctly receive the signal, there has to be no obstruction between such parties for this technology to function correctly. As a consequence of this limitation, it is not plausible to use GPS signals to localize someone inside a building. For this reason alternatives have to be developed to effectively locate someone indoors.

From finding a shop in a large mall to finding a gate at an airport, indoor positioning systems (IPS) have proven their importance. Several indoor localization technologies and techniques have been proposed in recent years to solve this problem. None of those technologies have become ubiquitous. The Institute of Electrical and Electronics Engineers (IEEE) have not yet released specifications regarding indoor technologies standards since, as of today, a solution does not exist that can solve this problem perfectly.

The technologies that can be used to solve the indoor localization problem, which ranges from using the wireless local area network (WLAN) infrastructure to using ultra wideband (UWB) technology, have their own advantages and disadvantages, which explains the lack of a clear winner at this point.

Localization and navigation of robots while indoors has also been a problem extensively studied. A technique called simultaneous localization and mapping (SLAM) calculates the localization of a robot and at the same time creates a map where the robot can navigate, it uses landmarks installed in the environment to solve the location and navigation task.

1.1. Motivation

Even though a standard for indoor positioning systems, is not available as of today the most widely used indoor positioning system technology is based on using the wireless local area network (WLAN) infrastructure deployed in buildings. The main reason behind this trend is because of its low cost and extensively use. Its original intention is to provide internet access to users within the building.

The work presented in this thesis is based on WLAN positioning systems and the use of smartphones to improve those systems. Smartphones have become increasingly popular in the last few years surpassing desktop computers in number of sales [2].

The main contribution of this thesis is based on the idea of taking advantage of the WIFI hotspot feature embedded in most smartphones. The hotspot feature is being used as a WIFI repeater that create temporary access points from personnel who are inside buildings. Indoor positioning systems that have been developed so far shortfall on taking advantage of this WIFI feature. The approach of this thesis contributes by yielding accuracy superior to that of current WLAN indoor systems.

This thesis also implements a SLAM technique called RGBD-SLAM in a mobile robot. The technique was developed by Felix et al.[9] which uses a kinect camera for obtaining video in real time for further processing to solve the SLAM problem.

1.2. Motivation for the Research

Improvement is needed in terms of accuracy for indoor positioning systems currently based on WLAN infrastructure. As is presented in this thesis, there are many indoor positioning systems that exist using a wide variety of technologies and there are advantages and disadvantages for each of those systems. The aim of this research is to show that the novel WLAN-based indoor positioning system presented in this thesis provides better accuracy compared to existing systems.

1.3. Relevance to the Field

Improving accuracy indoors using the WLAN infrastructure has been one of the main challenges needed to be addressed for indoor positioning systems; the amount of accuracy obtained varies depending on the approach and technology considered to solve this problem.

1.4. Relevance to Society

The relevance to the society in general lies in that there are several market segments and emerging technologies that can take full advantage of an indoor positioning system. Indoor positioning systems can be used for applications such as location based shopping and advertising, and even for emergency response applications.

Several smartphone applications and websites are using location based services based on GPS technology to obtain more information about the user and consequently to show user-tailored information when located at a certain location. For example an application designed to show information regarding restaurants can take advantage of the current user location to infer the closest restaurant and show directions to the user automatically.

1.5. Overview

Chapter 2 presents the background and literature review of indoor positioning technologies, techniques and algorithms with great emphasis in the WLAN positioning systems, since the WLAN technology approach was selected for the work presented in this thesis. Chapter 3 presents the implementation of a room-level accuracy Indoor positioning system using MATLAB. Chapter 4 presents the main contribution of this thesis, which is the dynamic WIFI fingerprinting indoor positioning system, The chapter explains the proposed technique, implementation of the server and android applications, and results. Finally, chapter 5 presents the SLAM technique implemented in a robot. All the testings were performed at the University of North Texas (UNT) Discovery Park building.

CHAPTER 2

BACKGROUND AND LITERATURE REVIEW OF POSITIONING TECHNOLOGIES, TECHNIQUES AND ALGORITHMS

In this chapter existing indoor positioning techniques, topologies and technologies are presented with their corresponding performance metrics, advantages and disadvantages; also definitions necessary to understand indoor positioning systems are presented.

2.1. Definitions

- Transmitter

  A transmitter is a device that produces radio waves to be propagated in a medium.
- Receiver

  A receiver is designed to receive and process the radio waves emitted by the transmitter.
- Positioning

  Positioning refers to the process of inferring the current location of a user in an environment. The environment can be indoors or outdoors.
- Radiolocation

  The principal method used to infer location, for both indoors and outdoors, is based on using radio frequency waves. The location is inferred by measuring parameters of the radio waves between a receiver and a transmitter [7]. In most systems the user acts as the transceiver and there exist several receivers that process the radio waves coming from the transceiver to infer the localization. In the case of outdoor positioning, when a mobile device needs to be localized, the process of radiolocation is performed by measuring signals coming from nearby cellular base stations or satellites for GPS.
- Wireless local area network (WLAN)

  The purpose of this type of network is to allow connectivity between 2 or more wireless devices using a wireless medium. WLAN is also used to provide internet connectivity to the devices.

2.2. Metrics

The parameters presented in the following section allow for the evaluation of the performance of an indoor positioning system.

- Accuracy

  Accuracy is a performance metric defined as the error that exists between the location estimated by the positioning system and the true location of the user. Accuracy is considered the most important metric of an indoor positioning system.

- Precision

  Precision refers to the degree of closeness between several location calculations with the same signal.

- Complexity

  The more complex the indoor positioning system is, the more time it takes to obtain a position and the more energy is needed to successfully localize the user, it is important to keep the complexity as low as possible to decrease the power consumption of the system.

- Responsiveness

  This metric calculates how fast the system performs calculations to infer the location and to display the result to the user needed to be localized; this performance depends on the complexity of the system.

- Robustness

  High robustness is achieved when an indoor positioning system can work even when some of the signals used for localization purposes are not available or are disturbed by noise in the environment. A robust indoor positioning system has to use the incomplete information efficiently to obtain correct location.

- Scalability

  This metric has to do with efficiency of the indoor positioning system when dealing with a large number of users. A high scalable system can efficiently process many requests coming from mobile devices that need to be localized, with all of them being requested at the same time.

- Coverage

  The coverage of an indoor positioning system is defined as the total area within the indoor floor plan of a building where the system is able to operate.

- Costs

  The cost of an indoor positioning system can be expressed in terms of infrastructure needed, desired lifetime, responsiveness of the system, whether beacons are needed or not, among other factors. Some indoor positioning systems are monetarily expensive to implement because as the area increases the more sensors are needed for positioning.

- Resource efficiency

  This parameter is related to efficiently using the knowledge of where the users to be localized are situated to optimize the use of resources related for indoor positioning; if a high density of users occupy a specific indoor area, it could be possible to reduce the resources at the places where the system is not being used.

2.3. System Topologies for Localization Systems

There are 4 system topologies [18] that can be used for indoor positioning. A topology is a geometric configuration for indoor localization.

- Remote positioning system

  Using this topology, the mobile device needed to be localized emits a signal that is received by remote sensors which further process the signal and the result of the calculated position is returned to the device.

- Self-positioning

  Using this topology, the position is determined on the mobile device itself. This unit receives the signal from several transmitters located in known locations and then the location is computed in the positioning device based on values obtained from the measured signals.

- Indirect remote positioning

  Using this topology, a device sends the measurement results from a self-positioning measuring unit to a remote server or receiver for future processing and then the information is further used to locate new users.

- Indirect self positioning

  Using this topology, the measurement result is stored in a database and then the information is sent from a remote positioning unit to a mobile device.

2.4. Analysis of WLAN Received Signal Strength Indicator (RSSI) at UNT's Discovery Park

In the following section, an analysis of the WLAN signals at 2 different locations are presented; The reason behind studying the signals behavior lies in the importance of the variation of the RSSI over time, this effect is of great importance for an indoor positioning system.

2.4.1. Multipath Fading

Multipath fading is an effect that occurs between a transmitter and a receiver in a noise environment. The effect causes a constructive and destructive interference in the signals that affect the signal strength. The strength increases or decreases the signals amplitude according to external factors such as noise [26].

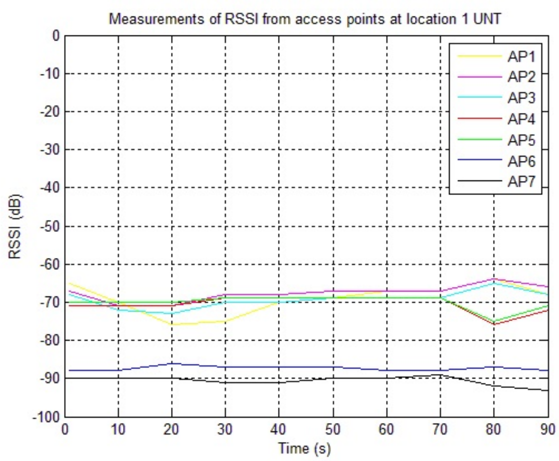
2.4.2. Shadowing

Shadowing is additional attenuation of signal power. Shadowing occurs when an object affects the propagation of a signal between a transmitter and a receiver in a noisy environment. This random effect depends also on human body presence that blocks the signal [26].

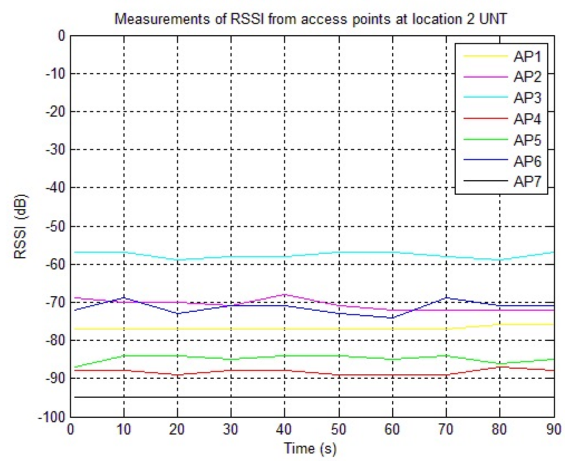
2.4.3. RSSI Measurements without Rotation

Signal measurements at UNT's Discovery Park were obtained from nearby access points (APs) installed in the infrastructure to observe if the fluctuation of the measurements was significant over time, since a notable change could drastically affect the functionality of the IPS. Ideally, a constant RSSI measurement is obtained from access points, but in real case the RSSI measurements are being affected by noise.

Figure 2.1a and Figure 2.1b illustrate the results obtained when considering 10 access points at 2 different locations, the locations are the second floor main hallway and the first floor main hallway of the Discovery Park building, respectively.



(A) Location 1: DP Second floor main hallway



(B) Location 2: DP First floor main hallway

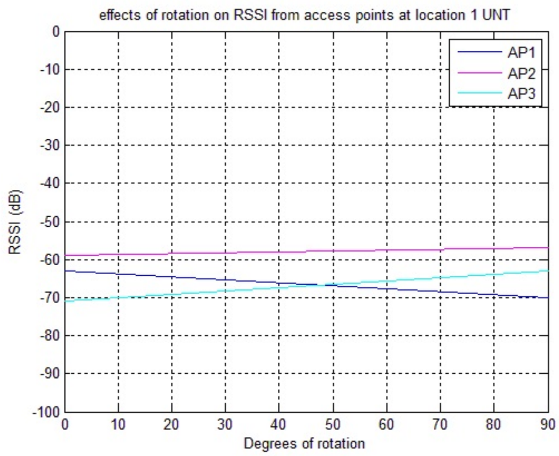
FIGURE 2.1. Offline and testing fingerprints

The RSSI results over a time period of 90 seconds did not have any drastic changes, which is desired for an IPS.

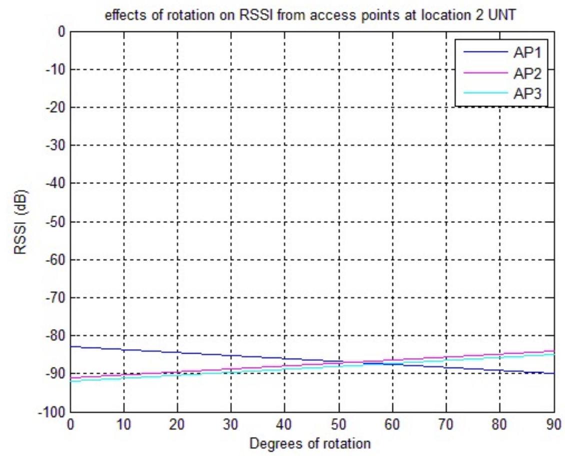
WLAN signals that stay at the same level are desired for any WIFI fingerprinting system since those signatures yield better overall performance.

2.4.4. RSSI Measurements with Rotation

Figure 2.2a and Figure 2.2b illustrate how rotation of the phone affected the RSSI readings from 0 to 90 degrees.



(A) Location 1: Second floor main hallway with rotation



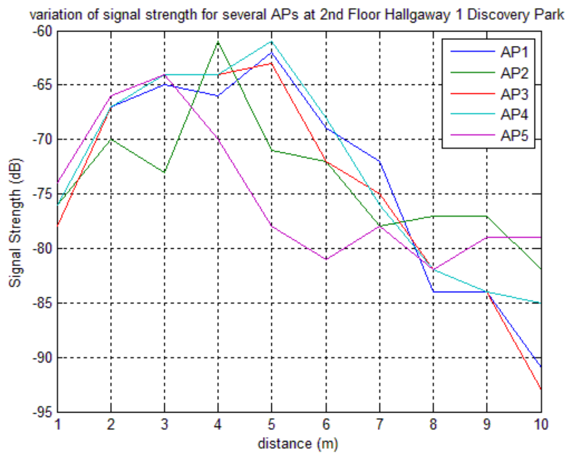
(B) Location 2: First floor main hallway with rotation

FIGURE 2.2. RSSI measurements with rotation

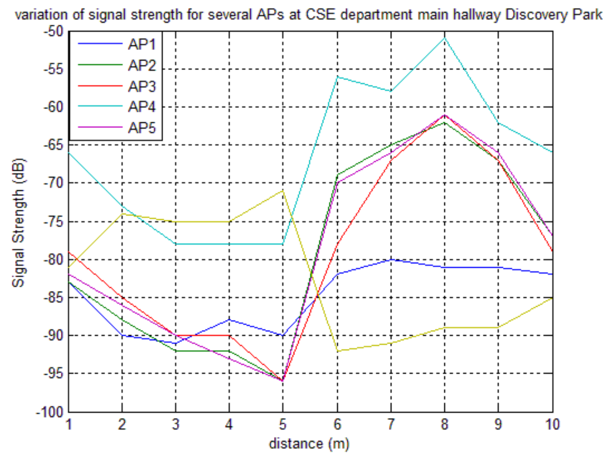
At location 1, the access point number 3 changes from a RSSI of -70 at 0 degrees to -60 at 90 degrees, this result is undesirable for an IPS since different RSSI readings at a same location decrease the accuracy of an IPS, this is the main reason why several readings are needed considering rotation. The multipath and shadowing effects are the main factors that degrade the strength.

2.4.5. Variation of the RSSI with Change of Distance

In the following subsection, the variation of the signal strength was recorded when change in distance is presented between the access point and the mobile device; the experiment was performed at the Discovery Park building. The experiment is recorded when changing from 1 to 10 meters in 2 different locations (Figure 2.3a). The locations selected were the Computer Science and Engineering (CSE) hallway and the second floor main hallway at the Discovery Park building. The peaks of the signal strength represents the best strength values that can be obtained from the WIFI signals coming from access points. In Figure 2.3a the peak of most of the signal is obtained at 5 meters, in Figure 2.3b most of the peak values of the signals are obtained at a distance of 8 meters. Those specific locations represent the closest distance to access points installed in the infrastructure.



(A) Location 1: Second floor main hallway



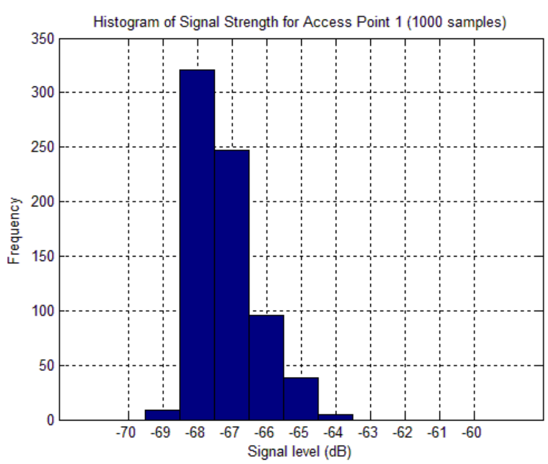
(B) Location 2: CSE main hallway

FIGURE 2.3. RSSI measurements with rotation at location 1 and 2

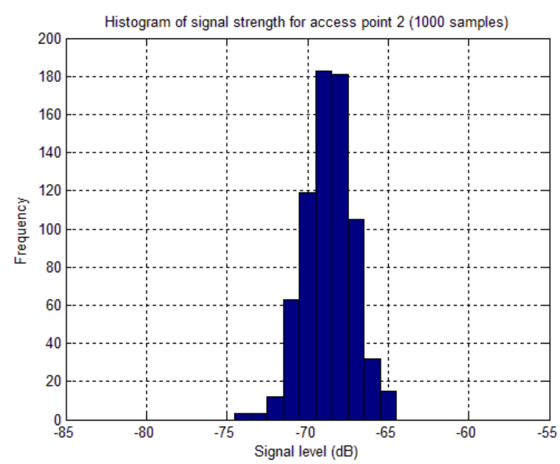
2.4.6. Histograms of RSSI Frequency

This subsection presents the analysis of the change of the signal strength when several readings are being made at the same location. For this experiment, 1000 samples of the received signal strength with sample period of 5 second were captured.

In Figure 2.4 and Figure 2.5 it can be observed that the signal changes in a 10 dB interval.

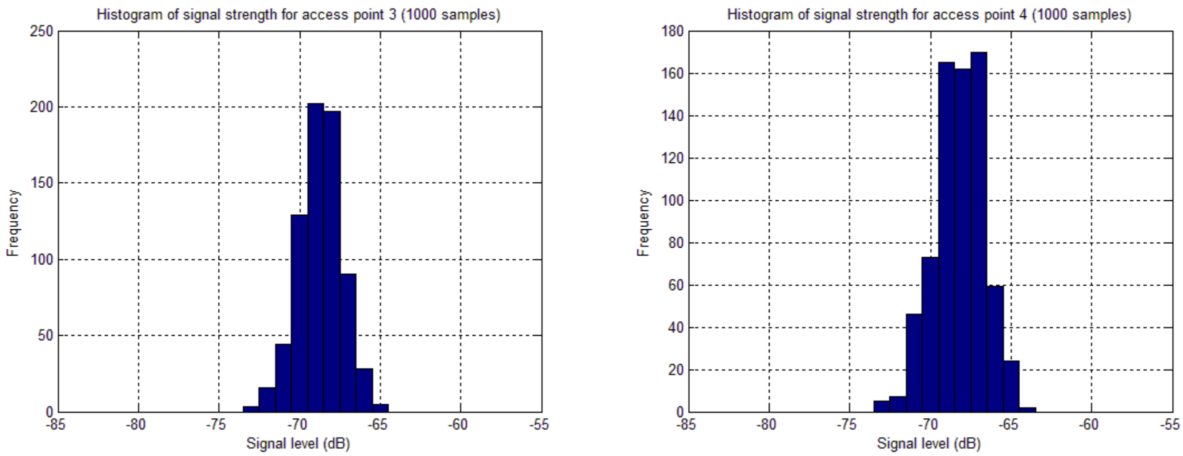


(A) Location 1: First floor Hallway



(B) Location 2: Second floor Hallway

FIGURE 2.4. RSSI for 1000 samples



(A) Location 1: First floor Hallway

(B) Location 2: Second floor Hallway

FIGURE 2.5. RSSI for 1000 samples obtained at Discovery Park

2.5. Positioning Techniques

The indoor positioning techniques can be categorized into triangulation and trilateration, scene analysis and proximity as it can be observed in Figure 2.6.

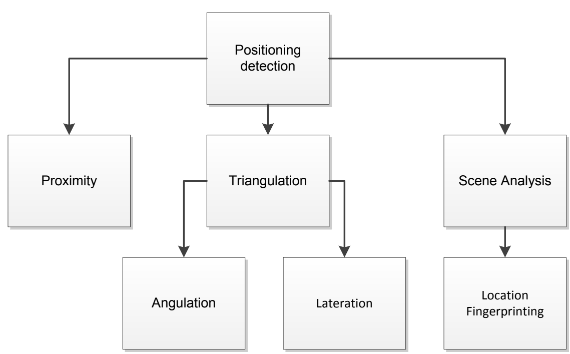


FIGURE 2.6. Classification of indoor positioning technologies and subcategories

2.5.1. Triangulation and Trilateration

Triangulation indoor methods use the angles of at least 3 known positions in respect to the user to be localized to infer location. GPS is an example of a system that uses such a method. In trilateration, 3 positions at least are needed to infer the location.

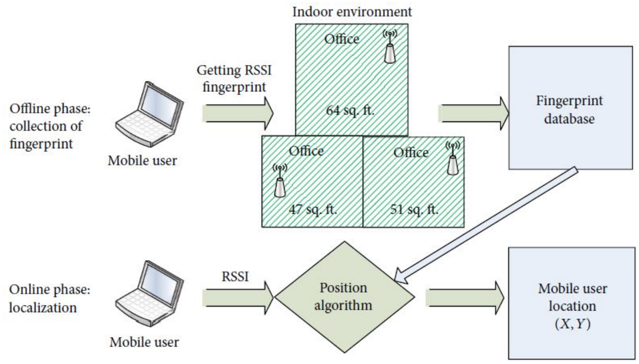


FIGURE 2.7. Positioning using a fingerprinting system

2.5.2. Scene Analysis

This positioning technique analyzes the features from an observed scene. The scene can be any type of signal, from a sensor or from the environment itself that can be measured and used to differentiate between locations. Using this approach, the location of a device can be calculated based on the similarity between scenes. A WLAN fingerprinting positioning system is considered to be part of the scene analysis positioning technique and it is the foundation of the IPS presented in this thesis.

2.5.3. Proximity

Using this technique, the device to be localized acts as a transmitter and there exist several receptors within the localization space, the localization of the device can be inferred according to the proximity of the device to the sensors.

2.6. WIFI Fingerprinting Indoor Positioning System

This section emphasizes the concepts behind the WIFI fingerprinting technique for indoor positioning-based systems as it is the approach being used for the system developed in this thesis. The approach is presented in Figure 2.7 [12]. WIFI fingerprinting is a scene analysis technique and is the most extensively used radio frequency based technology for indoor localization. WIFI fingerprinting has been shown to be a reliable way to localize people indoors since it uses infrastructure already deployed indoors.

2.6.1. Obtaining Indoor Position from Fingerprints

WIFI fingerprinting associates a unique location inside a building to a fingerprint that gives that location a specific identifier. The fingerprint is usually a feature of a signal in the indoor environment.

The received signals at the mobile device emitted by one or several transmitters can be used to infer the location of the user. The location can be computed locally or remotely. To obtain the position of a mobile device, a match needs to be performed between the signal being read at the mobile device in real time and those signals previously saved in a database.

2.6.2. RSSI for Fingerprinting

Any type of signal that can help differentiate a location inside a building can be used as a fingerprint. For this thesis, the received signal strength obtained from nearby WIFI access points is used to characterize the fingerprint. The RSSI in noise free environments can be modeled with the help of the following equation:

$$RSSI = P - R - 10\alpha log_{10} d \tag{1}$$

P is the transmitted power, α is the path loss exponent which falls linearly and R is a constant that depends on the conditions of the environment [26]. Due to noise in the environment, this equation cannot be used for trilateration localization purposes. The RSSI from multiple access points can be employed to infer the localization of the mobile device, which is the core idea of the fingerprinting method.

Single samples taken from the RSSI received from nearby access points are not sufficient to characterize a fingerprint. It is necessary to obtain an average of the readings to successfully identify a fingerprint. The collection of access point average readings at one location is what characterizes one fingerprint location. Table 2.1 presents an example of RSSI average readings.

BSSID	Mac Address	Signal Level (dB)
eduroam	00:1a:1e:1a:6a:d2	-63
eduroam	00:1a:1e:1a:58:d2	-73
eduroam	00:1a:1e:1a:78:02	-70
UNT	00:1a:1e:1a:6b:71	-89
UNT	00:1a:1e:1a:02:f1	-88
eduroam	00:1a:1e:1a:6b:72	-91
UNT	00:1a:1e:1a:70:b1	-95

TABLE 2.1. Example of RSSI at Discovery Park building

2.6.3. Derivation of Position from RSSI Fingerprints

As explained in the previous section, a matching between the fingerprints from a training set and the fingerprints being read in real time on the mobile device needs to be performed. This process is called offline and online phases, respectively.

2.6.4. Offline Phase

During this phase, a survey of the indoor area where the indoor positioning system is going to be deployed is obtained to create a training set of offline fingerprints. Each fingerprint contains a set of averages values from the nearby access points that characterize that location, as presented in Table 2.2.

2.6.5. Online Phase

During this phase, the mobile device is within the indoor positioning system coverage. At the beginning the position of the mobile device is currently unknown. To calculate the position, the device reads the RSSI measurements from the near access points and creates a vector with the average of these readings; then it compares the values obtained with the ones saved on the offline survey using a positioning algorithm, the algorithm returns the

Location	Average AP 1 (dB)	Average AP 2 (dB)	Average AP 3 (dB)	Fingerprints
(0,0)	-78	-75	-95	Fingerprint 1
(1,1)	-86	-89	-95	Fingerprint 2
(2,2)	-80	-86	-83	Fingerprint 3
(3,3)	-82	-89	-91	Fingerprint 4

Table 2.2. Collection of fingerprint vectors

approximated location. The process of obtaining a location via a positioning algorithm are explained in the following section.

2.7. Adjacent Channel Interference in WLAN Networks

Tan et al.[25] investigate the effect of adjacent channel interference in neighboring nodes of wireless WLAN mesh networks operating in the same or adjacent frequency channels. They reach the conclusion that 37 dB to 45 dB less attenuation is needed to eliminate the adjacent channel interference between 2 nodes. These results show that when new WLAN infrastructure is installed in a building, adjacent channel interference is created between new nodes and neighboring WIFI nodes already installed in the WLAN infrastructure, this effect will modify the RSSI pattern at the fingerprints of a WIFI fingerprinting indoor positioning system. A new survey is required if new access points are installed in the infrastructure.

2.8. Indoor Positioning Algorithms

Once data is captured with one of the techniques presented in the previous section (offline phase), an algorithm capable of processing the data to approximate the true location of the user is needed; this is a very important aspect of an indoor positioning system, since according to the algorithm chosen, the performance of the positioning system will be affected substantially.

2.8.1. WLAN Fingerprint Positioning Algorithms

For WLAN-based indoor positioning systems there are 2 types of algorithms to infer the location of the user given the data obtained by the positioning technology, the deterministic and probabilistic approaches.

2.8.1.1. Deterministic Algorithms

A fundamental property of a deterministic algorithm is that by giving the same set of input signals the output of the algorithm will always be the same.

In order to obtain the location of a user using a deterministic approach, the Euclidean distances between the offline fingerprints and the online fingerprint needs to be obtained.

Assuming M offline fingerprints, the Euclidean distance (D) between the ith measured online fingerprint f_i and the ith offline fingerprint can be calculated as:

$$D = \sqrt{\sum_{i=1}^{N} |r_i - f_i|^2} \tag{2}$$

This distance must be calculated between the online fingerprint and all the existing offline fingerprints, the smallest distance is used to infer which offline fingerprint is selected to infer the location of the user, as a consequence, the coordinates of the selected offline fingerprint determine the location of the user.

Bahl et al.[4] proposed the first indoor positioning system based on WIFI fingerprinting. They used a deterministic approach for their proposed system.

- K-Nearest Neighboor Algorithm

The K nearest neighbor algorithm is similar to the Euclidean distance approach, with the difference that K nearest offline fingerprints is used to infer the location of the online fingerprint. The euclidean distance is obtained for the case of $K = 1$.

Considering N access points deployed in the environment, the online fingerprint vector as $\mathbf{r}$ and the offline fingerprint vector as $\mathbf{f}$ the K nearest neighbors can be selected using the following distance metric:

$$D_k = (\sum_{i=1}^{N} |r_i - f_i|^q)^{\frac{1}{q}} \tag{3}$$

where D_k is called the Manhattan distance.

From the selected set of smallest K distances and the same set of K locations with pairs (x_k, y_k) The approximated location of the user (x, y) from K nearest neighbors can be obtained by using the following equation:

$$(x, y) = \frac{1}{k}\sum_{i=1}^{N}(x_i, y_i) \tag{4}$$

(x_i, y_i) are the coordinates of the ith fingerprint.

2.8.1.2. Probabilistic Algorithms

Roos et al.[22] were the first to purpose a probabilistic-based approach algorithm. They estimate the likelihood of a fingerprint distribution to obtain the approximated location of the user.

Given a vector of locations $\mathbf{v}$ of fingerprints and a signal vector $\mathbf{s}$, the element from $\mathbf{v}$ selected is the one obtained from:

$$\text{if } P(v_i|s) > P(v_j|s) \text{ for } i, j = 1, 2, 3, ..., n, j \neq i \tag{5}$$

$P(v_i|s)$ denotes the probability that the user is located at position v_i given the online fingerprint s.

$p(v_i)$ denotes the probability that the user is in location v_i. The selection of the fingerprint is based on posteriori probability.

If Bayes rule is used to solve this problem, and assuming an equal probability between locations expressed as: $P(v_i|s) = P(v_j|s)\ for\ i, j = 1, 2, 3, ..., n$ The following decision rule based on the likelihood that $P(s|v_i)$ is the probability that the signal $\mathbf{s}$ is received. $P(v_i)$ given that the user is located at location v_i.

The estimation of the location v_i can be obtained with:

$$P(v_i|s) = \frac{P(S|v_i)P(v_i)}{P(S)} \tag{6}$$

If considering that $P(S)$ is constant for all v the previous equation can be rewritten as:

$$P(V_i|S) = P(S|v_i)P(V_i) \tag{7}$$

The estimated location v is the one that attains the maximum probability when

$$v = \arg\max_{v_i}[P(v_i|s)] = \arg\max_{v_i}[P(S|v_i)P(v_i)] \tag{8}$$

2.8.2. Support Vector Machine (SVM)

SVM is a machine learning method used mainly for classification purposes, an SVM separates points in space by a hyperplane that separates positive samples from negative samples, using this algorithm it is necessary to maximize the distance between the samples.

For our purposes the training set containing all the fingerprints needs to be separated by location.

2.8.3. Propagation-based Algorithms

Using propagation-based algorithms a direct line of sight between transmitter and receiver is required for positioning.

2.8.3.1. Time of Arrival (ToA)

This algorithm measures the travel time of a radio signal from a transmitter to several receivers or vice versa. At least 3 reference points are required for a system implementing a TOA algorithm to function properly. The Figure 2.8 [12] shows a TOA system.

As in [18] a system considering this algorithm considers a device located at points (x_0, y_0) at time t_0 it transmits a signal to the N base stations located at $(x_1, y_1), (x_2, y_2)$, $(x_3, y_3)..., (x_N, y_N)$ and consequently the base stations receives a signal at times $t_1, t_2, ..., t_N$

The cost function of the system can be expressed as:

$$\mathbf{F(x)} = \sum_{i=1}^{n} \alpha_i^2 f_i^2(x)$$

α_i is the reliability of the signal received at the measuring unit i and $f_i(x)$ is defined as $f_i(x) = c(t_i - t) - \sqrt{((x_i - x)^2 + (y_i - y)^2}$,where c is the speed of light.

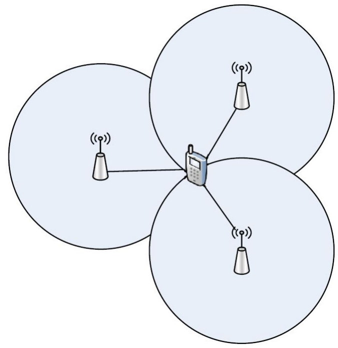


Figure 2.8. Calculation of position using TOA

2.8.3.2. Time Difference of Arrival (TDOA)

TDOA compares the relative difference in time at which the signal is received from several transmitter and receivers.

As opposed to TOA, in TDOA the receivers need to be synchronized.

Figure 2.9 [12] shows a TDOA system.

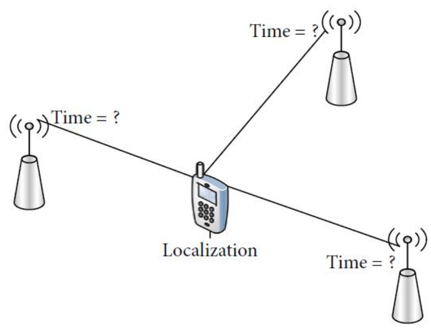


Figure 2.9. Calculation of position using TDOA

2.8.3.3. Angle of Arrival (AoA)

This technique is based on finding the angle of arrival between a transmitter and several receivers.

As an example, consider the following AoA system presented in Figure 2.10 [12], the angles at the receiver from 2 of the transmitters are considered as θ_1 and θ_2 respectively, and their locations are (x_1, y_1) and (x_2, y_2)

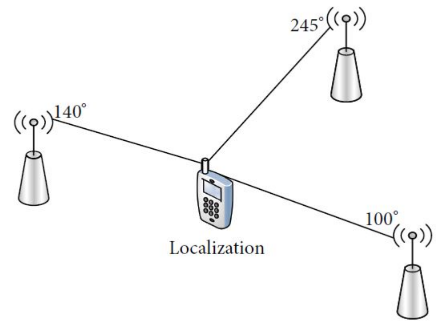


FIGURE 2.10. Calculation of position using AOA

2.8.3.4. Propagation-based Algorithms for WLAN positioning

Propagation-based algorithms calculate the position by obtaining measurements of the Received Signal Strength with path loss.

In [17] the Received Signal Strength can be measured using the following equation:

$$R = r - 10\alpha \log_{10}(d) - L \tag{9}$$

where r is the initial value of the Received Signal Strength, d is the distance from the desired location to a selected access point, α is the path loss exponent, L is the sum of the losses contributed by each wall in the building that affects the propagation of the signal strength.

2.9. WLAN Access Point Selection Strategies

Each AP available in the environment has its own contribution to the positioning system, there are APs that help in the performance of the system and there are APs that decrease the performance of the system. Discarding the APs that decrease the performance is needed, it is only necessary to include the APs that improves the performance.

A variety of access point selection strategies have been extensively studied in the existing WLAN fingerprinting location literature [30] [6] [10].

Youseff et al.[30] propose a joint probabilistic technique for indoor positioning; and presents an AP selection strategy called $MaxMean$, were a few access points from all that are available in the environment with the strongest RSSI are selected for positioning. Chen et al.[6] presents an access point selection strategy called $InfoGain$ based on selecting the APs with the highest discriminating power. The discriminative power of the ith AP is obtained as the reduction of entropy as described in the following equation:

$$InfoGain(AP_i) = H(G) - H(G|AP_i) \tag{10}$$

An AP with high discriminative power helps in efficiently differentiate fingerprints from one another. Chen et al. also introduces a random access point selection strategy that is independent from the signal strength called $RandMean$.

The $MaxMean$ and the $RandMean$ approaches are used in this thesis as access point selection strategies with the aim of increasing the accuracy of the system.

2.10. Indoor Navigation Technologies

In Figure 2.11, the technologies that can be used for indoor navigation are presented.

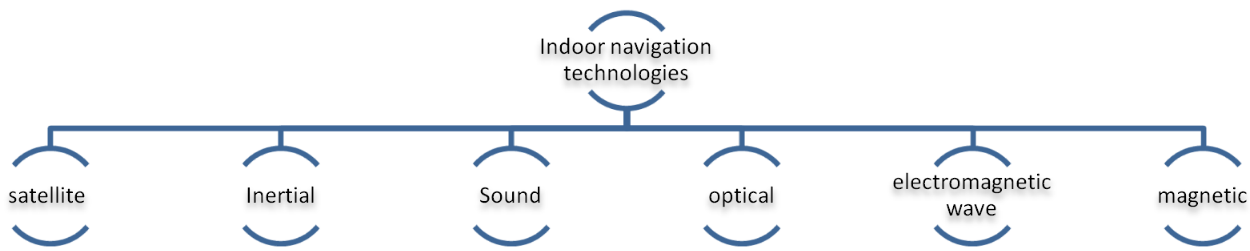


FIGURE 2.11. Indoor navigation technologies

2.11. Positioning Technologies

2.11.1. Dead Reckoning

This technology estimates the user location based on a previously known position, The techonology uses the aggregation of odometry readings to estimate the user location, as presented in [20]. The readings can be obtained with the built-in sensors included in smartphones as accelerometers, gyroscopes, magnetometers or compasses.

The nature of the aggregation of error over time makes this system inaccurate by itself, it needs to be combined with other technologies. However, this technology is infrastructureless, so there is no cost associated with deployment.

2.11.2. Infrared (IR)

This technology uses the infrared light for localization purposes, it requires a line-of-sight between transmitter and receiver to infer the location of the user. The usual configuration for this system is to install IR receivers on the walls of a building and an IR transmitter is carried out by the user to be localized, the transmitter is called a tag that emits a beacon of information that helps to localize the user. [28].

In Table 2.3 are presented the performance metrics for infrared signals.

Accuracy Range	Cost	Power consumption
1m to 2m	Medium	Low

TABLE 2.3. Performance metrics for IR

2.11.3. RFID

Radio frequency identification is a technology that deploys high-response RFID sensors in the indoor environment. The sensors are then able to localize users that carry either a passive or active RFID tag, the localization is being inferred by proximity which makes it hard to integrate with other non-proximity technologies.

The RFID tags carried by the users calculate the distance between each of the RFID tags from the users to obtain position.

In Table 2.4 are presented the performance metrics for RFID signals.

Accuracy Range	Cost	Power consumption
1m to 2m	Low	Low

TABLE 2.4. Performance metrics for RFID

2.11.4. Wireless Sensors Networks

A wireless sensor network (WSN) is installed inside buildings with the main purpose of capturing and processing environmental conditions as temperature sound or pressure.

A common example of a WSN is the Zigbee technology. The technology has low power and low data transmission rate. It can be used as an ad-hoc network for positioning as proposed in [13].

Zhang et al.[31] propose a WSN in which the localization of only a few nodes is known from a large set of nodes. The localization of the unknown nodes is inferred sequentially from the position of the known nodes. The authors conclude that the performance of this technique is largely dependent on the localization algorithm used.

In Table 2.5 are presented the performance metrics for WSNs.

Accuracy Range	Cost	Power consumption
3m to 5m	Low	Low

Table 2.5. Performance metrics for WSN

2.11.5. FM Radio Based Systems

The frequency modulated (FM) radio waves signals emitted by local FM radio stations commonly found all over the world can also be used for positioning. Chen et al.[5] combine RSSI values coming from WIFI and FM systems to create a unique fingerprint. They report that localization is improved by 83% when compared to WIFI Fingerprinting alone.

The main disadvantage of this approach is that smartphones nowadays do not include FM antennas which makes the approach unreliable.

In Table 2.6 are presented the performance metrics for FM.

2.11.6. Ultrasound

Ultrasound waves with a short wave length can be used for localization purposes indoors. The system requires a direct line of sight between a transmitter and a receiver.

Accuracy Range	Cost	Power consumption
2m to 4m	Low	Low

Table 2.6. Performance metrics for FM technology

An indoor positionig system using ultrasund-based positioning is developed in [21], it uses a combination of RF and ultrasound technologies.

In Table 2.7 are presented the performance metrics for ultrasound.

Accuracy Range	Cost	Power consumption
1cm to 1m	Medium	Low

Table 2.7. Performance metrics for ultrasound technology

2.11.7. Bluetooth

Bluetooth technology is a wireless technology that enables data sharing among devices in a short range. Bluetooth devices are low cost, low power, and integrated in most smartphones. A Bluetooth device tag has a unique ID, the ID can be used for localization. The main disadvantage of the technology is the discovery latency time that ranges from 10 to 30 seconds, which slows down an indoor positioning system.

In Table 2.8 are presented the performance metrics for bluetooth.

Accuracy Range	Cost	Power consumption
2m to 5m	Low	Low

Table 2.8. Performance metrics for bluetooth technology

2.11.8. Light Emitting Diode (LED) Lamps

LED lamps, can be used for indoor localization purposes. Using triangulation techniques the light can be processed by a mobile device using a light sensor to infer the localization of the users.

2.11.9. Magnetic Fields

Anomalies of ambient magnetic fields inside buildings can also be used for the purpose of location indoors. The convenience of a systems based on magnetic fields is that it does not need any physical infrastructure. The way the system works is by using a magnetometer embedded in smartphones that compares the sensor measurement with the magnetic map that has been created for the building in advance. Since the magnetic fields are high susceptible to noise, a low accuracy is obtained using this system.

Pathapati et al.[24] present an indoor localization method based on ambient magnetic fields. the authors classify the ambient signatures of indoor hallways using a dynamic time warping approach; they report an accuracy of 92.6% for a building with 26 hallways and 91.1% for a building with 15 hallways.

2.11.10. UWB Ultra Wide-band

This technology is characterized as having high accuracy capabilities in the range of 20 to 30 cm. and it uses TDoA, ToA and traversed time to determine the range between the UWB transmitter and UWB receiver. UWB is based on using ultra short pulses, typically less than 1 nanosecond, with a low duty cycle.

2.11.11. SmartSlam

SmartSlam [11] is a smartphone-based SLAM approach to solve the indoor positioning problem. The system developed in the paper has the particularity that it can work in a completely unknown environment and built its own map with use. SmartSlam selects from a variety of algorithms one that is best suited for the current scenario. The system accuracy depends on the environment, for the results shown the accuracy vary from 2 meters to 3 meters.

CHAPTER 3

MATLAB SIMULATION OF ROOM-LEVEL ACCURACY OF A WIFI INDOOR LOCALIZATION SYSTEM AT UNT'S DISCOVERY PARK

In this chapter is presented a room-level precision indoor localization system using a deterministic-algorithm based approach. The selected rooms is located at the Discovery Park building first floor. The implementation of the IPS was using the MATLAB software.

3.1. Weighted K-Nearest Neighbor Algorithm

Assuming M offline fingerprints, the euclidean distance vector $\mathbf{D}$ between the measured online fingerprint vector represented as $\mathbf{r}$ and the offline fingerprint vector represented as $\mathbf{f}$ can be calculated as:

$$\mathbf{D} = \sqrt{\sum_{i=1}^{N} |r_i - f_i|^2} \tag{11}$$

The weighted vector can be define as $\mathbf{W} = \dfrac{1}{\mathbf{D}}$. The k nearest offline points are sorted from low to high and the best weights are selected for positioning.

Then the estimated position coordinates are:

$$x = K \sum_{i=1}^{n} w_i x_i \tag{12}$$

$$y = K \sum_{i=1}^{n} w_i y_i \tag{13}$$

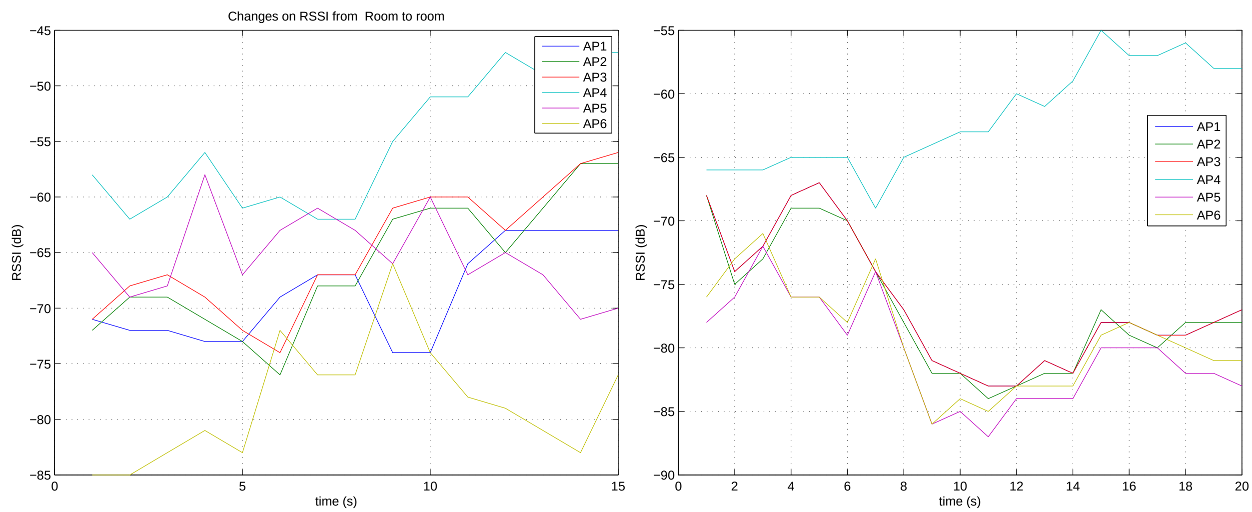
where the normalization factor K becomes:

$$K = \frac{1}{\sum_{i=1}^{n} w_i} \tag{14}$$

3.2. Characterization of Rooms Using RSSI Readings from Nearby Access Points

At UNT's computer science and enginering and electrical engineering departments, the changes on RSSI readings coming from available access points were recorded using a Nexus 4 smartphone, the experiment considered changes in locations of the mobile device in adjacency rooms, the results are presented in Figure 3.1

According to the observations in Figure 3.1, a change of approximately 10 dB was observed on each access point reading when going from one room to another. This change shows that the same access point can have a significant change at 2 adjacent rooms; The experiment let us to infer that room-level accuracy can be obtained with only fingerprints located inside rooms.



(A) CSE department room 244 & room 245 (B) EE department room F220 & room F221

Figure 3.1. RSSI variation from room to room

f	1	2	3	4	5	6	7	8	9	10
x	158	156	160	160	192	188	189	185	218	218
y	729	700	653	623	733	702	653	622	622	657

Table 3.1. Fingerprints from 1 to 10 for the MATLAB simulation

f	11	12	13	14	15	16	17	18
x	223	220	162	182	160	189	218	220
y	703	729	584	554	553	581	553	580

Table 3.2. Fingerprints from 11 to 18 for the MATLAB simulation

In Table 3.1 and Table 3.2 can be appreciated the 18 fingerprints (f) that are shown in Figure 3.2. The 18 fingerprints have a location associated with each fingerprint; the location is expressed as a pair of coordinates (x, y).Figure 3.2 shows the mapping of the fingerprints and calculated localization of the user at the first floor of Discovery Park.

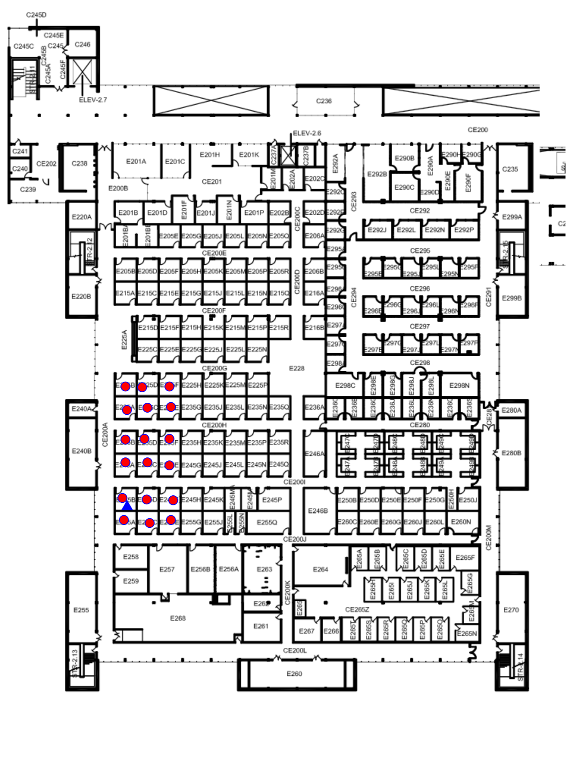


Figure 3.2. Mapping of the fingerprints

In Figure 3.2 is presented the distribution of the fingerprints (red dots). In the same figure, the blue rectangle represents the current location of the user as calculated by the weighted K-Nearest neighbor algorithm.

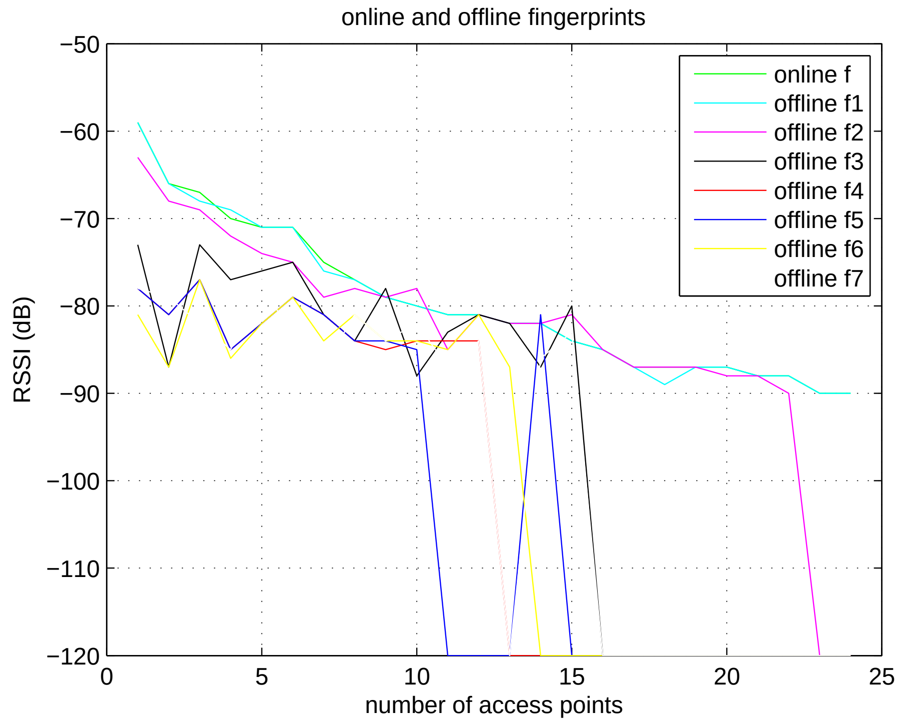


FIGURE 3.3. RSSI levels from 7 fingerprints

In Figure 3.3 8 signal samples, 7 from offline fingerprints and 1 from an online fingerprint are presented with their corresponding RSSI values. In the figure it can be observed that the level of the signal changes according to the location; the signal more similar to the online fingerprint is selected.

For this experiment, the signal most similar to the online fingerprint is the signal from offline fingerprint, $f2$, so for $K = 1$ the location assigned to fingerprint f2 is consequently assigned to the online fingerprint.

Figure 3.3 shows that the number of access points detected change per fingerprint, the value depends on the number of available access points at the location (x, y) of the fingerprint.

3.3. Conclusions

The Indoor positioning system presented in this section successfully provided room level accuracy as it can be observed in Figure 3.2.

Although the system is able to localize persons with room level accuracy; when considering $K = 1$, even a small error in the localization system assigns the location of the user to an incorrect room, which carries a large localization error. Another problem is that since localization is only achieved in rooms, peoples in hallways are never localized.

The problems presented in this indoor positioning system are overcome with the main contribution of this thesis, as presented in the following chapter.

In the next chapter, the main contribution of this thesis is presented, it considers fingerprints that have a different number of set of access points, as shown in Figure 3.3. The fingerprints are created and modified according to a set of dynamic access points.

CHAPTER 4

DYNAMIC WLAN FINGERPRINTING INDOOR POSITIONING SYSTEM

4.1. Introduction

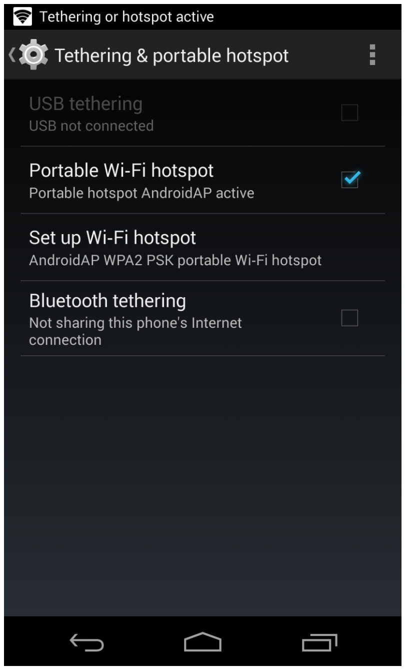
It is commonplace that people who perform daily activities indoors stay at the same place, for example, students in a classroom, professors at their offices, receptionists in a lobby area and so on. Moreover, they usually frequent the same places while indoors, that is, students going to the same classroom and professors having a meeting in the same room at a specific time of the day. The work of this thesis aims to take advantage of those observations and patterns to create an IPS that relies solely on WIFI hotspot signals from devices carried by occupants and also to improve the accuracy of existing IPS based on WIFI fingerprinting.

The adoption of smartphones has grown exponentially all over the world [2], due to their extensive functionality and decline of price. The level of adoption allows for the inference that a large number of people who stay indoors use smartphones. In this text the term *passive users* refer to people that perform their daily routine activites indoors.

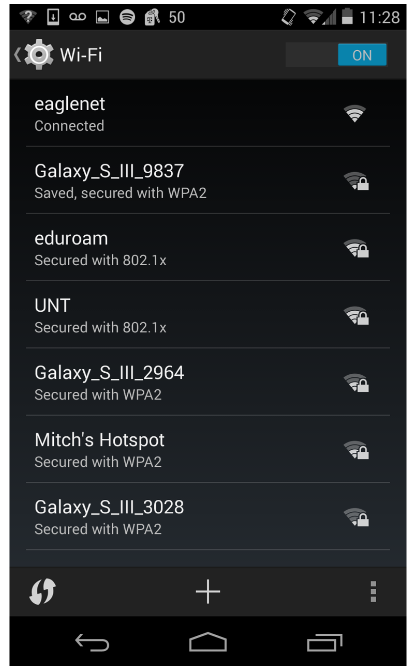
Most smartphones have a WIFI hotspot feature that allows sharing the smartphone's internet connectivity with nearby devices via WIFI; in this scenario, the smartphone behaves as a WIFI repeater. Unlocked smarphones as the Nexus 4, that have installed the Android original firmware developed by Google can activate the WIFI hotspot feature without having a WIFI hotspot plan with the phone's carrier; Figure 4.1 illustrates the concept. In these scenarios the user is asked to activate the service once a request to connect to the internet is being made from a device using the WIFI hotspot feature.

4.2. Thesis Contribution

The contribution of this thesis is based on the use of the smartphone's WIFI hotspot feature not only as a feature to extend the internet connectivity to other users as it is originally intended, but also as a tool to increase the accuracy of a WLAN IPS; furthermore,



(A) WIFI hotspot activation



(B) Visibility of 4 wifi hotspot devices

FIGURE 4.1. Thetering option and WIFI hotspot in a nexus 4 smarpthone

the walking patterns of users is analyzed using machine learning algorithms for prediction of the dynamics of the system.

4.2.1. Dynamic Access Points and Fingerprints (DAF)

At every place where passive users become stationary, they create a dynamic fingerprint. The dynamic fingerprint contain a set of access points already available in the infrastructure and a set of dynamic access points, which are created from the signals coming from other passive users when they are in a stationary state. DAFs are modified in real time according to the dynamics of the IPS.

4.2.2. Improvement of the Performance of an IPS

The performance of the proposed dynamic IPS was evaluated in terms of accuracy as the dynamics of the system changes, compared with the standard WIFI fingerprinting positioning system and when more users contribute to the IPS.

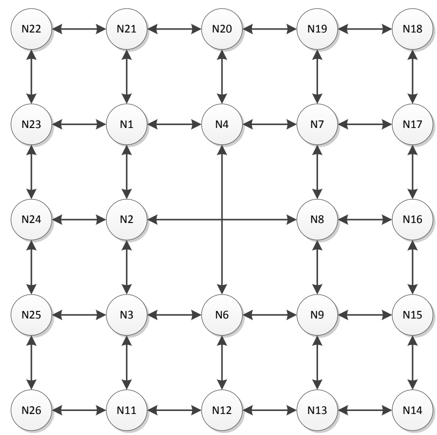


FIGURE 4.2. Ad-hoc wireless network consisting of WIFI hotspot nodes

The accuracy of the system was compared using various access point selection strategies and indoor positioning algorithms which include probabilistic and non probabilistic approaches. The testing was performed at UNT's Discovery Park electrical engineering and computer science and engineering departments. The results in terms of accuracy are presented in the last section of this chapter.

4.2.3. Configurations of the Proposed IPS

There are many buildings were there is no WLAN infrastructure or where not many fixed WIFI access points are deployed; in this case, an IPS can be created with the approach presented in this thesis. 2 configurations are presented using this approach.

4.2.3.1. No WLAN infrastructure deployed and passive users available

An ad-hoc wireless network can be created from only WIFI hotspots when there is no WIFI infrastructure deployed. The particular network is presented in Figure 4.2. Nodes from N1 to N22 represent a static stationary user sharing the mobile WIFI hotspot connectivity from his phone. Every node can detect the signal of every other node when all the nodes become static.

In order to determine when a user is stationary or dynamic, machine learning algorithm were used for prediction.

Source node	MAC address	RSSI
N2	00:24:6c:c1:c1:80	-53
N3	00:1a:1e:85:a4:11	-67
N4	00:1a:1e:87:04:c2	-67
N5	00:1a:1e:85:a4:02	-60

TABLE 4.1. Pair of mac address and RSSI from 4 nodes at node 1

The nodes have only 2 states:

- Moving Node (Deactivated) when the nodes start moving they are not being considered as in the system.
- Static Node (Activated) when a node is static its received signal strength is considered to help localize other nodes.

Table 4.1 is an example of a fingerprint created at N1 when signals from 4 other nodes are available (static). This case implies that the rest of the nodes are in a moving state (dynamic).

Chen et al.[6] reports that at least 3 access points are required for a WLAN fingerprinting to function correctly, so fingerprints with fewer than 3 fingerprints were not considered for positioning.

All the fingerprints combinations of at least 3 other nodes are calculated at the server, the server coordinates the content of the dynamic fingerprints according to the movement pattern of the users.

4.2.3.2. WLAN infrastructure and stationary users available

Figure 4.3 shows a combination of fixed WIFI nodes with WIFI hotspot nodes. The white color represents WIFI hotspot nodes and black color represents fixed WIFI nodes.

This configuration has the following types of fingerprints:

- Fixed WIFI fingerprints

  Fingerprints created solely from fixed WIFI nodes are deployed as a backup system

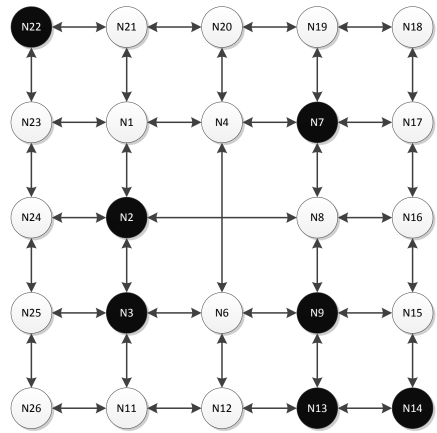


Figure 4.3. Combination of fixed WIFI nodes with WIFI hotspot nodes

when no WIFI hotspot nodes are available. This case exemplifies a standard non-dynamic fingerprinting indoor positioning system.

- WIFI hotspot fingerprints

  The fingerprints created from signals coming from WIFI hotspots nodes from stationary users contain fixed WIFI access points as well, which are already deployed in the infrastructure. These types of fingerprints have the particularity to be updated in real time, as users move. The configuration require the creation of all possible combinations of dynamic fingerprints available.

4.2.4. Applications

4.2.4.1. Natural disasters

The WLAN infrastructure in buildings can severally be affected as a result of a natural disaster; in those situations, the approach presented in this thesis could be the only available option for indoor positioning. The position system can help a rescue team find people trapped indoors. A rescue team would be able to use an existing dynamic positioning system if the trapped users indoors are able to share their mobile WIFI hotspot connectivity when the WIFI infrastructure is damaged.

4.2.4.2. Increase the coverage of an existing IPS

The coverage of an already deployed IPS can be expanded if the approach presented in this thesis is considered in the case when creating WIFI hotspot nodes where the WLAN infrastructure is not available.

4.3. Step Detection Using the Accelerometer

The accelerometer embedded in most smartphones measures the acceleration of the smartphone in the x (lateral), y (longitudinal), and z (vertical) axes. The value obtained from the 3 axes are represented by floating point values and are expressed in meters per second squared. The data is further analyzed to help yield better accuracy in the localization task. Figure 4.4 presents the variation of the acceleration over time when the user starts walking using a tri-axis Nexus 4 smartphone. The sudden start of change in the amplitude corresponds to the user going from a stationary state to a walking state.

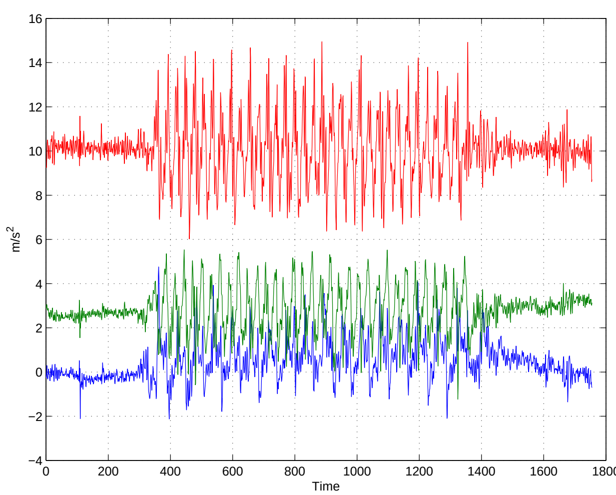


FIGURE 4.4. Change of acceleration in the X,Y and Z axes when movement is detected

In order to infer when a user is moving or not moving and to eliminate small changes in the step detection, Jimenez et al. [14] present a robust approach to step detection:

First, the magnitude of the acceleration a_i for every obtained acceleration sample i is expressed as:

$$a_i = \sqrt{a_{x_i}^2 + a_{y_i}^2 + a_{z_i}^2} \tag{15}$$

The local acceleration variance is obtained to improve step detection and to remove gravity (noise):

$$\sigma_{ai}^2 = \frac{1}{2w+1} \sum_{j=i-w}^{i+w} (a_j - \bar{a_j})^2 \tag{16}$$

where $\bar{a_j}$ is a local mean acceleration value, obtained by $\bar{a_j} = \frac{1}{2w+1} \sum_{q=i-w}^{i+w} a_q$, and w defines the size of the averaging window. for detecting the swing phase and stance phase a threshold is expressed as:

$$A_{1i} = \begin{cases} \text{Threshold 1}, & \sigma_{ai} > \text{Threshold 1} \\ 0, & \text{otherwise} \end{cases} \tag{17}$$

for the swing phase, and as:

$$A_{2i} = \text{Threshold 2, if } \sigma_{ai} < \text{Threshold 2} \tag{18}$$

for the stance phase.

A step is detected in an acceleration sample i at the end of a swing phase and when the stance phase starts. Also, the following conditions need to be satisfied for the acceleration to be successfully detected as a step:

- transition from high to low acceleration $A_{1i-1} < A_{1i}$
- $max(A_{2i:i+w})$ =Threshold 2

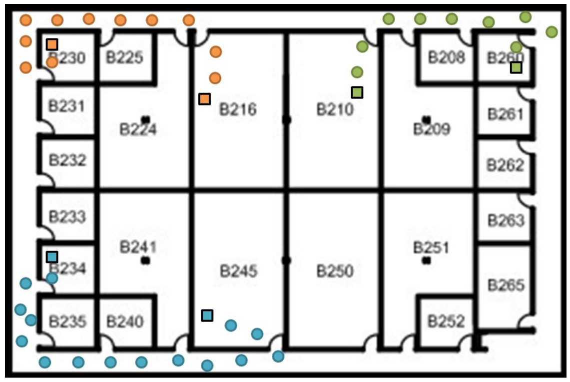


FIGURE 4.5. Electrical engineering department user pattern movement

4.4. User Movement Pattern

Figure 4.5 shows the movement pattern of 3 users that are represented with 3 different colors (Red, Green and Blue). The circles represent when users are moving and the squares represent when the users are static.

As presented in the previous subsection, the moving pattern of the user can be inferred with the accelerometer data coming from the mobile device.

The system is designed in such a way (as it is presented in section 4.3) that small changes in the acceleration are ignored, For example, when the user decides to use the phone or when a trivial change in position of the phone is made.

- Stationary user is static
  The WIFI hotspot feature is only activated when the user is at a static position, This allows for having a constant WIFI signal strength received at the fingerprints to characterize them correctly.
- Stationary user is moving
  When the user starts moving between places, the WIFI repeater is switched off and the access point signal is removed at the fingerprint, since variation on the received signal strength is not useful for positioning.

4.5. User Activity Recognition

A sample of acceleration signal from a mobile device is shown in Figure 4.6, The signal was labeled as dynamic or static according to the movement pattern of a user; each label of the signal is called a feature.

As it can be observed from Figure 4.6; when the user is at a static position the acceleration of the phone changes; those changes express short movement of the mobile device; for example, when the user decides to talk over the mobile device or when the user moves the phone to a different position.

4.5.1. Classification

In this section classification techniques are presented to determine if the user is static or dynamic, the technique that returns the best accuracy is selected for prediction.

Witten et al.[27] define data mining as a tool for finding and describing patterns in a large set of data to make future predictions based on patterns found on the observed data.

Based on data that describes a process or an activity; data mining can make future predictions for a given set of unknown data. The accuracy describes the level of correct predictions that can be made within the current data set.

Several machine learning algorithms from the WEKA [3] machine learning software were used for classification and prediction of the movement patterns of the user.

Figure 4.7 shows the statistics of the Y axis sample set using the WEKA software, which gives the minimum and maximum value of the set. The mean and standard deviation values of the sample set are also presented.

The labeled patterns which are activities of the user categorized as dynamic or static are shown in Figure 4.6, those acceleration samples are used to train the system for prediction.

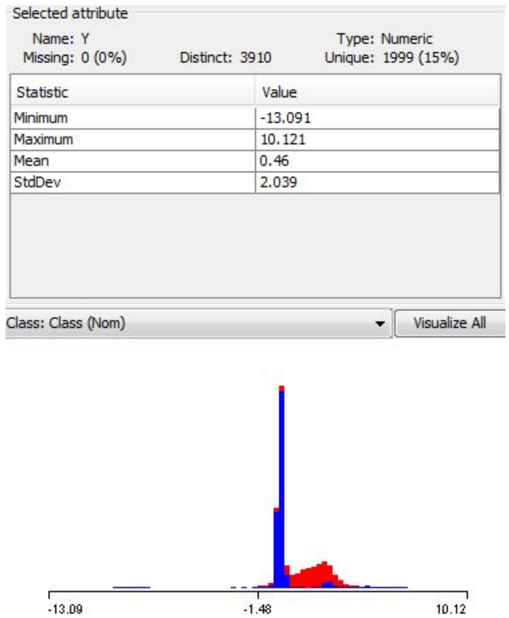


FIGURE 4.7. Statistics of the Y axis features

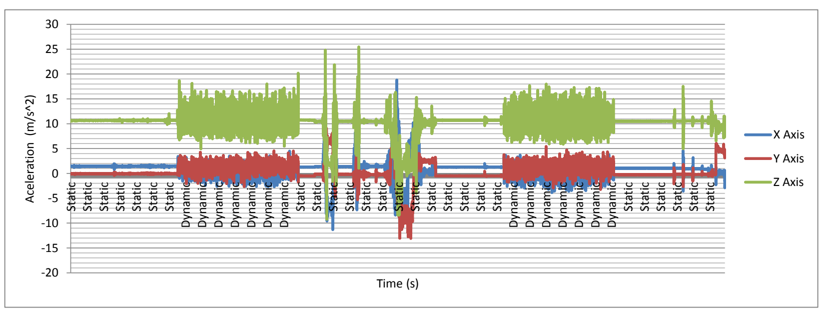


FIGURE 4.6. Acceleration signals from user movement pattern for experiment 1

4.5.2. Parameters Used for Classification

The classification results from WEKA are expressed with the following parameters:

TP Rate: Rate of true positives. It returns the instances correctly classified.

FP Rate: Rate of false positives. It returns the instances falsely classified.

Precision: Instances that are actually from a class divided by all the instances classified as that class

Recall: Fraction of relevant instances that are fetched

ROC area: The area under the ROC curve. An area of 1 represents a perfect classification.

The classification results from WEKA are presented using the Bayes network, naive Bayes, support vector machine and random forest classifiers. All the results were obtained with a cross validation of 10 folds and 66% of percentage split.

4.5.3. Classification Results for Experiment 1

Figure 4.8, Figure 4.9, Figure 4.10 and Figure 4.11 present the classification results for the experiment 1.

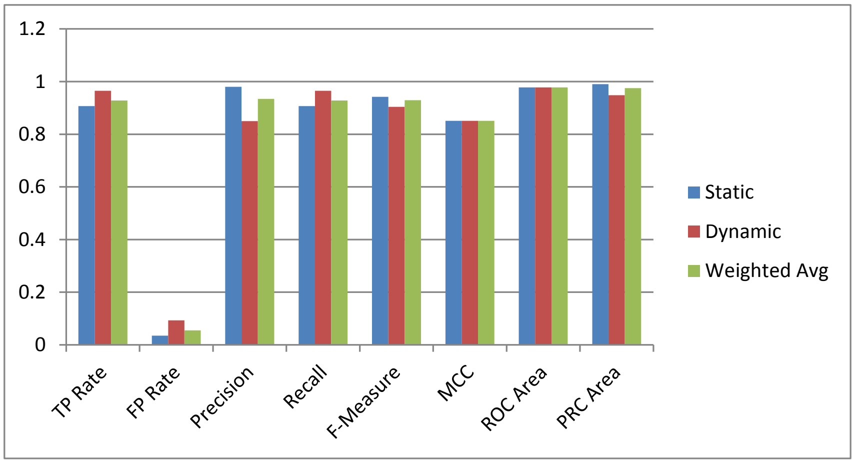


FIGURE 4.8. Classification results for experiment 1 using Bayes network classifier

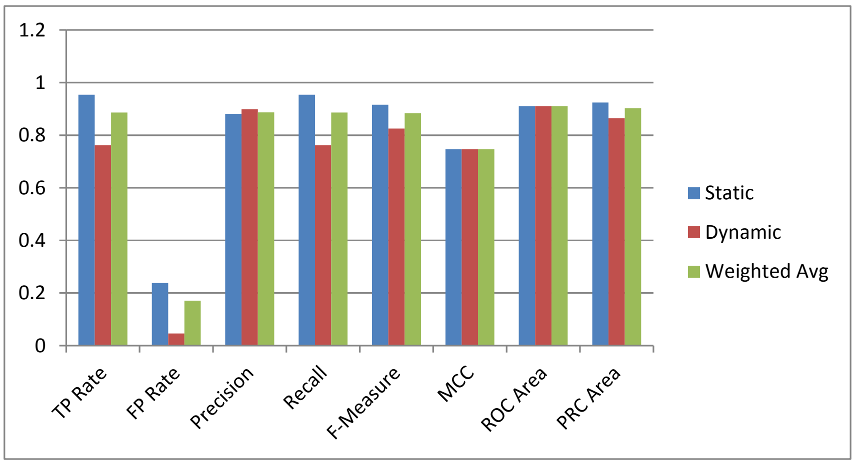


FIGURE 4.9. Classification results for experiment 1 using naive Bayes classifier

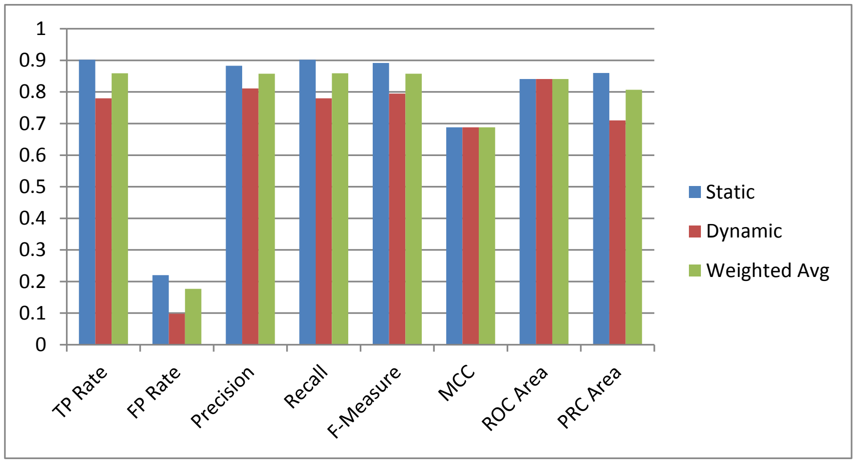


FIGURE 4.10. Classification results for experiment 1 using support vector machine classifier

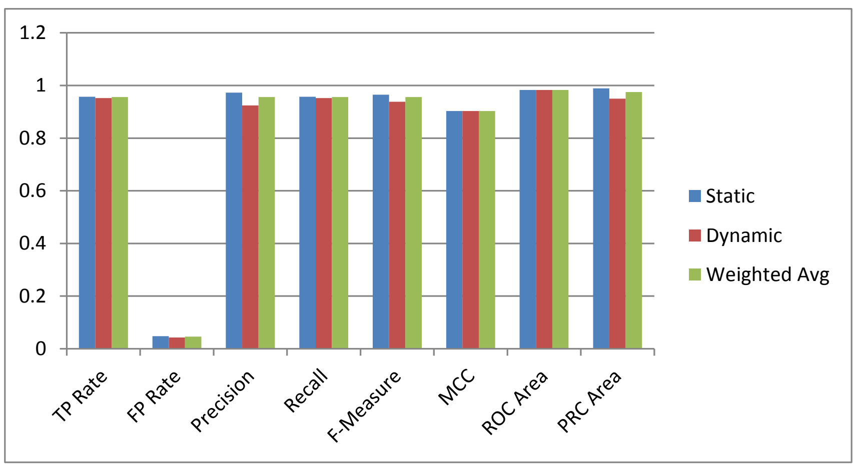


FIGURE 4.11. Classification results for experiment 1 using random forest classifier

4.5.4. Classification Results for Experiment 2

Figure 4.12 shows the acceleration signals from user movement pattern for experiment number 2. Figure 4.13, Figure 4.14, Figure 4.15 and Figure 4.16 present the classification results for the experiment 1.

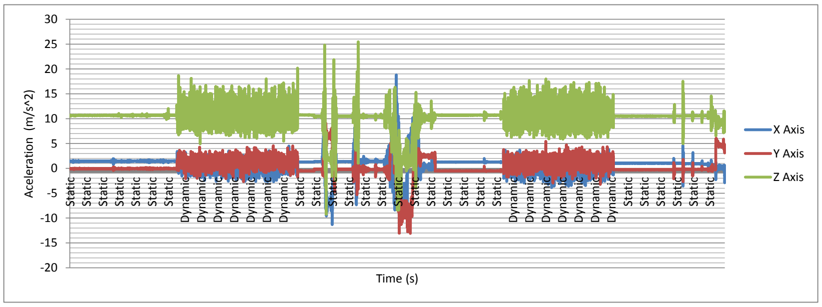


FIGURE 4.12. Acceleration signals from user movement pattern for experiment 2

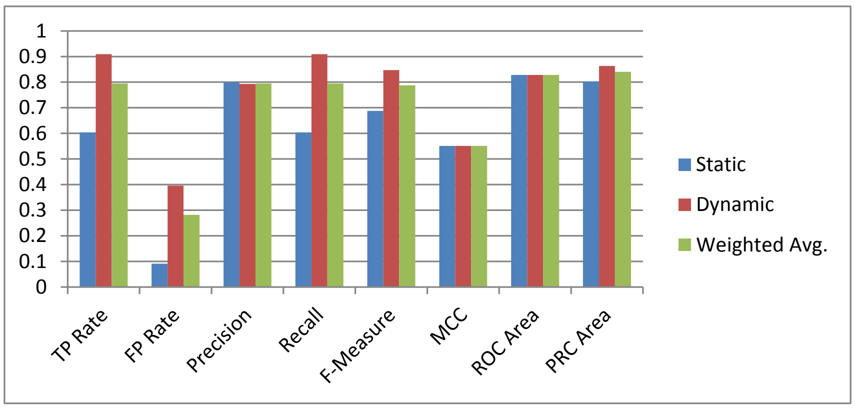


FIGURE 4.13. Classification results for experiment 2 using Bayes network classifier

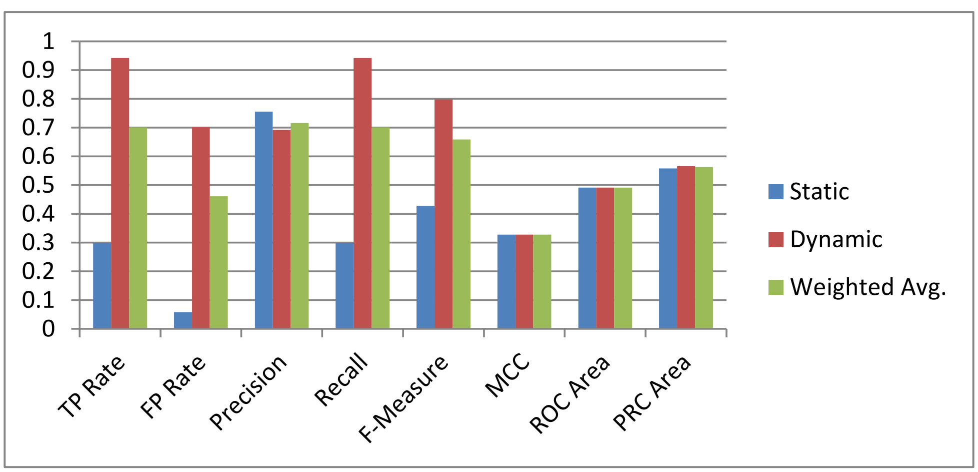


FIGURE 4.14. Classification results for experiment 2 using naive Bayes classifier

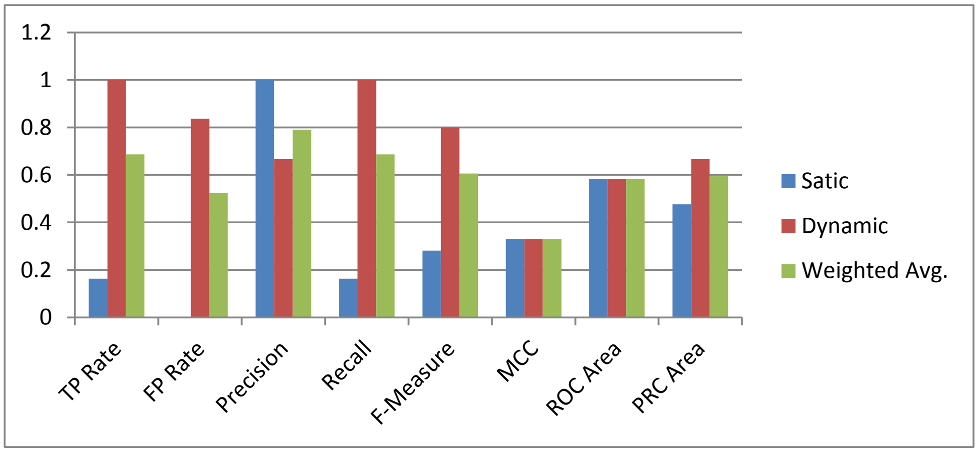


FIGURE 4.15. Classification results for experiment 2 using support vector machine classifier

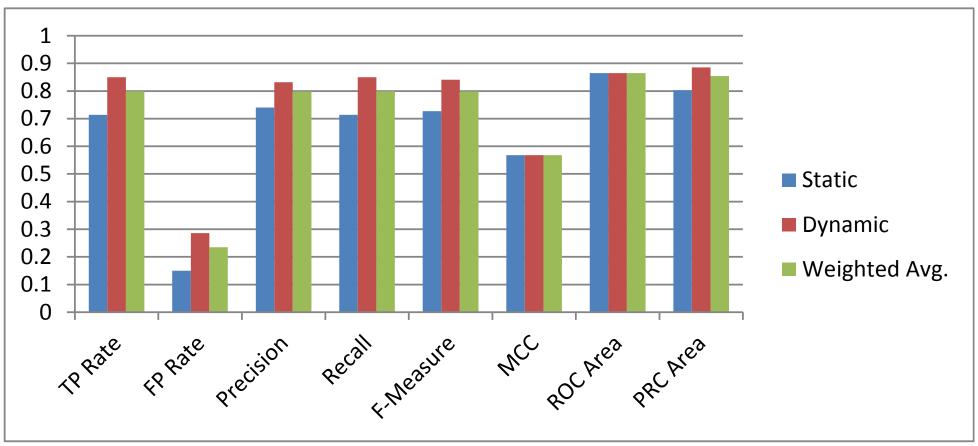


FIGURE 4.16. Classification results for experiment 2 using random gorest classifier

4.5.5. Conclusions of the Classification of the Experiments

According to experiment 1, the random forest classifier returns the best accuracy when compared with the accuracy of the other classifiers, The classifier returns a 95.5504% of correctly classified instances of the data set as presented in Figure 4.6.

According to the data from experiment 2, as it can be observed in Figure 4.12, the random forest algorithm is also the classifier that returns the best accuracy when compared with others classifiers with 79.9195% of correctly classified instances.

Since random forest is the classifier that returns the best accuracy among the classifiers used for experiment 1 and 2, it is the classifier selected for prediction of the user activity as dynamic or static.

4.5.6. Selected Buildings

The selected buildings to test the system is the electrical engineering department main section and the CSE building located at UNT's Discovery Park.

4.6. System Description and Design

This section presents the dynamic WLAN fingerprinting system architecture and its design.

4.6.1. Dynamic Fingerprints

- Creation

  The sum of fixed WIFI access points and the temporary access points created by stationary users constitute a dynamic fingerprint. The term dynamic is used as change is being made to the system every time a stationary user is added or removed, which depends on the movement pattern of the stationary users. Each dynamic fingerprint contains a set of dynamic access points that are added to, or removed from the fingerprint according to the moving pattern of the stationary user. A dynamic access point is associated to the availability of the WIFI signal coming from a passive user. The dynamic fingerprints have the particularity that they are only created at location where passive users are static; those locations can be inferred by the analysis of the movement patterns of the users.

- Update

  The dynamic fingerprints are updated when a user is added into the system and considering the change of state of the stationary users.

A record of fingerprints and accelerometer data can be expressed as $R_t = f(F, Ac)$ where F represents the WIFI signal fingerprint and Ac represents the recorded accelerometer data. if n access points are available in the building, then the fingerprint can be represented as $F = [f_1, f_2, ..., f_n]$ where f_i denotes the RSS value of the ith access point.

In order to categorize the motion of the user as dynamic or static the accelerometer readings Ac was saved over time for further analysis using the Random forest algorithm.

The training data for our system is the initial collection of data containing the dynamic fingerprints associated with the static and dynamic accelerometer data, after several samples are obtained from stationary users.

4.6.2. Enhancing an Existing WLAN Fingerprinting System

An existing WLAN fingerprinting system can be improved in terms of accuracy when dynamic fingerprints are considered. The WLAN fingerprinting system presented in Figure 4.17 considers only fixed WIFI fingerprints by red dots. The dynamic fingerprints help by providing temporary signals that yield better accuracy.

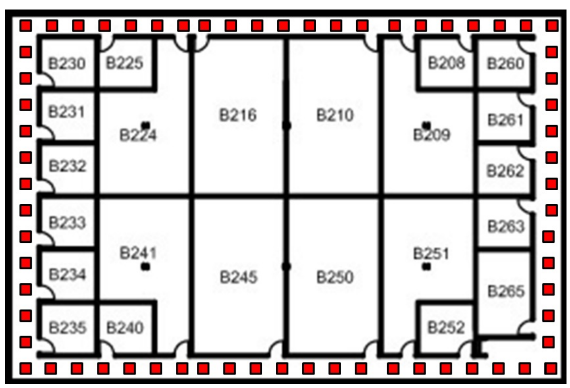


FIGURE 4.17. Conventional WLAN fingerprinting system

4.7. Android Mobile Application for the Online and Offline Phase

The project was developed using the Java programming language for Android devices. Part of the source code is based on the open-source platform Airplace [15]. Android was selected as the mobile platform for the IPS since it provides more hardware manipulation than iOS; another reason for the selection is that Android devices are used more extensively in the world than iOS devices.

4.7.1. Introduction

For this thesis 2 separated Android applications and a server were developed, as shown in Figure 4.18.

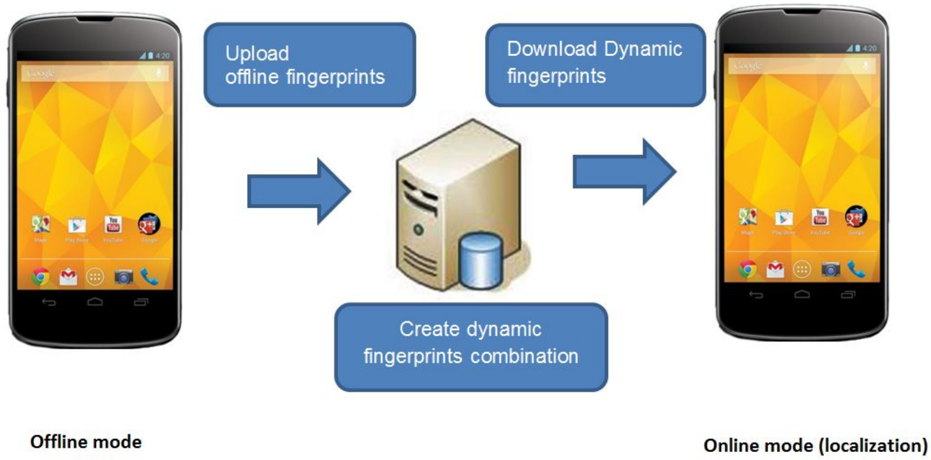


FIGURE 4.18. Android applications and server interaction

The offline and online applications were developed to achieve the following constraints:

- Low communication overhead and minimum database update

  In order to minimize the communication overhead between the device and the server and to prevent the database to be updated when obtaining new fingerprints, the server-device interaction was reduced as much as possible to lower the communication cost as much as possible.

- Privacy

  In order to encourage user privacy, the algorithms for positioning was running entirely on the mobile devices to prevent any undesired intrusion form the server, once the dynamic fingerprints are downloaded by the user, no other connection between the server and the user is necessary.

4.7.2. Offline Phase

The offline phase generates a *radiomap* which contains a set of fingerprints. The fingerprints can contain fixed WIFI access points and dynamic access points. The fingerprints can be created via crowdsourcing (several users) or by a single user. In the case of crowdsourcing, the server appends and combines all the fingerprints created by several users to create the final radio map that is used for positioning. The users can select how many samples per fingerprint they want to collect and the time between samples can also be adjusted. The users then must upload the collected fingerprints to a main server for further analysis and distribution.

4.7.3. Online Phase

During the online phase, the user has to connect to a main server and download the stored fingerprints that were created in the offline phase; also, specific-algorithm parameters that yield the lowest calculated accuracy from a testing data set are used for real time positioning.

The user can select the algorithm used for positioning in real time. The parameters for the lowest accuracy provided by the server cannot be modified.

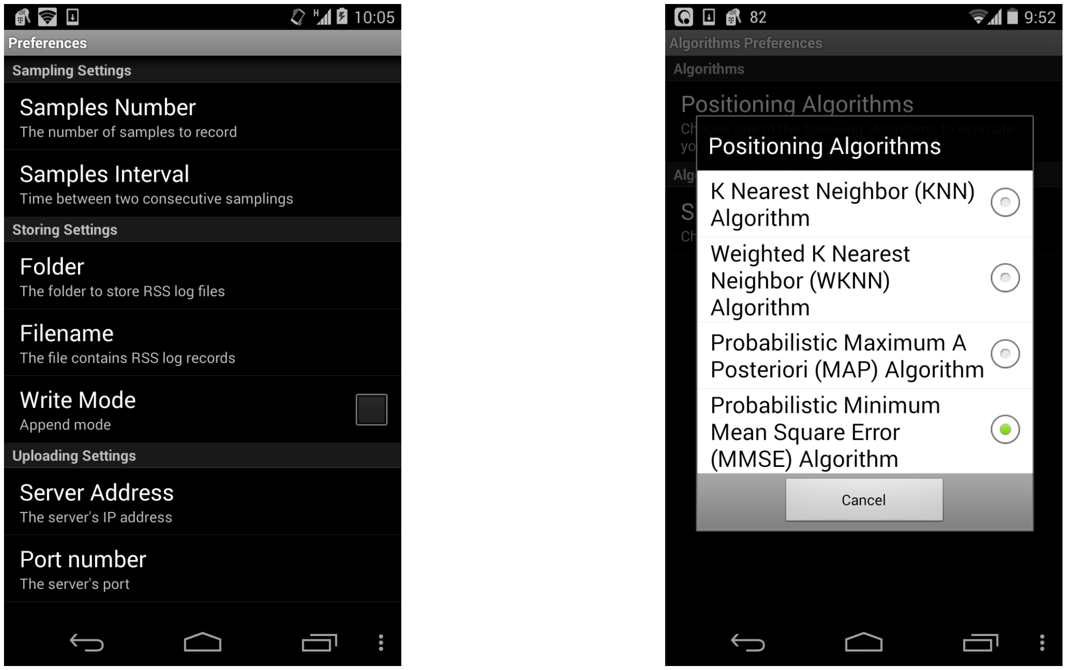
Figure 4.19 shows the settings of the offline android application and the algorithm selection of the online android application.

4.8. Server

The server, also developed in Java, receives the dynamic and non-dynamic fingerprints captured by the users. The server processes the raw fingerprint data by calculating the mean of the fingerprints obtained at the same location and then it calculates the algorithm-specific parameters for the probabilistic and deterministic approaches; depending on the value of those parameters the accuracy of the system change as it is presented in the results section. The parameters with the best accuracy for each algorithm is returned to the user for real-time positioning. The parameters selected cannot be changed by the online application.

4.9. Algorithms Used for Positioning

In this subsection, the characteristics of the positioning algorithms used for the IPS of this thesis are presented.



(A) Offline application settings

(B) Online application algorithm selection

FIGURE 4.19. Offline and online android applications

To evaluate the results of this thesis 4 algorithms are used for positioning. 2 deterministic, namely the k-Nearest Neighbor (KNN) [4] and the weighted k-Nearest Neighbor (WKNN) [16] algorithms and 2 probabilistic namely the maximum a posteriori (MAP) [29] and the minimum mean square error (MMSE) [22] algorithms.

Those algorithms were selected since they exemplify the deterministic and probabilistic approaches in a weighted and non-weighted manner.

4.9.1. K-Nearest Neighbor and Weighted Nearest Neighbor Algorithms

For this thesis, 2 deterministic algorithms, the k-Nearest Neighbor algorithm and the weighted nearest neighbor algorithms, were implemented for positioning.

The algorithm considers N offline fingerprints $\mathbf{f}$, the online fingerprint vector is represented by $\mathbf{r}$, and a vector of approximated locations $\mathbf{l}$ can be obtained by calculating the euclidean distance between each element i of the online and offline fingerprints.

Calculating the inverse of the euclidean distance the weight w of each approximated location can be obtained as:

(19) $$w_i = \frac{1}{|f_i - r_i|}$$

The approximated location $\hat{l}$ can be obtained as:

(20) $$\hat{l} = \sum_{i=1}^{K} \frac{w_i}{\sum_{j=1}^{K} w_j} l_i$$

The approximated location $\hat{l}$ is ordered according to increasing the distance between the offline and online fingerprints $|f_i - r_i|$

For the K-Nearest Neighbor algorithm:

(21) $$K \geq 1 \text{ and w is expressed as } w_i = \frac{1}{k}$$

For the weighted k-Nearest Neighbor algorithm:

(22) $$K \geq 1 \text{ and w is expressed as } w_i = \frac{1}{|f_i - r_i|}$$

4.9.2. Variation of Parameters for Deterministic Algorithms

The value of K neighbors is varied from 1 to 15 at the server; then the K that returns the less positioning error is selected, and the value changes according to the estimated positioning values obtained by the deterministic algorithms from the testing data.

4.9.3. Maximum A Posteriori and Minimum Mean Square Error Algorithms

For the probabilistic approach, 2 algorithms were implemented, namely, the maximum a posteriori and the mean square error algorithms.

The probabilistic approach is based on calculating the probability of location $\hat{l}$ given the signal s as:

(23) $$P(l_i|s) = \frac{P(S|l_i)P(l_i)}{P(S)} = \frac{P(S|l_i)P(l_i)}{\sum_{i=1}^{l} P(S|l_i)P(l_i)}$$

The MAP algorithm obtains the estimated location $\hat{l}$ as:

$$\hat{l} = \arg\max_{l_i}[P(s|l_i)p(l_i)] \tag{24}$$

The MSSE algorithm obtains the estimated location $\hat{l}$ as:

$$\hat{l} = E(l|s) = \sum_{i=1}^{l} l_i p(l_i|s) \tag{25}$$

4.9.4. Variation of Parameters for Probabilistic Algorithms

For each location of the radiomap a *probability* or likelihood of the user being at that specific location is assigned according to the similarity between the online and offline fingerprints.

The probability P is obtained using the following equation:

$$P = \prod_{i=1}^{n} e^{\dfrac{-(v_i^1 - v_i^2)^2}{\sigma^2}} \tag{26}$$

where v_i^1 and v_i^2 are the ith values from the RSSI of the radiomap (offline fingerprint) and the RSSI values being observed (online fingerprint), respectively.

For the experiments performed in this thesis, the equation (26) is used to vary the values of the parameter σ from 1 to 15. The σ that returns the less positioning error is selected. The value of σ changes according to the estimated positioning values obtained by the probabilistic algorithms. The testing data is used to determine which parameter returns the fewest errors for the positioning algorithm.

4.9.5. Importance of the variation of Parameters in the Deterministic and Probabilistic Algorithms

Shin et al.[23] proposed the enhanced weighted k-Nearest Neighbor algorithm that improves accuracy of an indoor positioning system by varying the number of K neighbors considered for positioning in real time, during the online phase of the system.

The premise for studying indoor positioning systems with a varying K is that there should only be considered K neighbors that are at a small distance to the location of the user. The number of nearest neighbors relevant for positioning depend on the type of indoor area. For example in a corridor there are fewer nearest neighbors relevant for positioning than the ones available in an open area.

In the case of the probabilistic algorithms the ith probability of a location l_i given the signal s changes according to the value of the parameter σ. The optimal value of σ also depends in the area used for positioning.

The accuracy results presented in this thesis, depends on the change of the K for the deterministic algorithms and σ for the probabilistic algorithms. The accuracy results are expressed as an average positioning error.

4.9.6. Performance of Deterministic and Probabilistic Algorithms

According to the literature [18], The probabilistic algorithms perform better than the deterministic algorithms. For this thesis, the type of algorithm is not the only characteristic that affects the performance of the algorithms, but also the parameters k for the deterministic and σ for the probabilistic approaches.

4.10. Calculation of Accuracy

The accuracy is measured by obtaining an average positioning error expressed in meters. Since it is unfeasible and impractical to obtain the average positioning error empirically, a large set of test data is used for this purpose.

The error is obtained by calculating the error of a large set of test data, the test data contain a collection of fingerprints associated to a location; each test data fingerprint is processed by a positioning algorithm. The obtained estimated location is compared with the real location by calculating the Euclidean distance between the *real* and the *estimated* positions to obtain the deviation between the original and the estimated value.

The *true* ith location are the coordinates (x_i, y_i) were the user captures an offline fingerprint to be used as a test data. An *estimated* jth location (x_j, y_j) is obtained by processing the true location of the user by the positioning algorithms. The relation can be expressed as follows:

$$(x_j, y_j) = f(x_i, y_i) \tag{27}$$

The average positioning error relies on the true and estimated location to calculate a value that correctly expresses the accuracy of the system.

The process to obtain the average positioning error is by the summation of positioning errors obtained per location and divided by the total number of locations, as it can be expressed in the following equation:

$$APE = \frac{\sum_{i=1}^{N} PE_i}{NP} m \tag{28}$$

APE represents the average positioning error expressed in meters, PE_i represents the positioning error of the ith location expressed in meters and NP is the total number of positions.

The position error is calculated as the euclidean distance between (x_i, y_i), which is the real location and (x_j, y_j) which is the approximated location. The PE is obtained as:

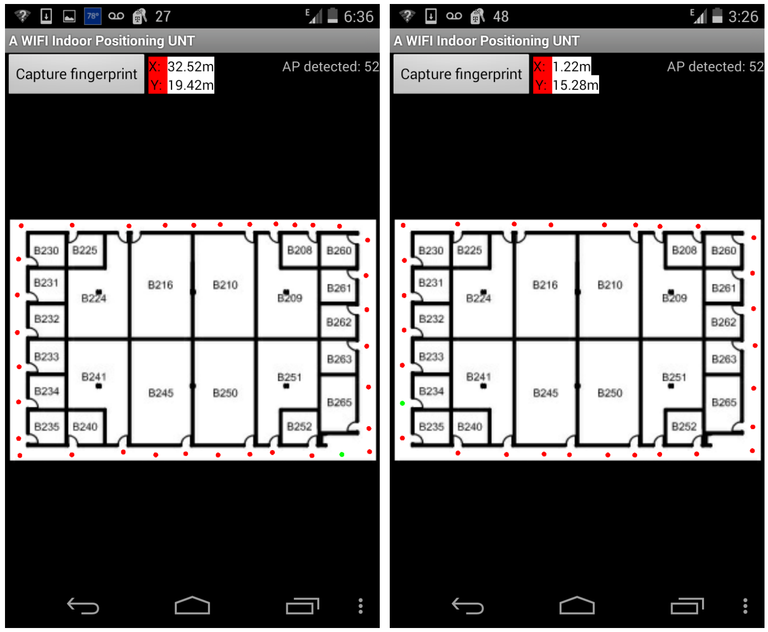
$$PE : \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \tag{29}$$

4.11. Experiment 1: Dynamic WIFI Fingerprinting System at UNT's Electrical Engineering Department Using MaxMean

The first experiment of the approach presented in this thesis was performed at the electrical engineering department of UNT.

The dimensions of the selected area of the electrical engineering department are 36.18 meters wide by 20.21 meters high. The dimensions are included in the android application to precisely calculate the accuracy error.

Figure 4.20a and Figure 4.20b shows the offline and testing fingerprints used for the positioning system.



(A) Offline fingerprints (B) Testing fingerprints

FIGURE 4.20. Offline and testing fingerprints

4.11.1. Dynamic Fingerprints and Access Points (DAF)

For this experiment the $MaxMean$ access point selection criterion is employed as described in section 2.9.

The results are expressed as the average positioning error, as the dynamic access points and the value of the parameters change.

10 access points from the UNT's Discovery Park building infrastructure and 10 WIFI hotspots from 10 volunteers were considered in the dynamic indoor positioning system.

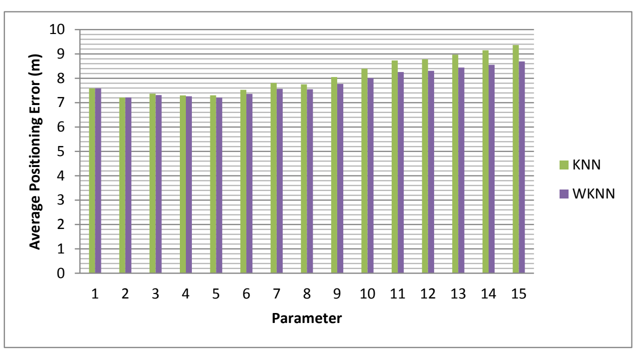
4.11.2. Location of the DAF

The locations of the WIFI hotspot users are shown in Figure 4.21. Also the pattern of the user to determine when the user is moving or not moving is studied and consequently classified.

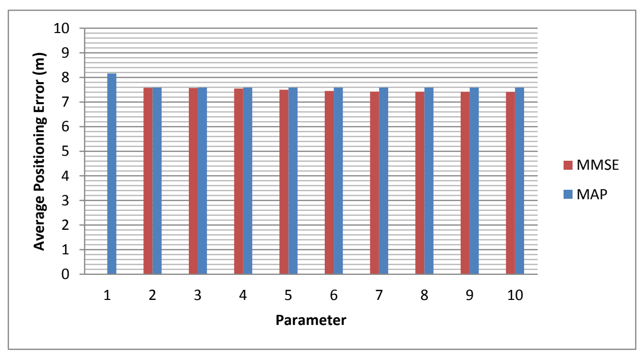
FIGURE 4.21. Location of dynamic access points and dynamic fingerprints

4.11.3. Results for the Non-Dynamic IPS

The APE of the positioning system is presented in this section without considering DAF. The APE is obtained for deterministic and probabilistic algorithms as shown in Figure 4.22a and Figure 4.22b. As it can be observed in Figure 4.22 The less APE considering the fixed WIFI access points deployed is of 7.2 m for the WKNN algorithm at parameter 2.



(A) Deterministic algorithms

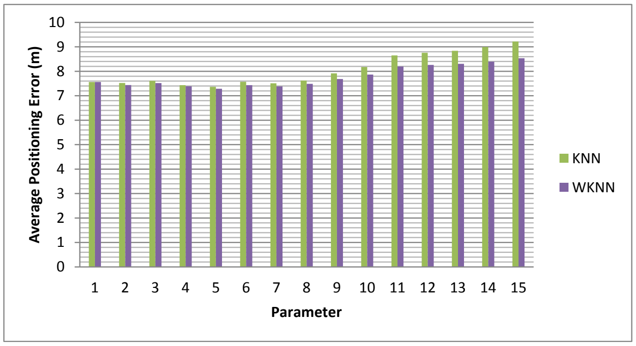


(B) Probabilistic algorithms

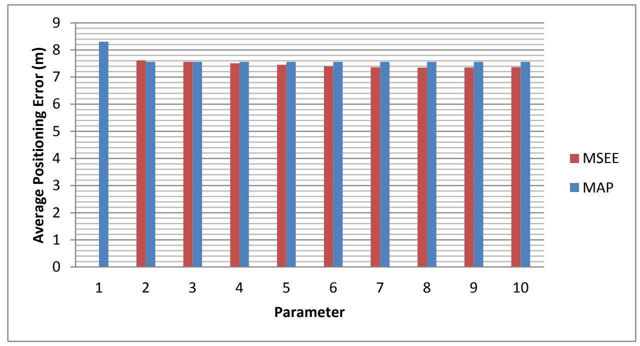
FIGURE 4.22. Average positioning error of the non-dynamic IPS

4.11.4. Adding DAF to the Original Positioning System

The figures from Figure 4.23 to Figure 4.27 show the results in terms of accuracy, using deterministic and probabilistic algorithms for the original system when considering from 2 to 10 DAF, a step of 2 DAF is considered.

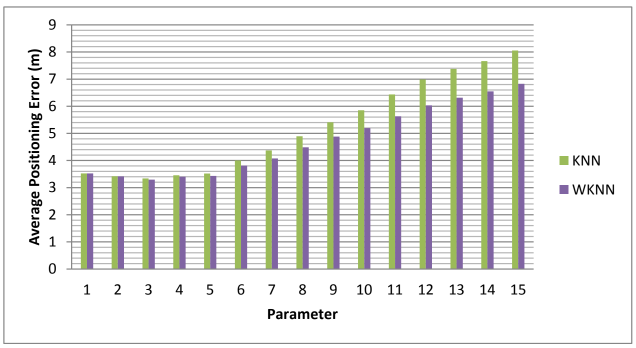


(A) Deterministic algorithms

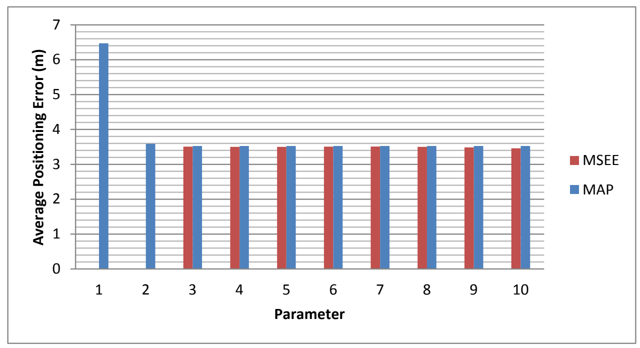


(B) Probabilistic algorithms

FIGURE 4.23. Average positioning error adding 2 DAF to the IPS

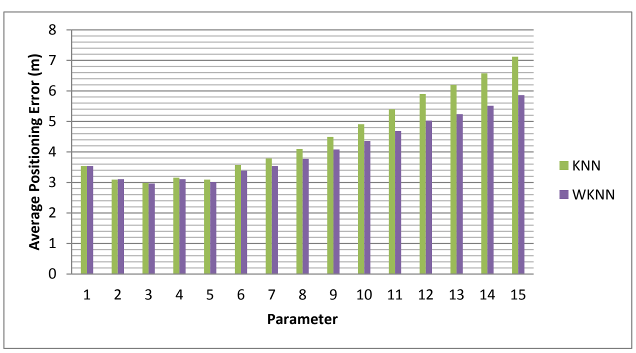


(A) Deterministic algorithms

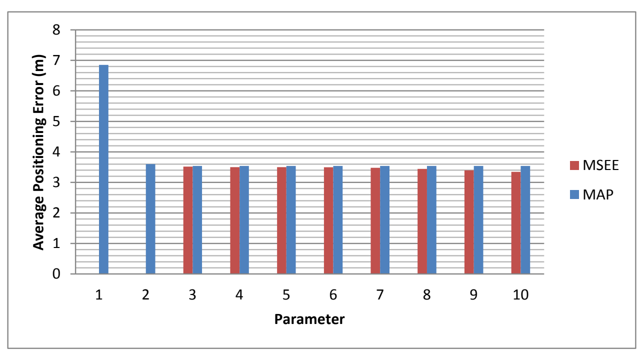


(B) Probabilistic algorithms

FIGURE 4.24. Average positioning error adding 4 DAF to the IPS

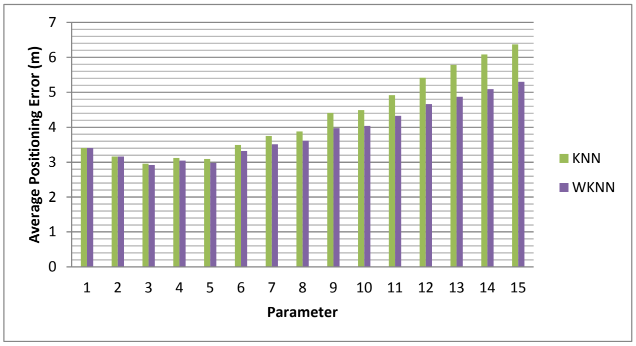


(A) Deterministic algorithms

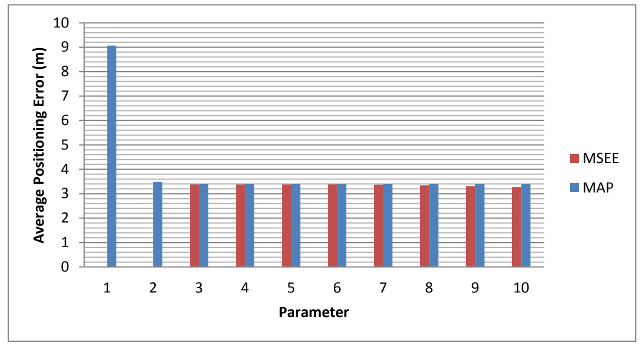


(B) Probabilistic algorithms

FIGURE 4.25. Average positioning error adding 6 DAF to the IPS

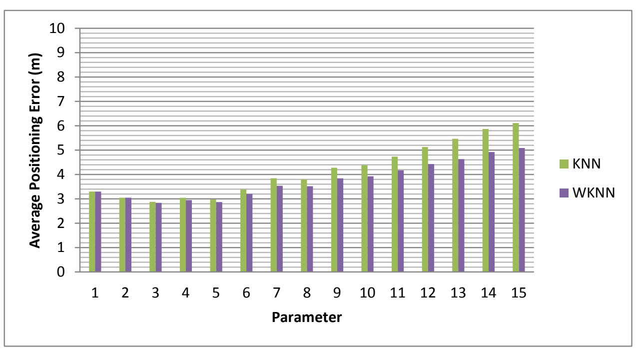


(A) Deterministic algorithms

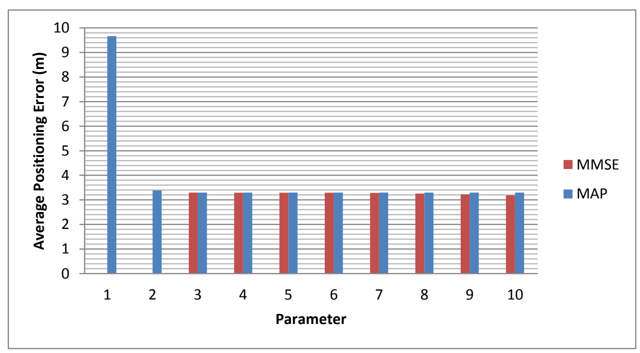


(B) Probabilistic algorithms

FIGURE 4.26. Average positioning error adding 8 DAF to the IPS



(A) Deterministic algorithms



(B) Probabilistic algorithms

FIGURE 4.27. Average positioning error considering 10 DAF in the IPS

4.11.5. Summary

A summary chart showing the best accuracy from all the previous experiments on this section is presented in Figure 4.28. The figure also presents the parameter and algorithm that yielded the fewest APE per experiment.

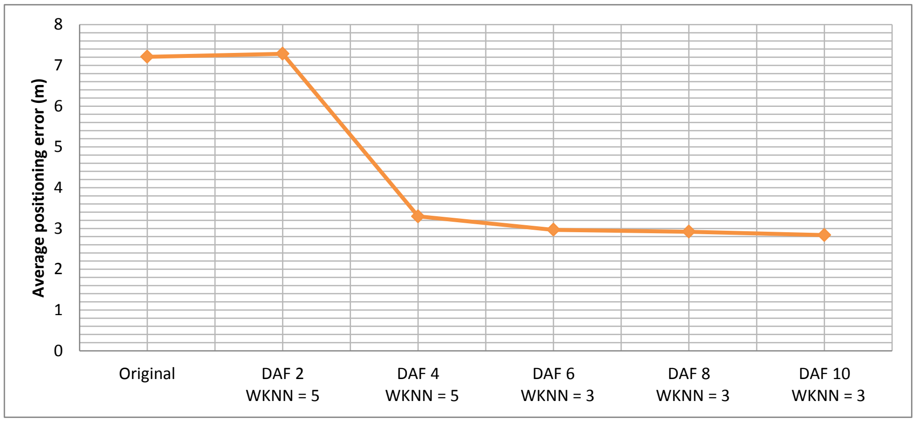


FIGURE 4.28. Best position per data set including each algorithm

4.11.6. Conclusions of the Experiment 1

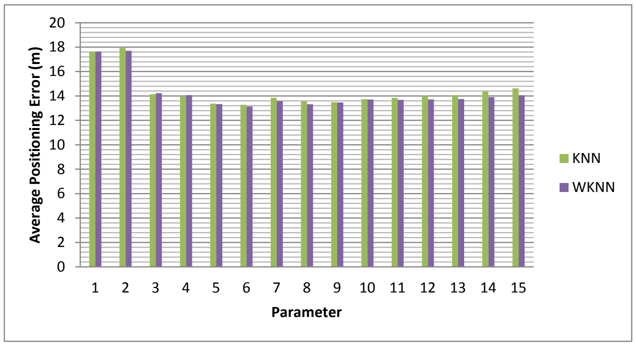
As it can be observed from the results, considering DAF improves the accuracy of an IPS at a certain point. In our case the variation was made from 2 to 10 DAF, as that was the maximum number of volunteers available to perform the experiment. The accuracy of the positioning system was improved as more DAF were added into the system to a limit of 6 DAF, adding more than 6 DAF the accuracy stays the same.

It is important to mention that the availability of the DAF depends on the movement pattern of the users. For this experiment The WKNN algorithm is the best algorithm for the given experiment using parameters 5 and 3.

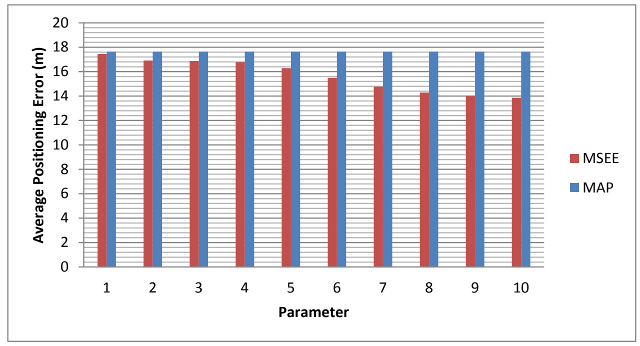
4.12. Experiment 2: Dynamic WIFI IPS Experiment at the Electrical Engineering Department Using DAF Only

The accuracy of the Experiment 1 is tested using a system consisting solely on DAF; the location of the DAF is the same as in the previous section as shown in Figure 4.21.

The figures from Figure 4.29 to Figure 4.34 illustrates the APE when DAF is varied from 2 to 10 considering a step of 2 DAF.

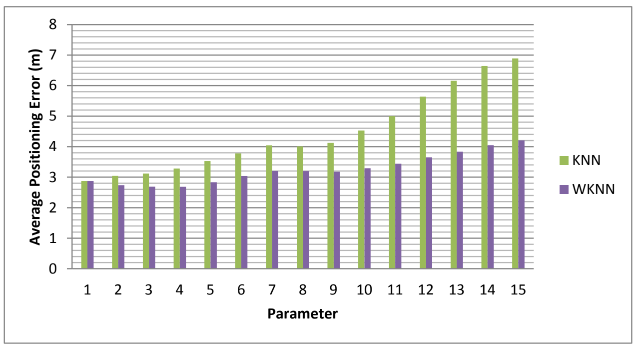


(A) Deterministic algorithms

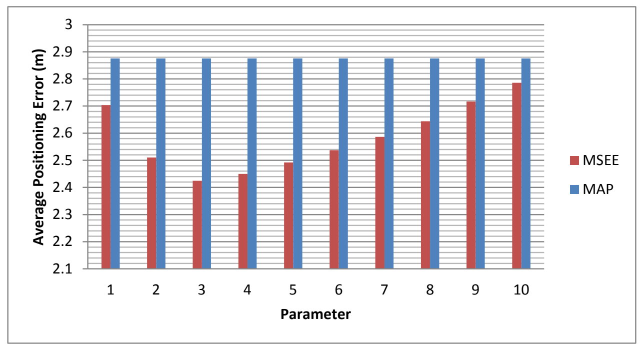


(B) Probabilistic algorithms

FIGURE 4.29. Average positioning error of 2 DAF IPS

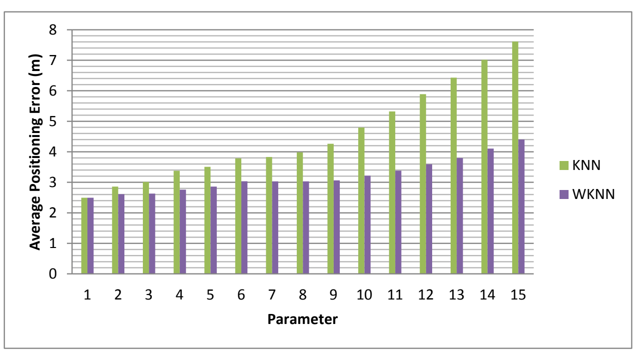


(A) Deterministic algorithms

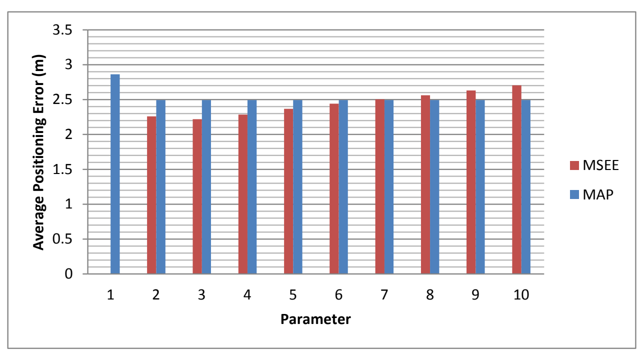


(B) Probabilistic algorithms

FIGURE 4.30. Average positioning error of 4 DAF IPS

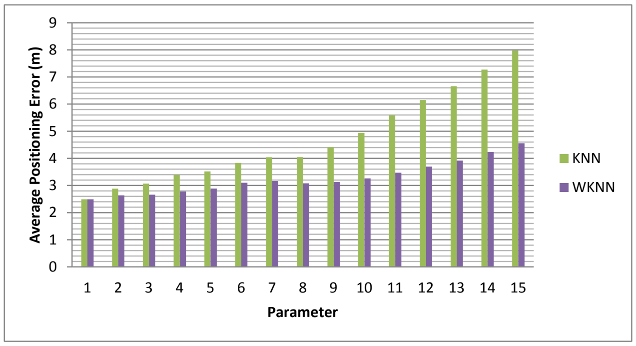


(A) Deterministic algorithms

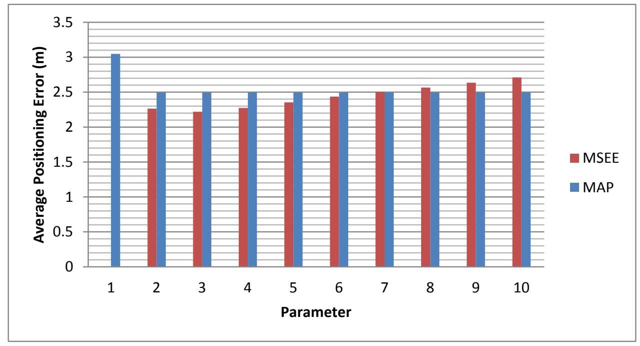


(B) Probabilistic algorithms

FIGURE 4.31. Average positioning error of 6 DAF IPS

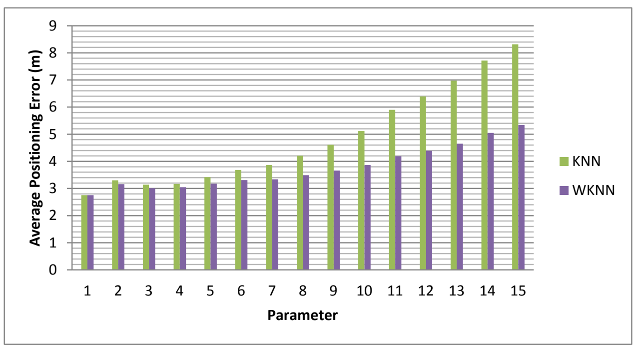


(A) Deterministic algorithms

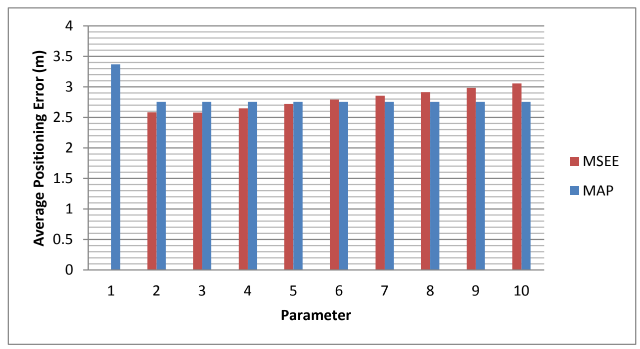


(B) Probabilistic algorithms

FIGURE 4.32. Average positioning error of 7 DAF IPS

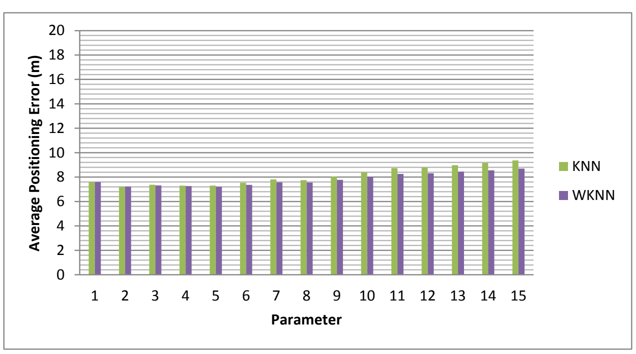


(A) Deterministic algorithms

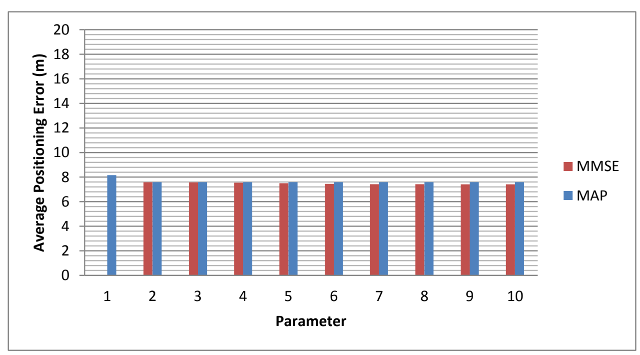


(B) Probabilistic algorithms

FIGURE 4.33. Average positioning error of 8 DAF IPS



(A) Deterministic algorithms



(B) Probabilistic algorithms

FIGURE 4.34. Average positioning error of 10 DAF IPS

4.12.1. Summary

In this subsection the best algorithm and parameter per DAF experiment that yield the least positioning error are presented in Figure 4.35. The algorithm and parameter are shown in the figure.

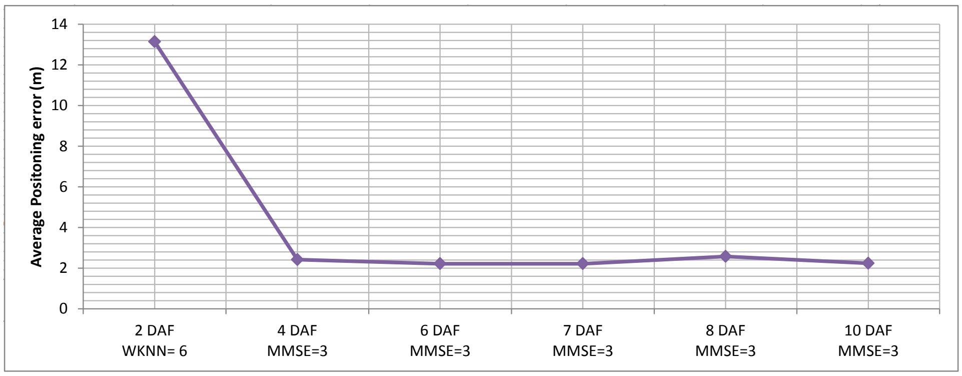


FIGURE 4.35. Best accuracy per DAF sample considering all parameters and all algorithms

4.12.2. Conclusions of the Experiment 2

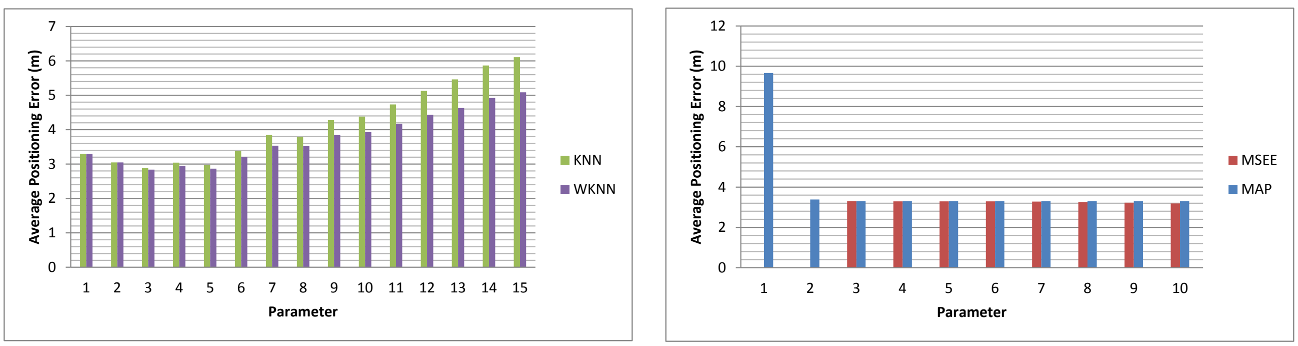
From Figure 4.35 it can be noted that the more drastic change is observed when the system changes from 2 DAF to 4 DAF. The results demonstrates that only 4 DAF are required for best accuracy. When considering more than 4 DAF, the accuracy of the system stays constant, it can also be noticed a slightly decrease in the accuracy at 8 DAF and then going back to the previous accuracy at 10 DAF.

4.13. Experiment 3: Dynamic WIFI Fingerprinting System at UNT's Electrical Engineering Department Using RandMean

In this section it is presented the results of an IPS using 5 different configurations of the *RandMean* access point selection strategy. As presented in the background section of this thesis, the *RandMean* strategy selects a random number of access points from a larger set of access points for the IPS. The 5 *RandMean* configurations have a different collection of access points. 1 access point cannot belong to more than one configuration.

In the discovery park building, using the location in Figure 4.20 46 access points were available for positioning. For this experiment from the 46 available access points 5 sets were constructed; 4 containing a set of 10 access points and 1 set containing a set of 6 access points. For the 5 sets the 10 dynamic access points were always available.

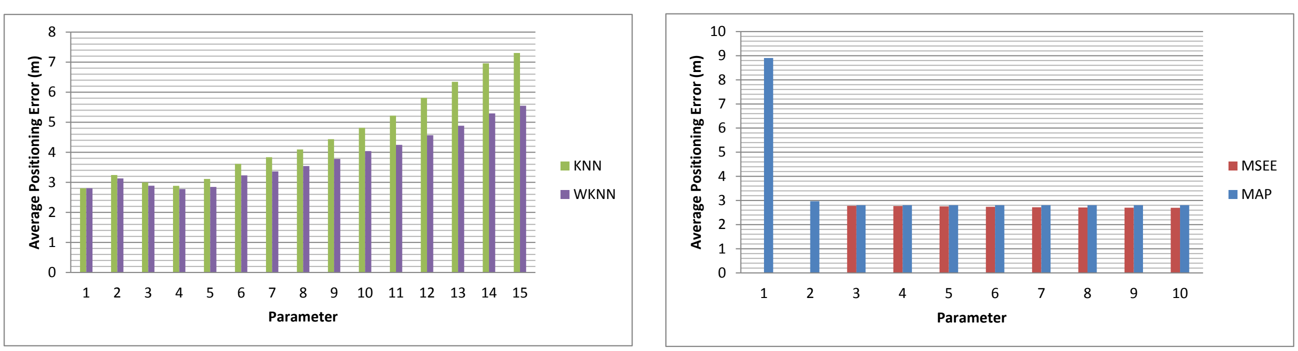
In this section results of each of the 5 configurations are presented for the deterministic and probabilistic approaches. The results are presented from Figure 4.36 to Figure 4.40.



(A) Deterministic algorithms

(B) Probabilistic algorithms

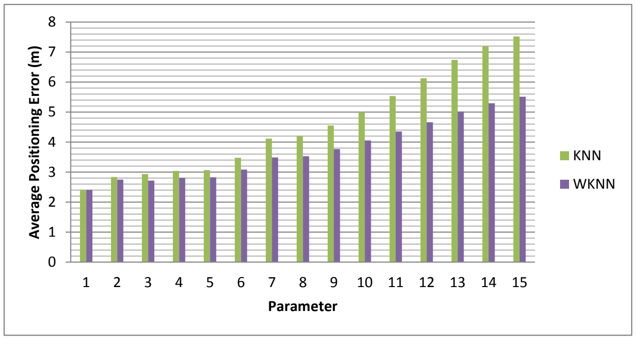
FIGURE 4.36. Average positioning error for the RandMean experiment 1



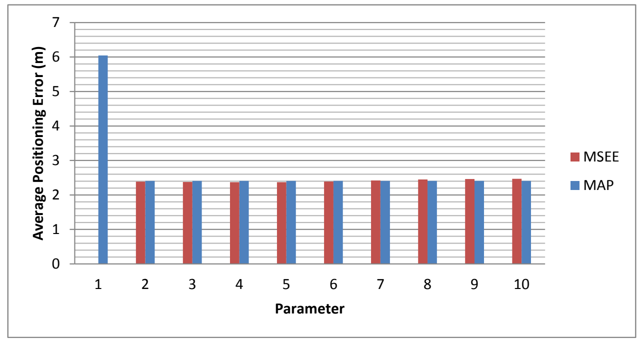
(A) Deterministic algorithms

(B) Probabilistic algorithms

FIGURE 4.37. Average positioning error for the RandMean experiment 2

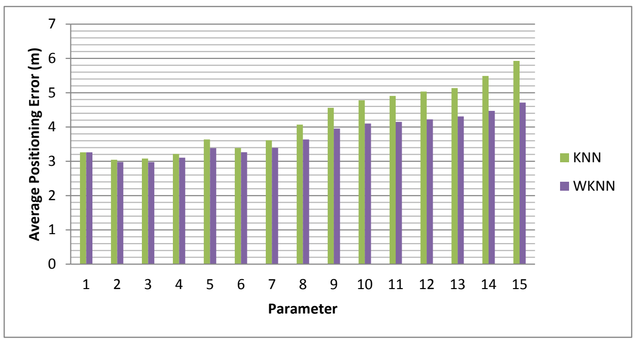


(A) Deterministic algorithms

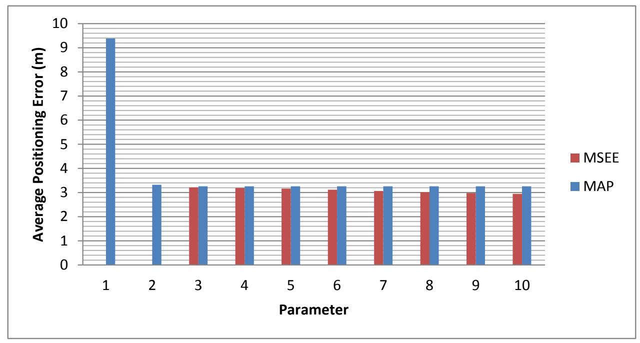


(B) Probabilistic algorithms

FIGURE 4.38. Average positioning error for the RandMean experiment 3

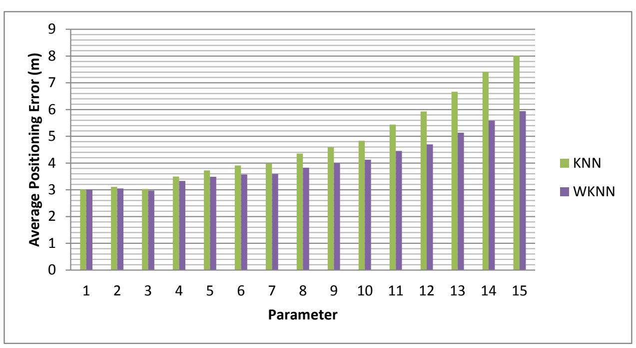


(A) Deterministic algorithms

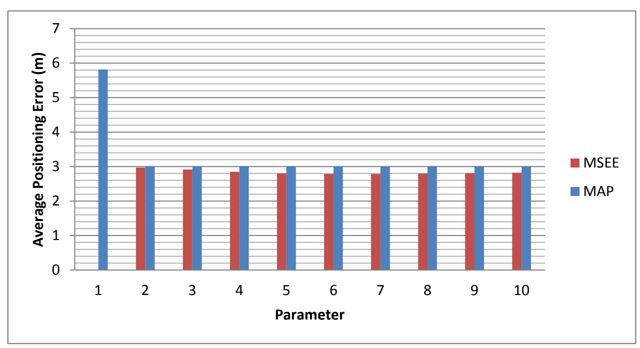


(B) Probabilistic algorithms

FIGURE 4.39. Average positioning error for the RandMean experiment 4



(A) Deterministic algorithms



(B) Probabilistic algorithms

FIGURE 4.40. Average positioning error for the RandMean experiment 5

4.13.1. Similarity of the Results for Experiment 3

Figure 4.41 shows the sorted results from the deterministic and probabilistic algorithms of the 5 sets according to the accuracy obtained when considering all the parameters. The results show that the accuracy vary from 2 meters to 10 meters according to the parameters selected. The figure also shows that similar accuracy results are obtained for the different sets of the random access points selection experiment. The best accuracy when using the RandMean approach is of 2.2 meters.

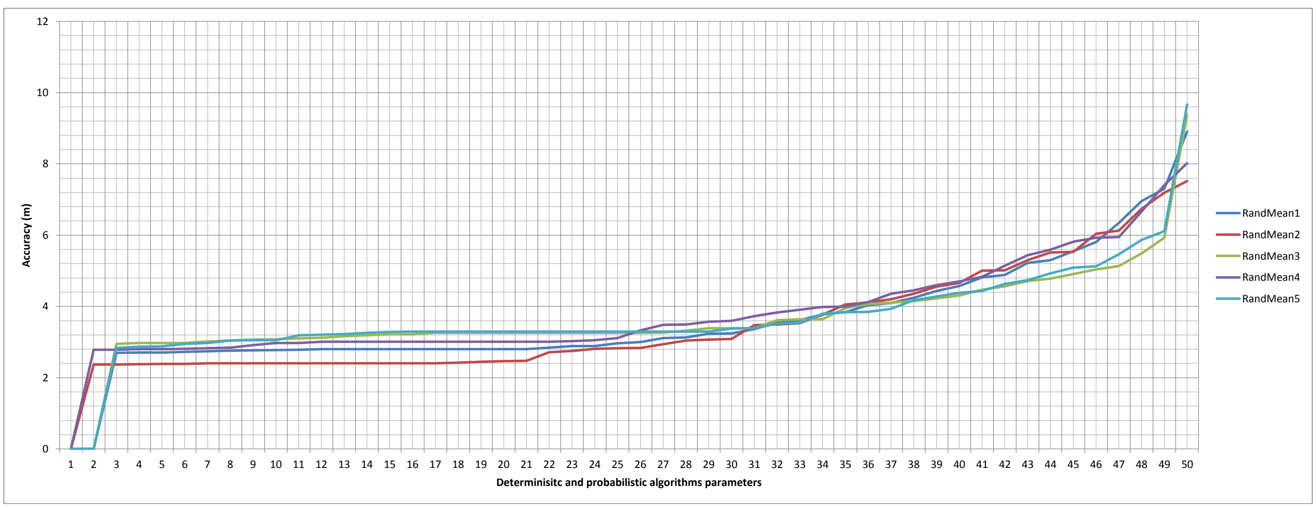


FIGURE 4.41. Sorted results according to the obtained accuracy from all experiments

4.13.2. Summary

In Figure 4.42 it can be observed the best accuracy per RandMean experiment considering all parameters and all algorithms. The results shows that the positioning error variate between 2 and 3 meters.

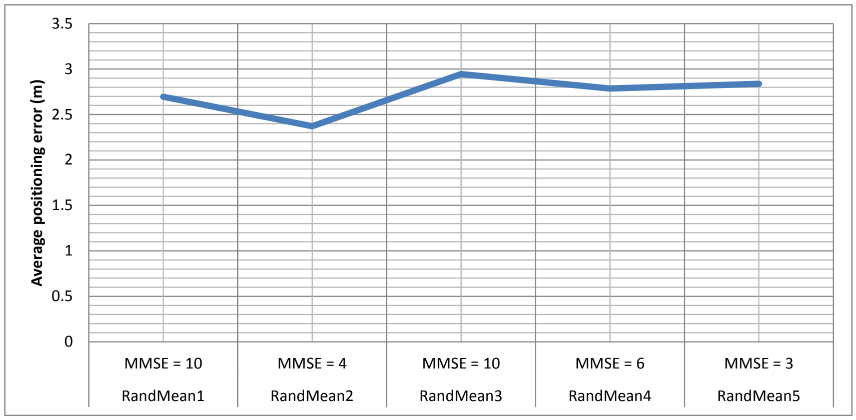


FIGURE 4.42. Best accuracy per RandMean experiment

4.14. Conclusion of the Results from all Experiments at the Electrical Engineering Department

(1) Original IPS

as it can be observed from Figure 4.28 the original system produces the best accuracy for an APE of 7 meters.

(2) Original IPS with 10 DAF using MaxMean access point selection strategy.

In Figure 4.28 is shown the APE for the original IPS when also considered 10 DAF, the best accuracy when considering both the original system and the DAF is obtained for an APE of 3 meters.

(3) IPS consisting only of DAF

In Figure 4.65a it can be observed the best accuracy results considering only DAF. the best accuracy is obtained for an APE of 2.6 meters approximately.

(4) Original IPS with 10 DAF using RandMean access point selection strategy.

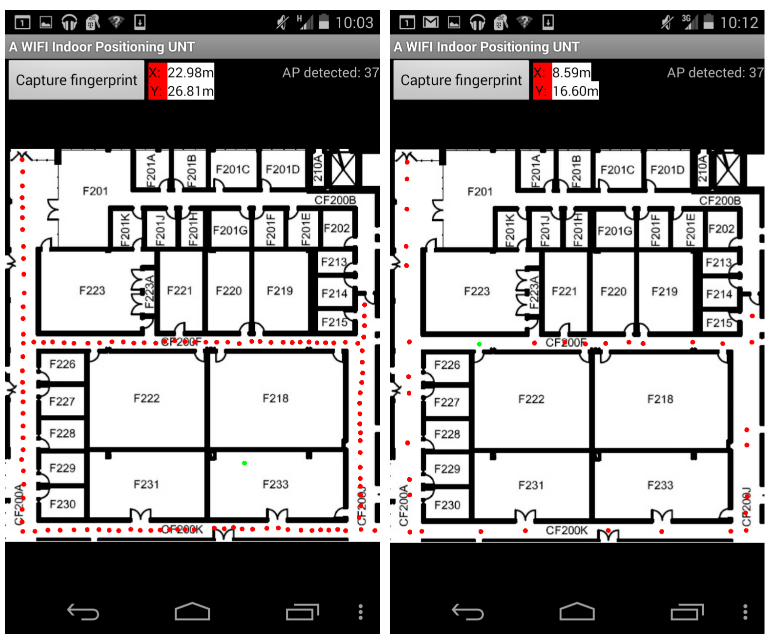
In Figure 4.42 is shown the APE for the original IPS when considering 10 DAF. The best accuracy when considering both the original system and the DAF is obtained for an APE of 2.4 meters.

As it can be noted from the previous results, a centimeter-difference APE is obtained for the IPS systems considering the MaxMean and RandMean access point selection criterion and the DAF only method; the difference ranges from 2.4 meters to 3 meters, all those IPS outperform the original system APE of 7 meters.

The results show that when considering a DAF only IPS system, the system outperforms the original system, when considering the same number of access points for the 2 systems. This effect is caused as the DAF are characterize more efficiently different fingerprints than an IPS without access points.

4.15. Experiment 4: Dynamic WIFI Fingerprinting System at UNT's CSE Department Using MaxMean

In this section a dynamic WIFI fingerprinting IPS deployed at the CSE department of UNT is presented. The selected area for the indoor positioning system is showing in Figure 4.43



(A) Offline fingerprints (B) Testing fingerprints

FIGURE 4.43. Offline and testing fingerprints at the CSE department

The dimensions of the selected area are 35.9664 meters wide by 33.528 high. The dimension are considered in the android application to calculate the average positioning error. Following the same approach as in experiment 1, only 10 access points are selected from all the fixed WIFI access points available using the *MaxMean* access point selection strategy. For this experiment 10 dynamic access points from volunteers were available for positioning. As in previous experiments, 4 algorithms are used to evaluate the accuracy of the system: 2 deterministic (KNN and WKNN) and 2 probabilistic (MAP and MMSE).

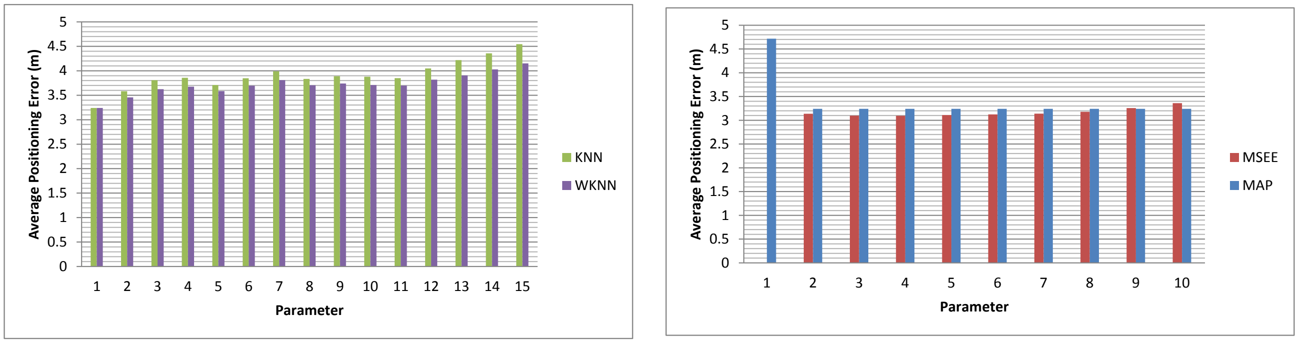
4.15.1. Location of the DAF

In Figure 4.44 are shown the location of the 10 DAF at the CSE department.

Figure 4.44. Location of DAF at the CSE department

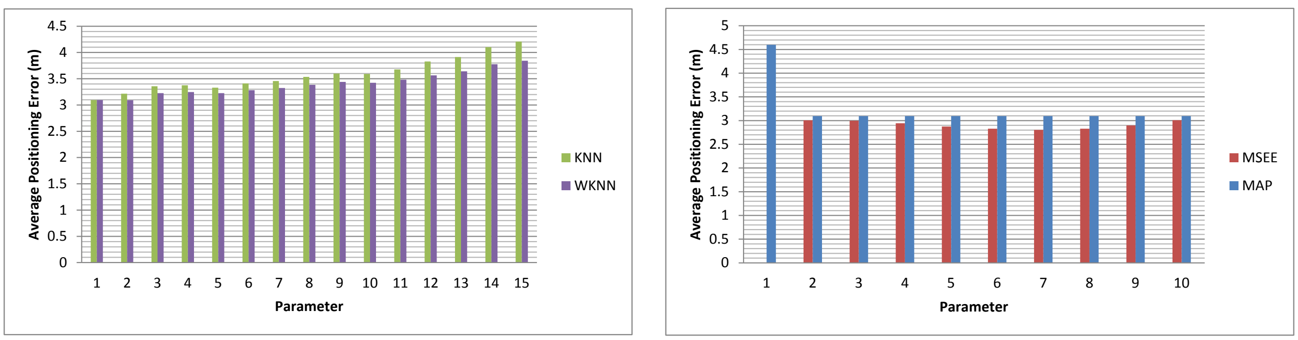
4.15.2. Results of the IPS from the Original (Non-Dynamic) System to Adding 10 DAF (Dynamic)

In this subsection the results from the non-dynamic system (original) to the addition of 10 DAF into the system are presented in Figure 4.45 to Figure 4.55.



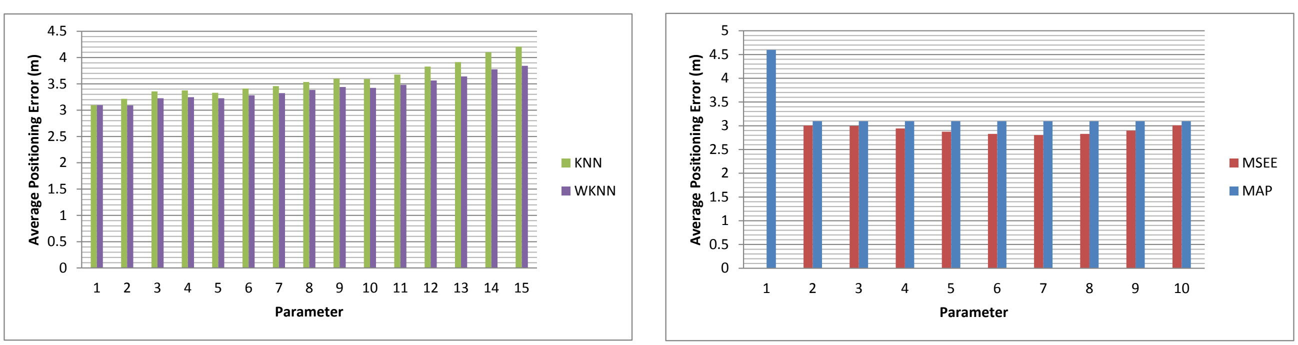
(A) Deterministic algorithms (B) Probabilistic algorithms

Figure 4.45. Average positioning error of the non-dynamic IPS



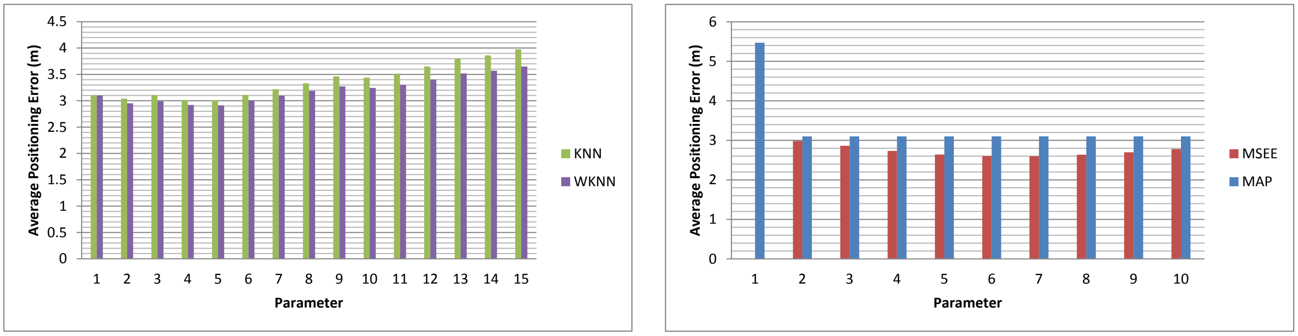
(A) Deterministic algorithms (B) Probabilistic algorithms

Figure 4.46. Average positioning error considering 1 DAF in the IPS



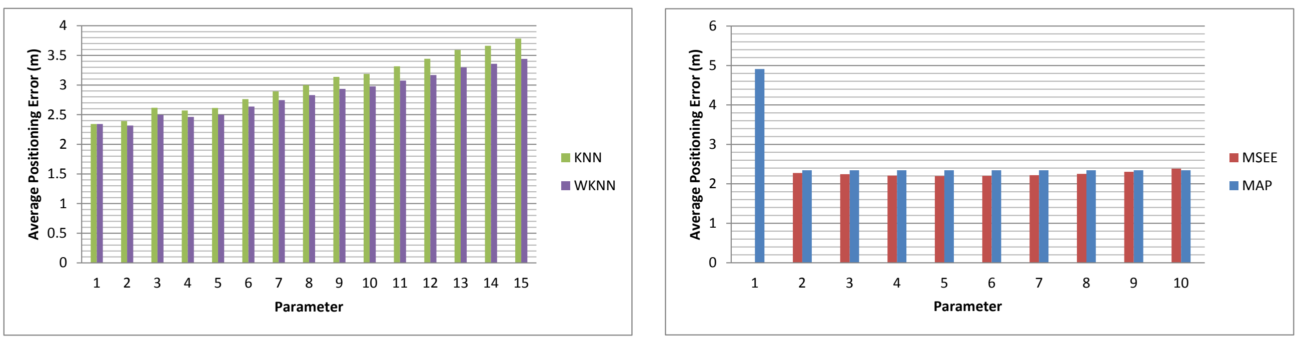
(A) Deterministic algorithms (B) Probabilistic algorithms

Figure 4.47. Average positioning error considering 2 DAF in the IPS



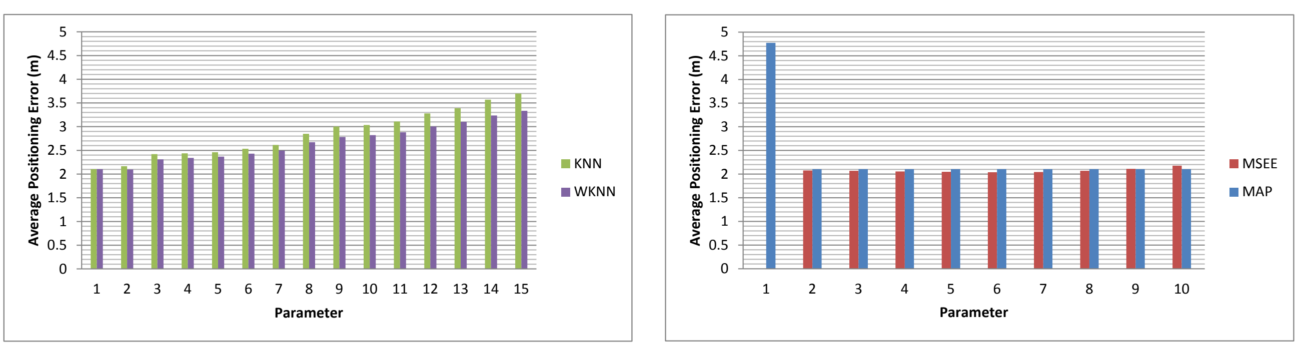
(A) Deterministic algorithms (B) Probabilistic algorithms

FIGURE 4.48. Average positioning error considering 3 DAF in the IPS



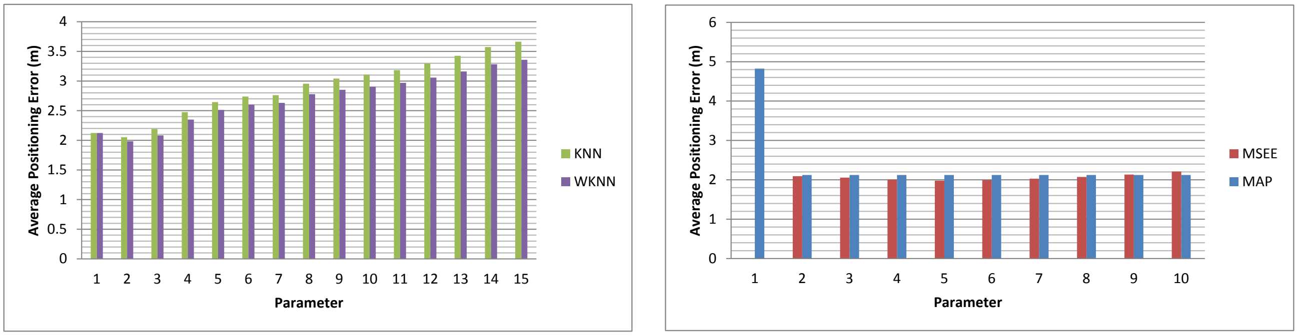
(A) Deterministic algorithms (B) Probabilistic algorithms

FIGURE 4.49. Average positioning error considering 4 DAF in the IPS



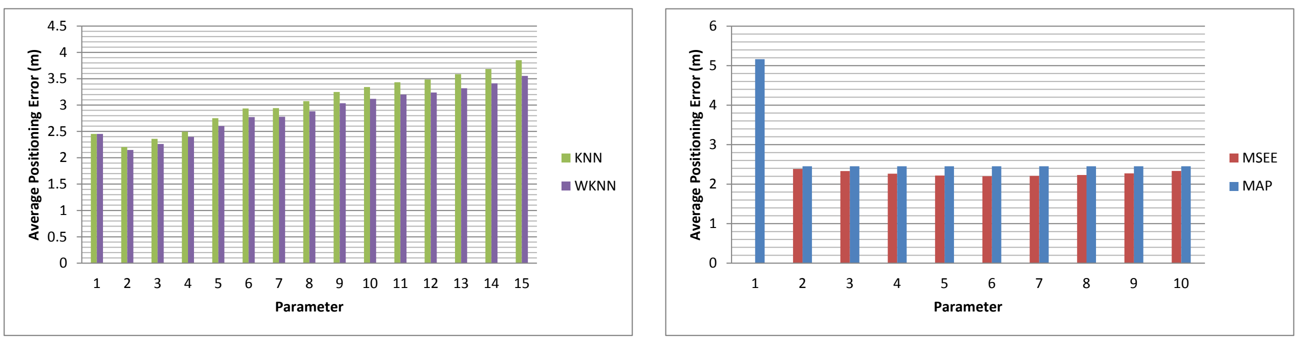
(A) Deterministic algorithms (B) Probabilistic algorithms

FIGURE 4.50. Average positioning error considering 5 DAF in the IPS



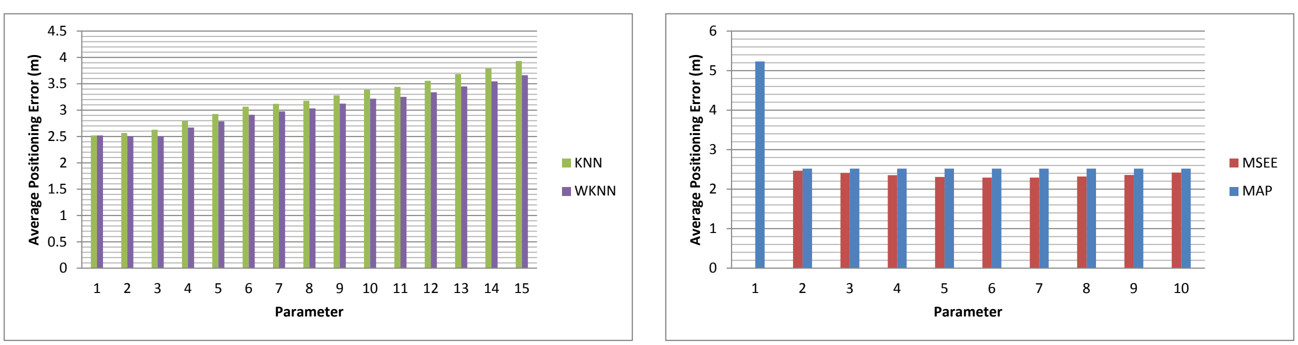
(A) Deterministic algorithms (B) Probabilistic algorithms

FIGURE 4.51. Average positioning error considering 6 DAF in the IPS



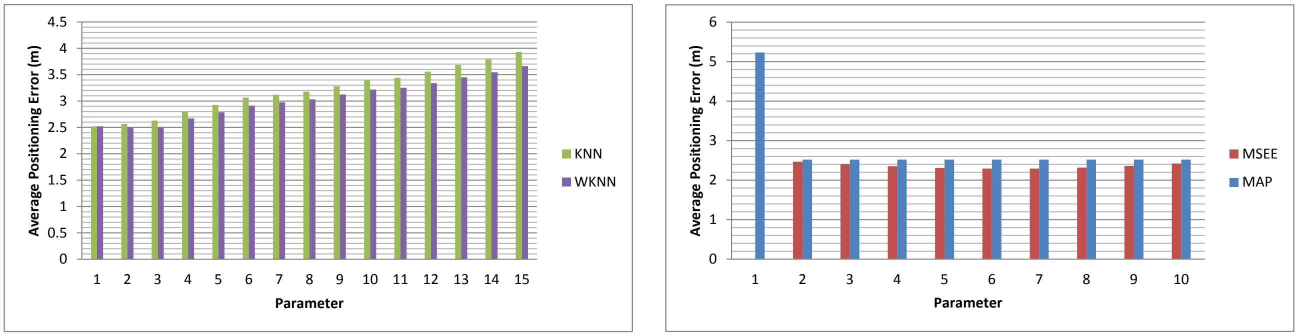
(A) Deterministic algorithms (B) Probabilistic algorithms

FIGURE 4.52. Average positioning error considering 7 DAF in the IPS



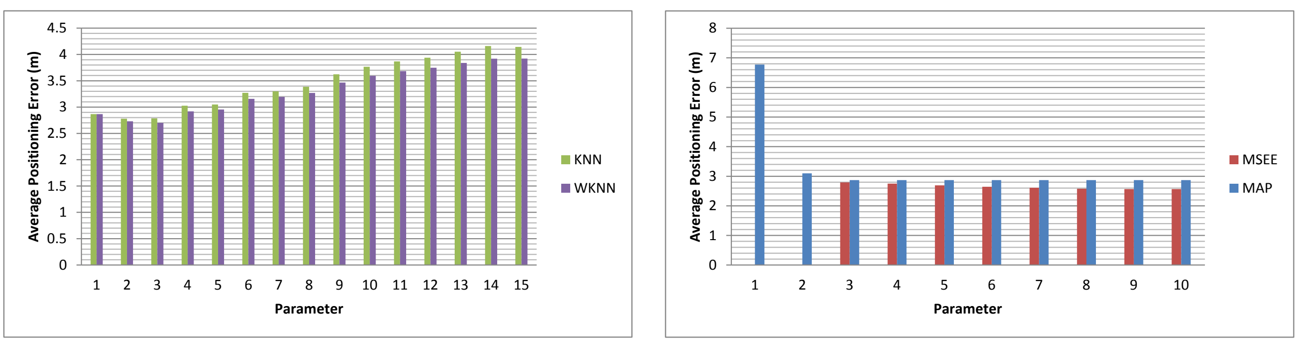
(A) Deterministic algorithms (B) Probabilistic algorithms

FIGURE 4.53. Average positioning error considering 8 DAF in the IPS



(A) Deterministic algorithms (B) Probabilistic algorithms

FIGURE 4.54. Average positioning error considering 9 DAF in the IPS



(A) Deterministic algorithms (B) Probabilistic algorithms

FIGURE 4.55. Average positioning error considering 10 DAF in the IPS

4.15.3. Summary

The best accuracy per experiment,from the original IPS to adding 10 DAF, are presented in Figure 4.56. The best accuracy for each trial is shown; also the parameter that yielded the best accuracy is presented.

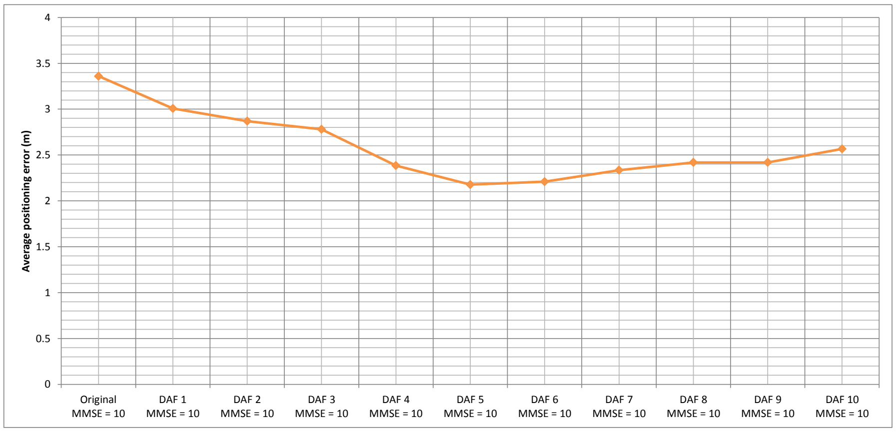


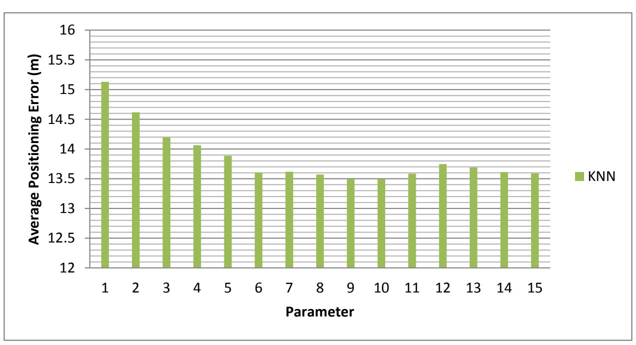
FIGURE 4.56. Best estimate of location in all experiments and the corresponding parameters

4.15.4. Conclusions of the Experiment 4

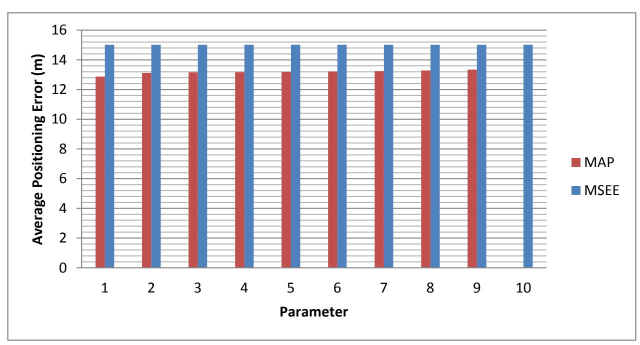
As it can be observed in Figure 4.56, the best accuracy for the IPS is obtained when 5 DAF are considered; considering more than 5 access points decrease the accuracy of the system. The MMSE at parameter 10 performed better in all the trials when compared with other algorithms as expected.

4.16. Experiment 5: Dynamic WIFI Fingerprinting Consisting of DAF Only

In this section, it is presented the results of the dynamic IPS consisting only of dynamic access points and fingerprints. The results are shown from Figure 4.57 to Figure 4.65. The availability of the DAF is varied from 1 to 10 in a step of 1 DAF.

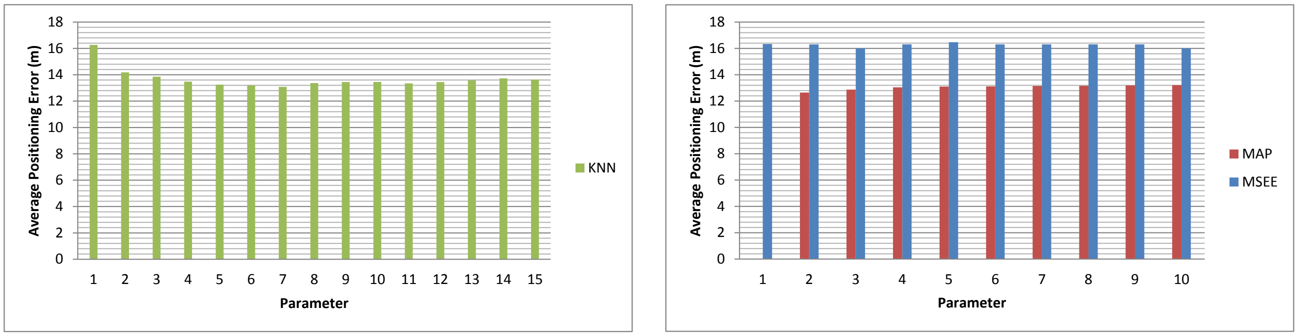


(A) Deterministic algorithms



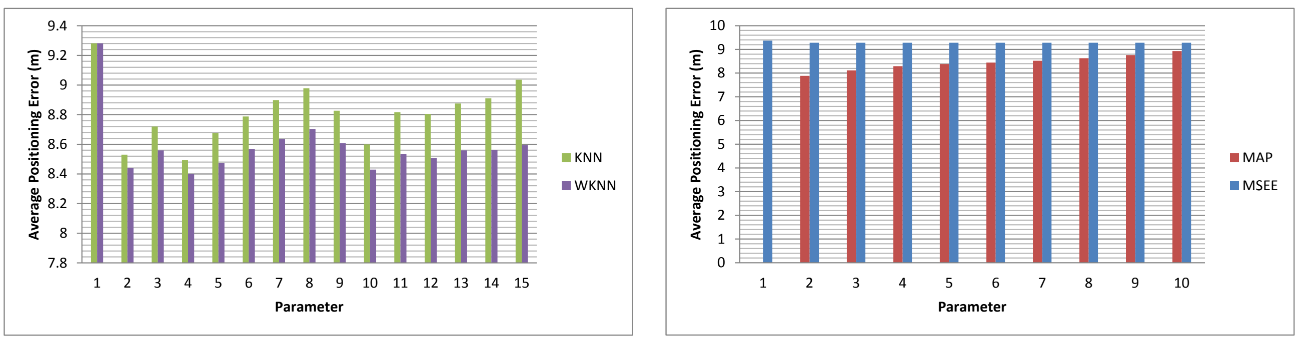
(B) Probabilistic algorithms

FIGURE 4.57. Average positioning error of the dynamic IPS considering 2 DAF



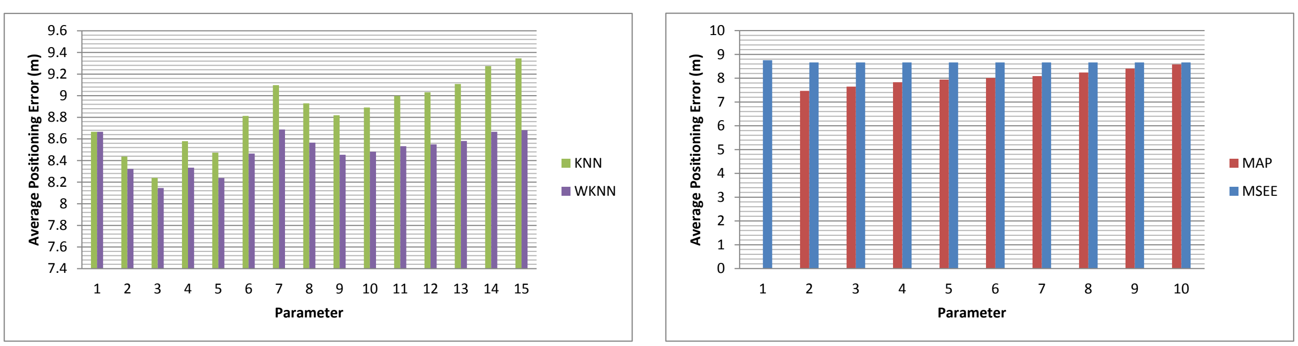
(A) Deterministic algorithms (B) Probabilistic algorithms

FIGURE 4.58. Average positioning error of the dynamic IPS considering 3 DAF



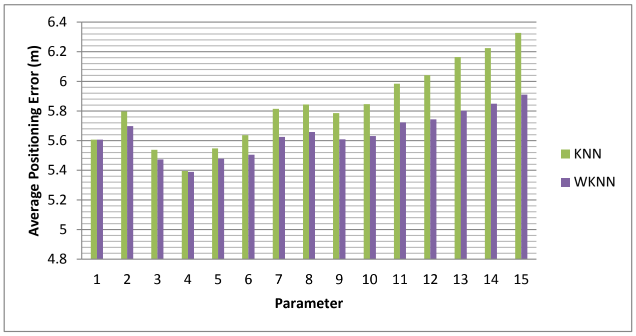
(A) Deterministic algorithms (B) Probabilistic algorithms

FIGURE 4.59. Average positioning error of the dynamic IPS considering 4 DAF

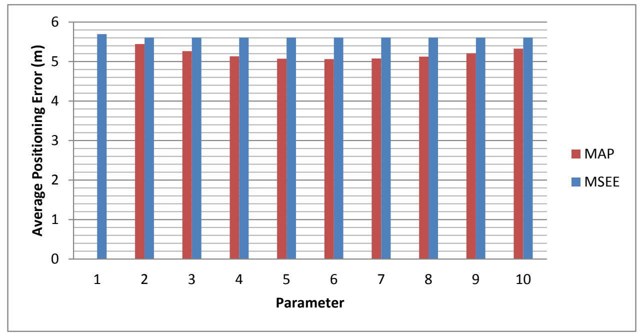


(A) Deterministic algorithms (B) Probabilistic algorithms

FIGURE 4.60. Average positioning error of the dynamic IPS considering 5 DAF

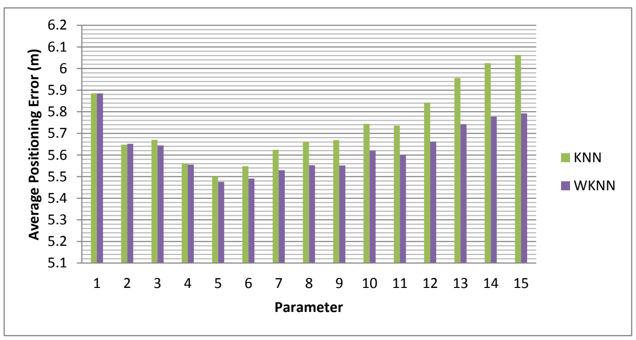


(A) Deterministic algorithms

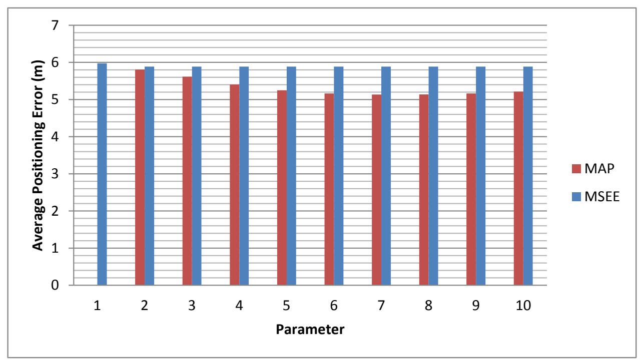


(B) Probabilistic algorithms

FIGURE 4.61. Average positioning error of the dynamic IPS considering 6 DAF

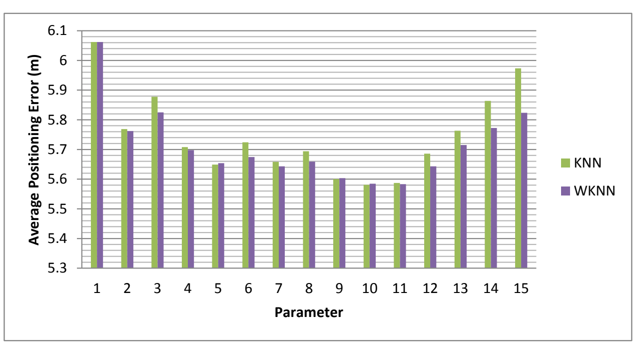


(A) Deterministic algorithms

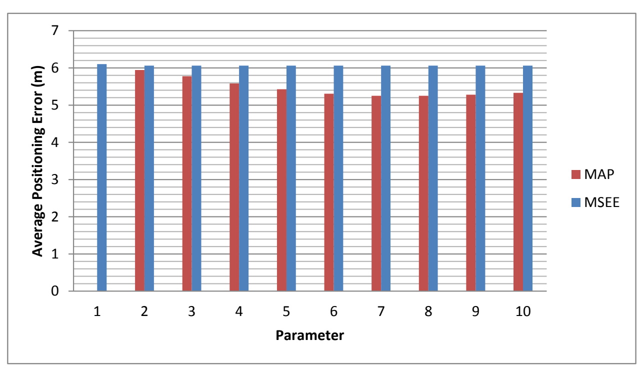


(B) Probabilistic algorithms

FIGURE 4.62. Average positioning error of the dynamic IPS considering 7 DAF

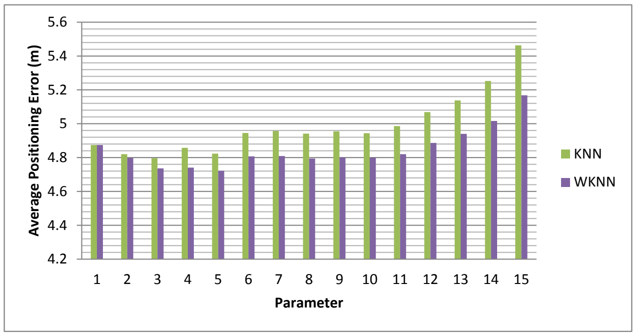


(A) Deterministic algorithms

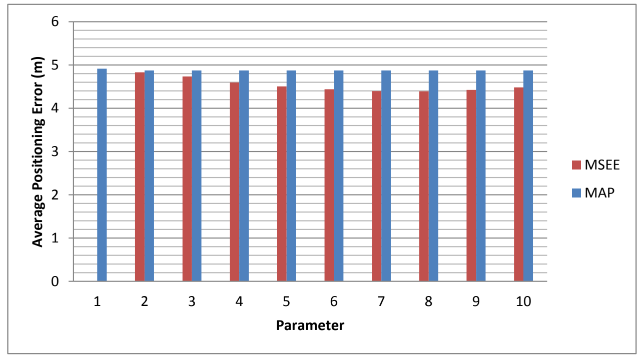


(B) Probabilistic algorithms

FIGURE 4.63. Average positioning error of the dynamic IPS considering 8 DAF

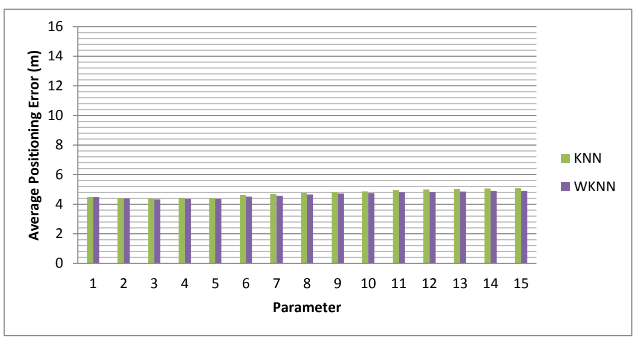


(A) Deterministic algorithms

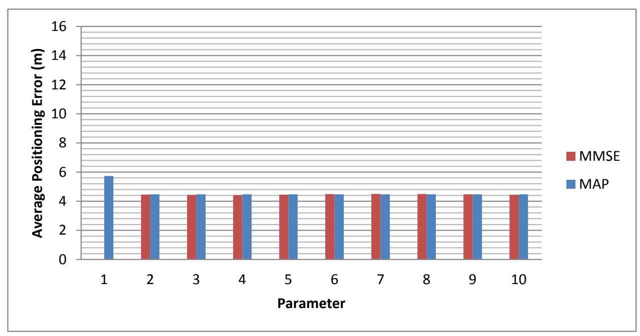


(B) Probabilistic algorithms

Figure 4.64. Average positioning error of the dynamic IPS considering 9 DAF



(A) Deterministic algorithms



(B) Probabilistic algorithms

Figure 4.65. Average positioning error of the dynamic IPS considering 10 DAF

4.16.1. Summary

In Figure 4.66 it can be observed the best results for all the experiments considering all parameters and all algorithms, the results show only the parameters that yielded the best accuracy per experiment; the accuracy is increased as more DAFs were considered into the system.

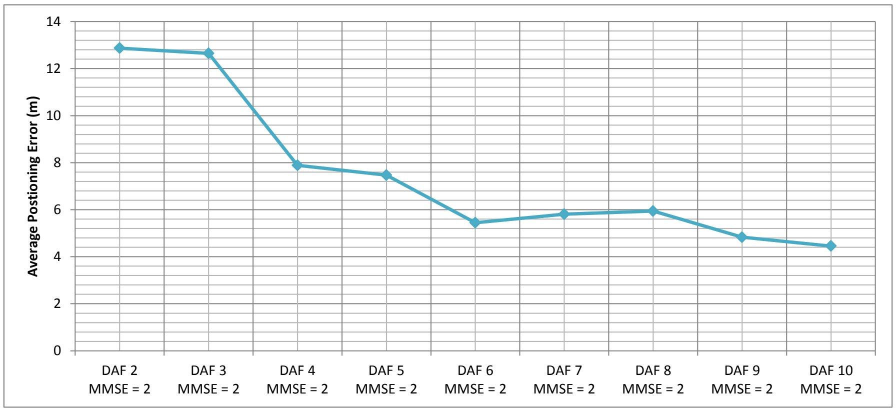


FIGURE 4.66. Best accuracy per DAF sample

4.16.2. Conclusions of the Experiment 5

The results presented in Figure 4.66 show that when considering a system solely of DAF the accuracy is better as the number of DAF increase. In the experiment performed for this thesis the DAF was varied from 1 to 10, the algorithm that performed better was the MMSE with parameter 2.

4.17. Conclusion of the Results from all Experiments at the CSE Department

The original IPS at the CSE department yields an APE of 3.4 meters. When adding DAF into the original system the accuracy of the system is increased, with a maximum of 2.4 meters for 5 DAF. When considering more than 5 DAF the accuracy of the system decreases.

When considering a DAF only IPS the APE varies from 13 meters to 4.4 meters for 2 and 10 DAF respectively. The results show that when adding more DAF into the system for the CSE department experiment the APE decreases.

CHAPTER 5

SLAM IMPLEMENTATION ON A ROBOT INDOORS

5.1. Introduction

In the robotics field, SLAM refers to the process that a robot uses to incrementally generate a map of an unknown environment and at the same time to calculate the position of the robot. This process is crucial for the success of the navigation of the Robot while staying at indoor environments.

5.2. Formulation of the Problem

If a robot is moving through the environment, at time k, the following concepts gives an approximate solution to the SLAM problem: [8] .

The $\mathbf{x}_k$ state vector describes the location of the robot, the record of robot locations can be expressed as: $\mathrm{X}_{0:k} = \{\mathrm{x}_0,\ \mathrm{x}_1,\ \cdots,\ \mathrm{x}_k\} = \{\mathrm{X}_{0:k-1},\ \mathrm{x}_k\}$:

The vector $\mathbf{u}_k$ is a control vector, it influences the behavior of the robot when going from position $\mathbf{x}_{k-1}$ to position $\mathbf{x}_k$, the record of control inputs can be expressed as: $\mathrm{U}_{0:k} = \{\mathrm{u}_1,\ \mathrm{u}_2,\ \cdots,\ \mathrm{u}_k\} = \{\mathrm{U}_{0:k-1},\ \mathrm{u}_k\}$

The vector $\mathbf{m}_i$ describes the location of the ith landmark. The set of all landmark observaions are expressed as: $\mathrm{m} = \{\mathrm{m}_1,\ \mathrm{m}_2,\ \cdots,\ \mathrm{m}_n\}$

The vector $\mathbf{z}_{ik}$ describes the observation of a landmark from the robot. The set of landmark observations are expressed as:$\mathrm{Z}_{0:k} = \{\mathrm{z}_1,\ \mathrm{z}_2,\ \cdots,\ \mathrm{z}_k\} = \{\mathrm{Z}_{0:k-1},\ \mathrm{z}_k\}$

The following probability distribution describes the joint posterior probability of the landmarks and the location of the robot, which needs to be computer for all times k:

$$P(\mathrm{x}_k,\ \mathrm{m}|\mathrm{z}_{0:k},\ \mathrm{U}_{0:k},\ \mathrm{x}_0) \tag{30}$$

Using this equation, a recursive solution to the slam problem is obtained. In particular at time $k-1$ the equation can be expressed as

$$P(\mathrm{x}_{k-1},\ \mathrm{m}|\mathrm{Z}_{0:k-1},\ \mathrm{U}_{0:k-1},\ \mathrm{x}_0) \tag{31}$$

which can be solved using Bayes theorem.

5.2.1. Models

In the following section, the observation and the motion models of the SLAM problem are presented.

- Observation Model: this model calculates the probability of observation z_k when the position and the landmark location are known to the Robot.

$$P(\mathrm{z}_k|\mathrm{x}_k,\ \mathrm{m}) \tag{32}$$

- Motion Model: this model obtains the probability of position x_k given the previous position of the robot, this equation can also be seen as a markov process.

$$P(\mathrm{x}_k|\mathrm{x}_{k-1},\ \mathrm{u}_k) \tag{33}$$

5.3. SLAM Algorithms

In this section 2 recursive algorithms are presented in the form of probabilistic estimations to calculate the position of the robot according to updates to the system of type time and measurement.

If, as presented in the previous section, the vector x_k is considered for the robot state, m for the map, k for the time and the observations $\mathrm{z}_{0:k}$ and control inputs $\mathrm{U}_{0:k}$ are also considered to estimate the probabilities.

The Figure 5.1 [8] presents the SLAM technique using landmarks.

- Time-Update

$$P(\mathrm{x}_k,\ \mathrm{m}|\mathrm{Z}_{0:k-1},\ \mathrm{U}_{0:k},\ \mathrm{x}_0) = \int P(x_k|x_{k-1},\ \mathrm{u}_k) \times P(\mathrm{x}_{k-1},\ \mathrm{m}|\mathrm{Z}_{0:k-1},\ \mathrm{U}_{0:k-1},\ \mathrm{x}_0)\mathrm{dx}_{k-1} \tag{34}$$

- Measurement-Update

$$P(\mathrm{x}_k,\ \mathrm{m}|\mathrm{Z}_{0:k},\ \mathrm{U}_{0:k},\ \mathrm{x}_0) = \frac{P(\mathrm{z}_k|\mathrm{x}_k, \mathrm{m})P(\mathrm{x}_k, \mathrm{m}|\mathrm{Z}_{0:k-1}, \mathrm{U}_{0:k}, \mathrm{x}_0)}{P(\mathrm{z}_k|\mathrm{Z}_{0:k-1}, \mathrm{U}_{0:k})} \tag{35}$$

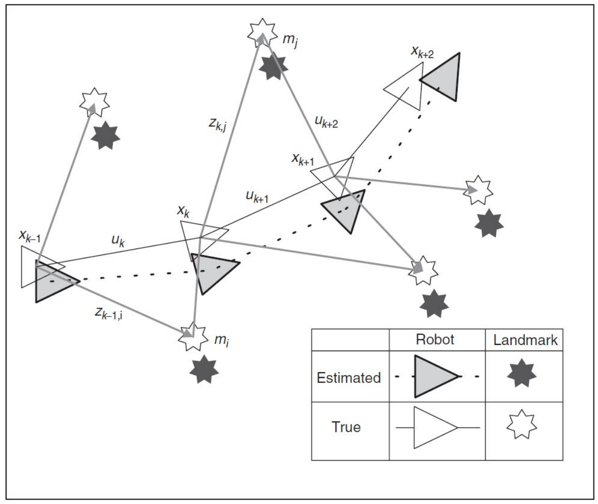


FIGURE 5.1. SLAM using landmarks

5.4. RGBD SLAM

RGBD SLAM (Red, Green, Blue & Deep Simultaneous Localization and Mapping) is an open source approach to solve the SLAM problem [9], based on cameras, the approach uses the Kinect Sensor to solve the slam problem and it also creates a 3D model representation of the environment.

In order to solve the SLAM problem the project is divided into frontend and backend. the frontend is in charge of obtaining relations between tha image frames obtained from the hand held device and the backend optimizes the pose of those observations in a pose graph.

5.4.1. Frontend

After obtaining a set of frames using the Microsoft Kinect sensor visual features are compared and matched, the OpenCV software, is used to perform the image processing.

A feature descriptor extract features from an image to perform matching between frames.

The following feature descriptors are implemented in the Frontend:

SIFT: The Scale Invariant Feature Transform (SIFT) has the properties that the features extracted are invariant to image scale and rotation. The transform provides robust matching in environments that include noise, affine distortion, change in 3D view point and changes in the illumination [19].

SURF: The Speeded-Up Robust Features (SURF) is fast feature extractor were the approach reduces the number of operations being performed. The authors implement a laplacian indexing strategy which improves the matching between images without affecting the performance of the method

ORB: ORB is a feature matching descriptor which outperforms in terms of speed to the previously mentioned descriptors (SIFT and SURF). This descriptor can resist noise and rotation invariant.

SIFTGPU: This descriptor is a parallelized version of the SIFT feature descriptor, it is implemented in the GLSL and CUDA GPU-oriented programming languages.

5.4.2. Backend

The SLAM backend constructs a globally consistency trajectory for the Robot as a pose graph. The pose graph is optimized using the g^2 framework. The framework performs a minimization of an error function that can be represented as a graph.

The Figure 5.2 [9] presents the RGBD SLAM process. Visual features are first captured and then it extract depth information that corresponds to 3D points. Pair of frames are register and a pose graph is built with this information. Finally a map is created.

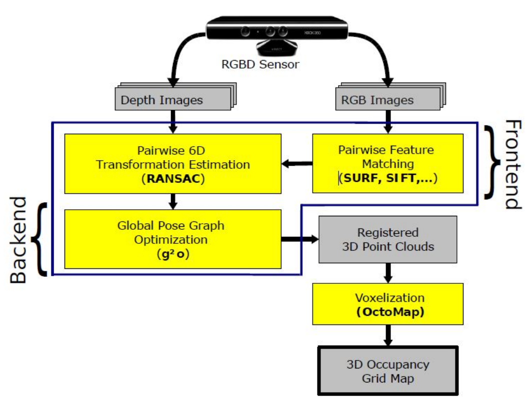


FIGURE 5.2. Schematic overview of RGBD SLAM

In this chapter are reported experiments conducted using SLAM with a Robot inside the hallways of the electrical engineering department.

5.5. Hardware Used

For the experiment conducted the robotic platform called *eddie* was used, as it can be observed in Figure 5.3 [1].

FIGURE 5.3. Eddie robot platform

5.6. Robot Operating System

ROS is defined [1] as a *"Flexible framework for writing robot software. It is a collection of tools, libraries, and conventions that aim to simplify the task of creating complex and robust robot behavior across a wide variety of robotic platforms"*. The software can also work in a multi-computer environment.

ROS is a widely used open-source framework among institutions and individuals to create Robotics applications. The main reason behind the extensive use of this software is because of its intention to be a general-purpose framework aiming to be compatible with several robotics platforms, which facilitates the porting of the code to other systems.

Those are basic concepts of the framework:

Node: Process that performs calculations The Figure 5.4 presents the interaction between nodes.

master: Provides communication among nodes

messages: The way the nodes communicate with each other

Topics: Provides a communication channel among nodes with a subscriber and publisher configuration

Services: Enables request and reply requests among nodes

Bags: Bags are used to store the data being shared.

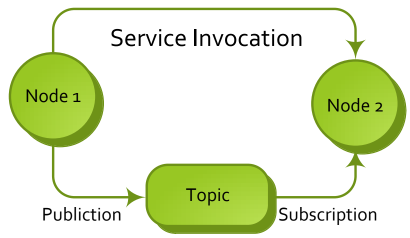


FIGURE 5.4. Node interaction

5.7. Eddiebot

The Eddie Robot was originally developed to work under the Microsoft Robotics Development Studio.

The *eddiebot* is an open-source ROS package that ports the basic drivers from the Eddie robot to provide functionality of the robot within the Robot Operating System, facilitating the interface between the software (ROS) and the hardware (Robot)

5.8. Conclusions and Results

The RGBD-SLAM package was implemented on a Eddie Robot for experimentation at the electrical engineering department hallways at the University of North Texas.

The Figure 5.5 and Figure 5.6 present the result of the experiments performed in the Discovery Park building.

A SLAM technique based on the Microsoft Kinect was successfully implemented in the Eddie robot for positioning an navigation at the electrical engineering department at UNT.

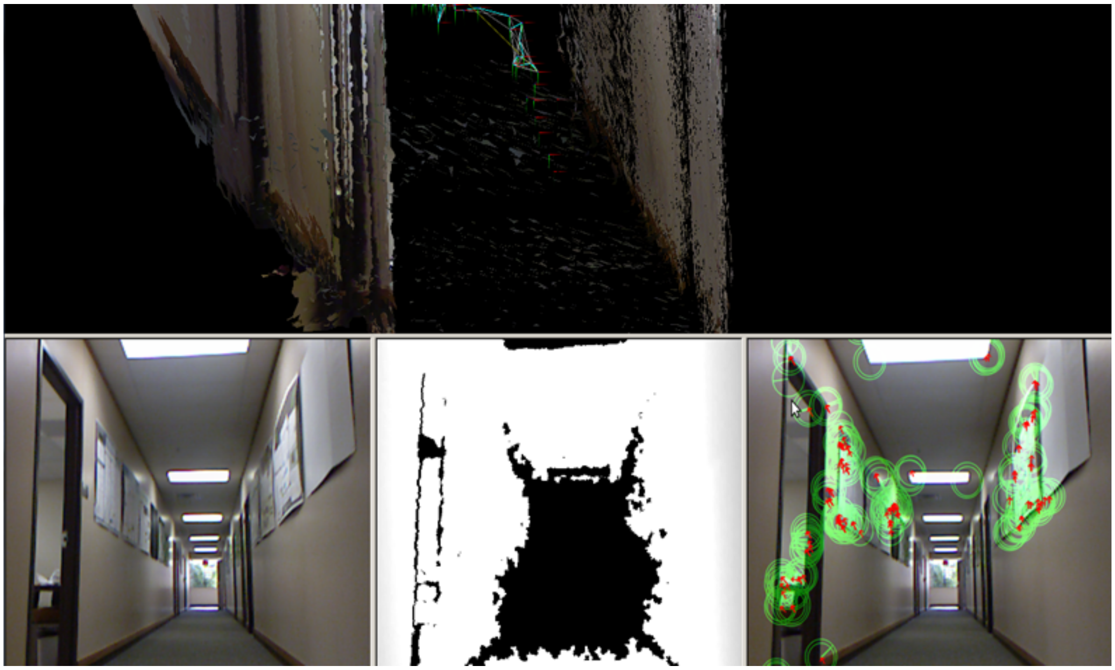


FIGURE 5.5. Extraction of visual features that are associated to 3D points

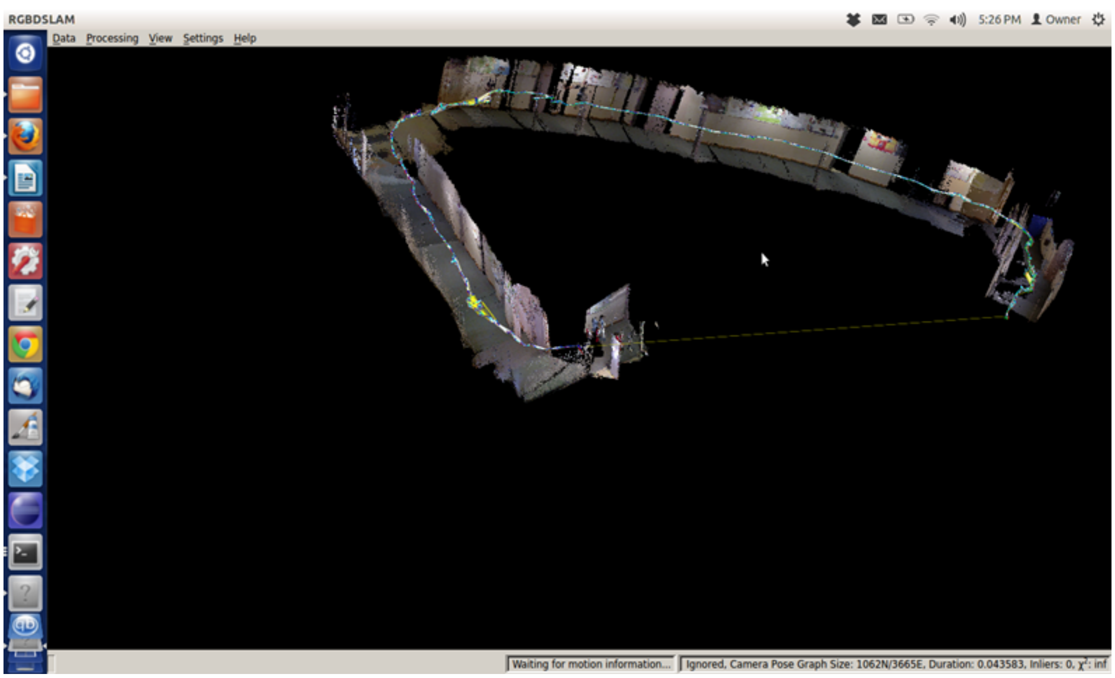


FIGURE 5.6. 3D model of the electrical engineering department at UNT

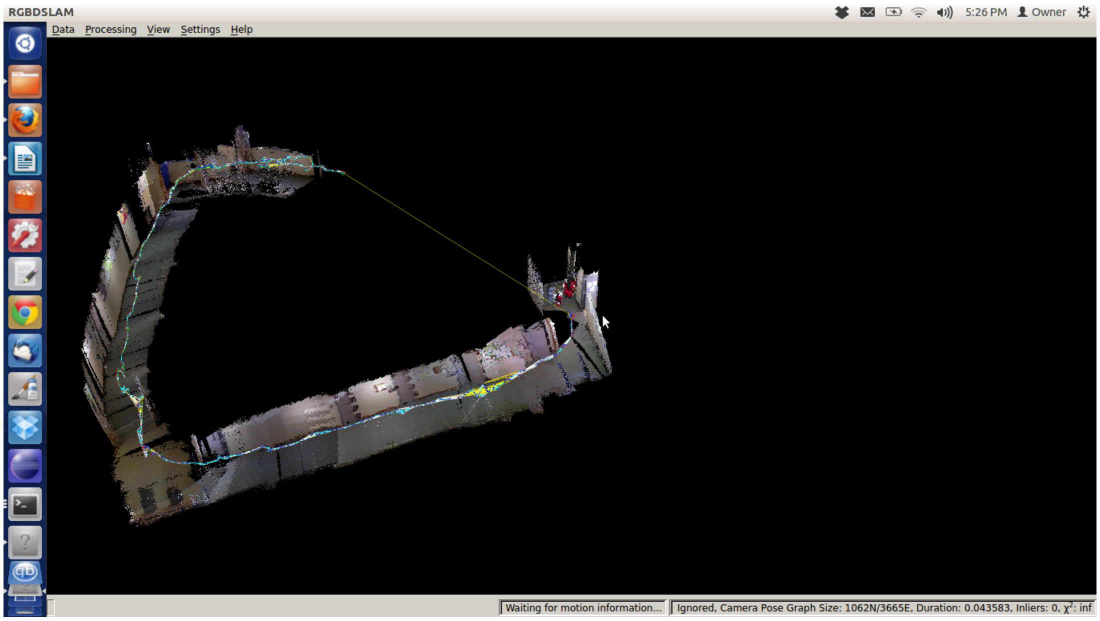


FIGURE 5.7. 3D Model electrical engineering department alternative view

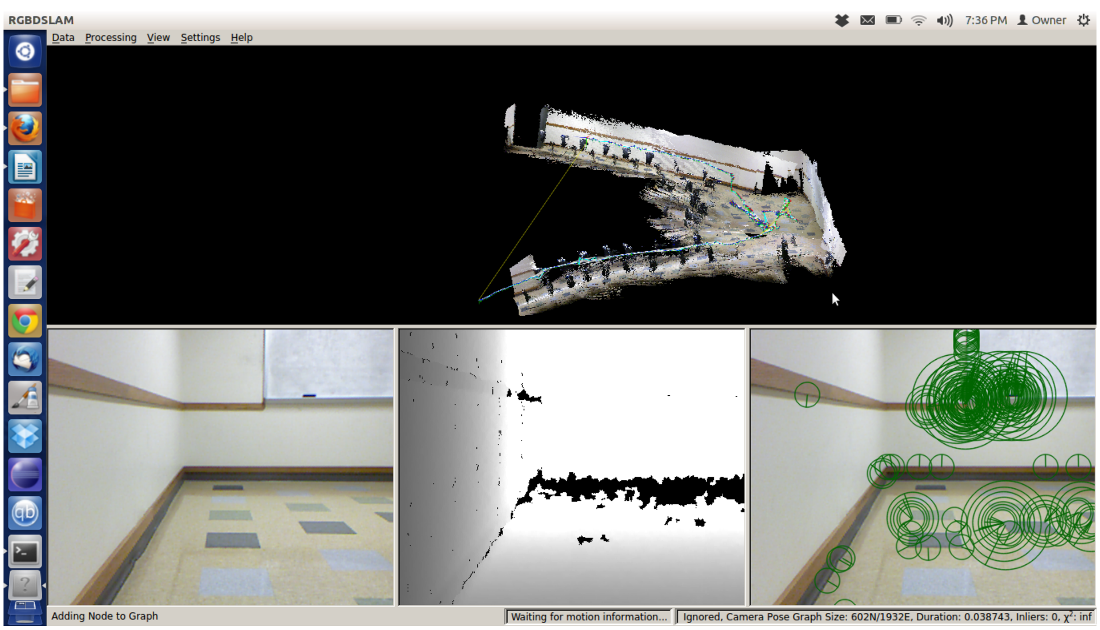


FIGURE 5.8. 3D modeling of a room at the electrical engineering department

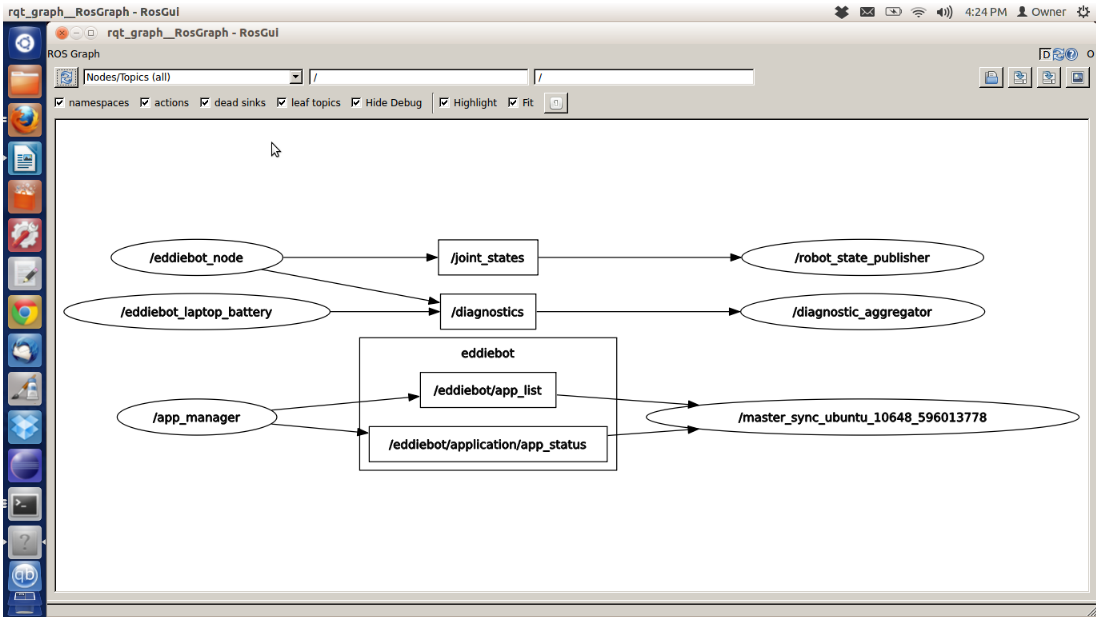


FIGURE 5.9. Node interaction in ROS

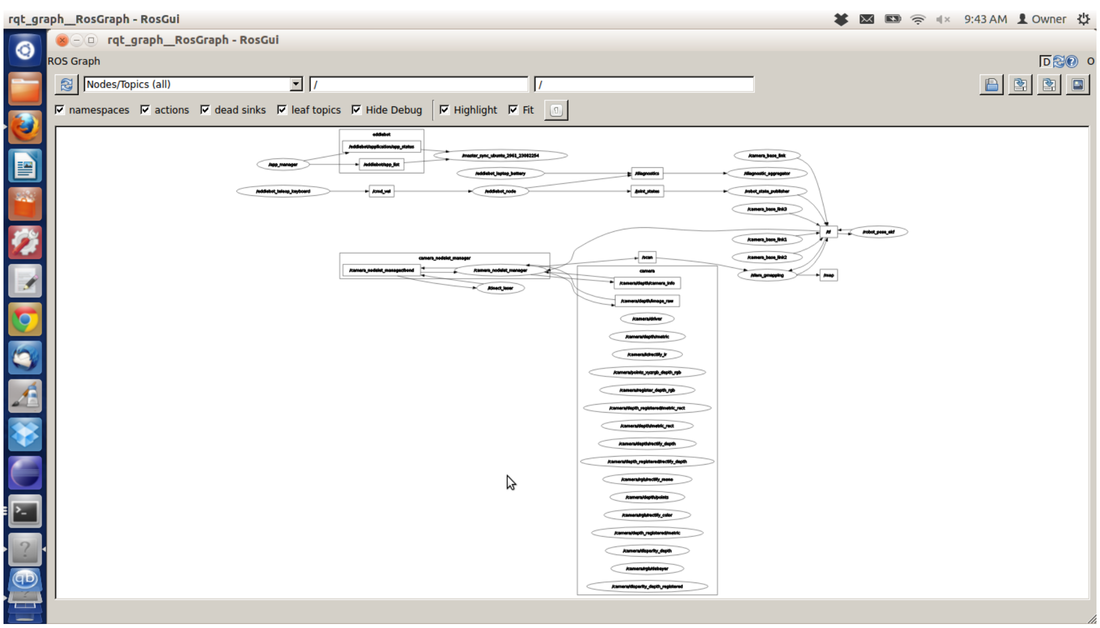


FIGURE 5.10. ROS node interaction

CHAPTER 6

SUMMARY, CONCLUSIONS AND FUTURE WORK

6.1. Summary and Conclusions

In this thesis, a novel approach to improve accuracy of an indoor positioning system based on WIFI fingerprinting was presented.

The approach takes advantage of the WIFI repeater option in most smartphones to create dynamic access points and fingerprints. The accelerometer embedded in most smartphones was used to predict the movement patterns of the users using machine learning algorithms.

2 deterministic and 2 probabilistic algorithms were used to calculate the position of the users while indoors at 2 locations in the Discovery Park building. For each algorithm a set of parameters where tested and the best parameter that returned the highest accuracy was the one used for real time indoor positioning.

In this thesis, 2 access point selection strategies, the MaxMean and RndMean, were used to select a set of access points from all the available in the Discovery Park building for the purpose of improving positioning.

For non-dynamic IPS, at location 1 the APE was of 7.2 meter and at location 2 of 3.5 meters.

For a non-dynamic IPS including DAFs, the best accuracy was obtained for an APE of 2.8 meters at location 1 considering 6 DAF, using the WKNN algorithm with a parameter of 3; at location 2, the best APE obtained was of 2.2 meters, considering 5 DAF, using the MMSE alogrithm with parameter 3.

For an IPS consisting of DAF only, the best accuracy was obtained for an APE of 2.2 meters considering 4 DAF at location 1 using the MMSE algorithm with a parameter of 3; at location 2 the best APE obtained was of 4.2 meters considering 10 DAF using the MMSE alogrithm with parameter 2.

Both experiments show that a limit of 5 DAF are needed to obtain improvement of an existing indoor positioning system. The accuracy was decreased or stayed constant when considering more than 5 DAFs. Since the availability of the dynamic fingerprints and access points change over time, as users are added or removed to the system, there exists an high probability that at least half of the total number of DAFs are available for positioning, which are the least number of DAFs needed to maximize the accuracy of the system.

In the case of an indoor positioning system consisting solely of DAFs; at location 1 it was shown that the best accuracy with the fewest DAFs is obtained with 4 DAFs. At location 2 the more DAFs were considered, the better the accuracy that was obtained. Those results show that if the accuracy is not increased it is constant for a DAF only indoor positioning system.

The factors that let to obtain a small change in the accuracy results at the 2 locations were influenced by the size of the area, number of fingerprints obtained and the environmental noise available.

Regarding the SLAM for robots project, the main section of the electrical engineering department was successfully used to provide navigation and localization of the robot indoors with the use of landmarks deployed in the environment.

6.2. Future Work

As future work of this thesis, the inclusion of other ambient signals can be studied to increase the efficiency when characterizing the fingerprints to decrease the chances to provide an incorrect location indoors. Example of those signals can be the earth magnetic fields or the frequency modulated signals, which are available in the environment.

The fusion of the dynamic WIFI indoor positioning system with other technologies can be studied, for example Bluetooth BLE, RFID tags or UWB sensors, in order to increase the performance of the positioning system. The selection of the technology should be made wisely according to the needs of the user, since each one of them has its own advantages and disadvantages.

A dynamic indoor positioning system was developed in 2 locations at Discovery Park with the help of volunteers to create the dynamic scenarios; The results of this thesis show improvement for an indoor positioning system. A next step for this project is to deploy the system in all the Discovery Park building using the WIFI hotspot signals from students and professors and after that continue with the deployment in buildings located in main campus.

Regarding the SLAM for robots project, the future work of the project is to create a cooperative slam with the robots available at the electrical engineering department. The aim of the project is to help localize and navigate the group of robots more efficiently.

[1] *Robot operating system*, `http://www.test.org/doe/`. 81

[2] *Smartphone sales overtake personal computers for the first time [study]*, `http://mashable.com/2012/02/03/smartphone-sales-overtake-pcs/`. 2, 31

[3] *Waikato environment for knowledge analysis, weka*, `http://www.cs.waikato.ac.nz/ml/weka/`. 39

[4] P. Bahl and V.N. Padmanabhan, *Radar: an in-building rf-based user location and tracking system*, INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE, vol. 2, 2000, pp. 775–784 vol.2. 16, 49

[5] Y. Chen, D. Lymberopoulos, J. Liu, and B. Priyantha, *Fm-based indoor localization*, Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services (New York, NY, USA), MobiSys '12, ACM, 2012, pp. 169–182. 23

[6] Y. Chen, Q. Yang, J. Yin, and X. Chai, *Power-efficient access-point selection for indoor location estimation*, Knowledge and Data Engineering, IEEE Transactions on 18 (2006), no. 7, 877–888. 20, 21, 34

[7] D. Dardari, E. Falletti, and M. Luise, *Satellite and terrestrial radio positioning techniques: A signal processing perspective*, Academic Press, Academic, 2012. 4

[8] H. Durrant-Whyte and Tim Bailey, *Simultaneous localization and mapping: part i*, Robotics Automation Magazine, IEEE 13 (2006), no. 2, 99–110. 77, 78

[9] F. Endres, J. Hess, N. Engelhard, J. Sturm, D. Cremers, and W. Burgard, *An evaluation of the rgb-d slam system*, Robotics and Automation (ICRA), 2012 IEEE International Conference on, May 2012, pp. 1691–1696. 2, 79, 80

[10] S. Fang and T. Lin, *Accurate indoor location estimation by incorporating the importance of access points in wireless local area networks*, Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE, Dec 2010, pp. 1–5. 20

[11] RM Faragher and RK Harle, *Smartslam–an efficient smartphone indoor positioning system exploiting machine learning and opportunistic sensing*. 25

[12] Z. Farid, R. Nordin, and M. Ismail, *Recent advances in wireless indoor localization techniques and system*, Journal of Computer Networks and Communications 2013 (2013). 12, 18, 19

[13] S. Fernandez, D. Gualda, J.C. Garcia, J.J. Garcia, J. Urena, and R. Gutierrez, *Indoor location system based on zigbee devices and metric description graphs*, Intelligent Signal Processing (WISP), 2011 IEEE 7th International Symposium on, Sept 2011, pp. 1–5. 23

[14] A.R. Jimenez, F. Seco, C. Prieto, and J. Guevara, *A comparison of pedestrian dead-reckoning algorithms using a low-cost mems imu*, Intelligent Signal Processing, 2009. WISP 2009. IEEE International Symposium on, Aug 2009, pp. 37–42. 36

[15] C. Laoudias, G. Constantinou, M. Constantinides, S. Nicolaou, D. Zeinalipour-Yazti, and C.G. Panayiotou, *The airplace indoor positioning platform for android smartphones*, Mobile Data Management (MDM), 2012 IEEE 13th International Conference on, July 2012, pp. 312–315. 46

[16] B. Li, J. Salter, A. G. Dempster, and C. Rizos, *Indoor positioning techniques based on wireless lan*, LAN, First IEEE International Conference on Wireless Broadband and Ultra Wideband Communications, pp. 13–16. 49

[17] B. Li, Y. Wang, H.K. Lee, A. Dempster, and C. Rizos, *Method for yielding a database of location fingerprints in wlan*, Communications, IEE Proceedings- 152 (2005), no. 5, 580–586. 20

[18] Hui Liu, H. Darabi, P. Banerjee, and Jing Liu, *Survey of wireless indoor positioning techniques and systems*, Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on 37 (2007), no. 6, 1067–1080. 6, 18, 52

[19] D. G. Lowe, *Distinctive image features from scale-invariant keypoints*, Int. J. Comput. Vision 60 (2004), no. 2, 91–110. 80

[20] A.R. Pratama, Widyawan, and R. Hidayat, *Smartphone-based pedestrian dead reckoning as an indoor positioning system*, System Engineering and Technology (ICSET), 2012 International Conference on, Sept 2012, pp. 1–6. 21

[21] N. B. Priyantha, A. Chakraborty, and H. Balakrishnan, *The cricket location-support system*, Proceedings of the 6th Annual International Conference on Mobile Computing and Networking (New York, NY, USA), MobiCom '00, ACM, 2000, pp. 32–43. 24

[22] T. Roos, P. Myllymki, H. Tirri, P. Misikangas, and J. Sievnen, *A probabilistic approach to wlan user location estimation*, International Journal of Wireless Information Networks 9 (2002), no. 3, 155–164 (English). 17, 49

[23] Beomju Shin, Jung Ho Lee, Taikjin Lee, and Hyung Seok Kim, *Enhanced weighted k-nearest neighbor algorithm for indoor wi-fi positioning systems*, Computing Technology and Information Management (ICCM), 2012 8th International Conference on, vol. 2, April 2012, pp. 574–577. 51

[24] K.P. Subbu, B. Gozick, and R. Dantu, *Indoor localization through dynamic time warping*, Systems, Man, and Cybernetics (SMC), 2011 IEEE International Conference on, Oct 2011, pp. 1639–1644. 25

[25] Wee Lum Tan, K. Bialkowski, and M. Portmann, *Evaluating adjacent channel interference in ieee 802.11 networks*, Vehicular Technology Conference (VTC 2010-Spring), 2010 IEEE 71st, May 2010, pp. 1–5. 15

[26] D. Tse, *Fundamentals of wireless communication*, Cambridge university press, 2005. 7, 13

[27] I. H. Witten and E. Frank, *Data mining: Practical machine learning tools and techniques*, Morgan Kaufmann, 2005. 39

[28] J. Xiao, Z. Liu, Y. Yang, D. Liu, and X. Han, *Comparison and analysis of indoor wireless positioning techniques*, Computer Science and Service System (CSSS), 2011 International Conference on, June 2011, pp. 293–296. 22

[29] M. Youssef and A. Agrawala, *The horus wlan location determination system*, Proceedings of the 3rd International Conference on Mobile Systems, Applications, and Services (New York, NY, USA), MobiSys '05, ACM, 2005, pp. 205–218. 49

[30] M.A. Youssef, A. Agrawala, and A. Udaya Shankar, *Wlan location determination via clustering and probability distributions*, Pervasive Computing and Communications,

2003. (PerCom 2003). Proceedings of the First IEEE International Conference on, March 2003, pp. 143–150. 20

[31] X. Zhang, M.K. Banavar, M. Willerton, A. Manikas, C. Tepedelenlioglu, A. Spanias, T. Thornton, E. Yeatman, and A.G. Constantinides, *Performance comparison of localization techniques for sequential wsn discovery*, Sensor Signal Processing for Defence (SSPD 2012), Sept 2012, pp. 1–5. 23