2012

WiFiPoz -- an accurate indoor positioning system

Xiaoyi Ye
Eastern Washington University

WIFIPOZ — AN ACCURATE INDOOR POSITIONING SYSTEM

A Thesis

Presented To

Eastern Washington University

Cheney, Washington

In Partial Fulfillment of the Requirements

for the Degree

Master of Computer Science

By

Xiaoyi Ye

Winter 2012

THESIS OF XIAOYI YE APPROVED BY

____ DATE________

CAROL TAYLOR, GRADUATE STUDY COMMITTEE

____ DATE________

KOSUKE IMAMURA, GRADUATE STUDY COMMITTEE

MASTER'S THESIS

In presenting this thesis in partial fulfillment of the requirements for a master's degree at Eastern Washington University, I agree that the JFK Library shall make copies freely available for inspection. I further agree that copying of this project in whole or in part is allowable only for scholarly purposes. It is understood, however, that any copying or publication of this thesis for commercial purposes, or for financial gain, shall not be allowed without my written permission.

Signature______________________

Date__________________________

Abstract

Location based services are becoming an important part of life. Wide adoption of GPS in mobile devices combined with cellular networks has practically solved the problem of outdoor localization needs. The problem of locating an indoor user has being studied only recently. Much research contributed to the innovative concept of an indoor positioning system. By analyzing different technologies and algorithms, this thesis concluded that, considering a trade-off between accuracy and cost, a Wi-Fi based Fingerprint method is proved to be the most promising approach to determine the location of a mobile device. However, the Fingerprint method works in two phases — an offline training phase (collection of Received Signal Strength signatures) and an online phase in which data from the first phase is used to determine the current position of a mobile user. The number of training points in a certain area has a direct impact on the accuracy of the system. As a result, the offline phase is a tedious and cumbersome process and the positioning systems are only as accurate as the offline training phase has been detailed. Moreover, the offline phase must be repeated every time a change in the environment occurs.

To avoid these limitations, we focus on improving the accuracy of the indoor positioning system, without increasing the number of training points. This thesis presents a Wi-Fi based system for locating a user inside a building. The system is named WiFiPoz, which means Wi-Fi positioning system based on the zoning method. WiFiPoz has a novel approach to Fingerprint method that incorporates Propagation and zoning methods. Experimental results show that WiFiPoz is highly efficient both in accuracy and costs. Compared to traditional Fingerprint methods, with the optimization of the accuracy of the location estimation, WiFiPoz reduces the number of training points. This feature makes it possible to quickly adapt to changes in the environment.

In order to explore another possible solution, this thesis also developed, implemented and tested an indoor positioning system named GIS (Geometric Information based positioning System), which is based on a model proposed by another researcher.

Several experiments were run in the offline phase and results were compared between the traditional Fingerprint method, GIS and proposed WiFiPoz. We concluded that WiFiPoz

is a more efficient and simple way to increase the accuracy of the location determination with fewer training points.

Acknowledgement

First, I would like to thank my advisor Dr. Carol Taylor for her valuable guidance and suggestions in academics and research. Dr. Taylor's help throughout the project ensured I was able to continue my work at all times. I also need to thanks Dr. Taylor for proofreading my thesis and her valuable comments. In another words, my work and accomplishments were only possible because of her help and encouragement.

My gratitude also goes out to members of the thesis commit, Dr. Imamura and Doris, who have provided me with useful suggestions and remarks on how to improve my work in various ways. James Lamphere, the systems administrator in CS department also provided me some useful information about the network in Computer Science building. Without that information, I couldn't start the experiments.

I also take the opportunity to express my gratitude for Dr. Schimpf and Dr. Rolfe for guiding my graduate studies.

I am also extremely thankful for other professors in computer science department: Dr. Li, Professor Capaul, Professor Steiner, Dr. Clark, Dr. Simmons, Professor Lemelin and Dr. Xu.

And last but definitely not least, I would like to thank my husband, Zhimin Wang, for his continuous support during my studies, even when we are 260 miles apart.

Table of Contents

Table of Figures

Table of Tables

List of abbreviations

- AOA (Angle of arrival)
- AP (Access Point)
- GPS (Global Positioning System)
- IPS (indoor positioning system)
- k-NN (k-nearest neighbor)
- LBS (Location-based service)
- LOS (line-of-sight)
- NLOS (Non line-of-sight)
- RF (Radio frequency)
- RFID (Radio frequency Identification)
- NNSS (Nearest Neighbor in Signal Space)
- RM (Radio Map)
- RSS (Received Signal Strength)
- RSSI (Received Signal Strength Indication)
- SSID (Service Set Identifier)
- TOA (time of arrival)
- Wi-Fi (Wireless fidelity)
- WLAN (wireless local area network)

Chapter 1: Introduction

1.1 Background and Motivation

Indoor Positioning Systems are very Useful

Location based services (LBSs) are becoming an important part of life [1] [2]. A survey estimated that LBSs will generate annual revenues on the order of US$15 billion worldwide [3]. Wide adoption of GPS in mobile devices, combined with cellular networks, has practically solved the problem of outdoor localization needs [4]. As GPS employs satellites orbiting earth which transmit radio signals describing their current position, these radio waves are difficult to pass through the walls and ceilings of buildings. Moreover, GPS is only able to locate devices as accurate as approximately 10 meters [5]. As a consequence, indoor scenarios require an alternate positioning technique which is able to position devices within a building. As part of Location based services, indoor location awareness is important for such fields as ambient intelligence, assisted daily living, behavior analysis, social interaction studies, and myriads of other context-aware applications [6].

There are many examples in which it would be favorable to have an indoor navigation system. For example, people might need to navigate through a public building like a mall or hospital. It can be hard to find your way in such a building. A system that would help you navigate from one room to the next would be very helpful. Another example is that, a visitor can access the online service at the entrance of a museum. The system can give information about the exhibits in the building based on the user's current location. For locating services, the indoor positioning system can help the visitor to find the "point of interest" inside a building. The system can display the location of the nearest printer or first aid box for the assisted person.

The basic application of the indoor positioning system is to help visitors find their way to a certain destination. This can be done in many different ways. One of the examples of this is the use of signs. If signs are properly placed throughout a building, a visitor should be able to find the information they need, but in some situations it is still confusing. Plus, the visitor might not be able to find out their position. It is much more convenient if the user only receives personal navigation instructions based on awareness of their current position [4].

Using handheld devices equipped with a wireless network based positioning system to give personal navigation instructions is a good choice. After a decade of hardware progress, ubiquitous computing is starting to become reality; handheld and wearable computers have been thoroughly integrated into everyday activities. Those mobile devices can be more powerful when they are able to get context-aware information, including position [1].

On the other hand, wireless networks are primarily designed for voice and data communications. The widespread availability of wireless nodes, however, makes it possible to utilize the networks for wireless location purposes as well. It is expected that location-based applications will play an important role in future wireless markets [3]. A mobile device with a wireless network connection would work just like a GPS navigation device that the user may be familiar with. Application level software will incorporate location information into its features, be more aware of the facilities and objects in the surroundings, and able to give the right information at the right moment to the user.

Existing Research and Products do not Fulfill the Needs

While it is important to have indoor navigation, the field of indoor positioning and navigation is still very limited. Over the past few years, significant effort has been dedicated to the development of indoor localization systems. Many people in academia and industry are currently involved in the research and development of these systems and have contributed creative and innovative concepts of indoor positioning systems with various platforms and architectures [7]. Despite these efforts, existing indoor positioning systems are still limited. The results vary in performance and cost as they either require expensive infrastructure or special custom made devices such as ultrasound and RFID) [8], have low accuracy [9] [10] , or require much manual effort to process the data.

On the other hand, indoor localization methods based on Wi-Fi signal strength are becoming more and more attractive because they don't require additional infrastructure costs beyond the existing Wi-Fi infrastructure and need no additional hardware. Apple or Google provide positioning services that employ the availability of Wi-Fi networks, which utilize a database that stores the geo-location of positions, where certain Wi-Fi networks are visible. If a mobile device is within the range of a network that is known to the system, it is positioned at the known geo-location of this network. As these systems are very inaccurate (Wi-Fi positioning on the apple iPhone 3G is accurate to 74 meters on average [11]), scenarios that require accuracy better than approximate 10 meters require more sophisticated positioning techniques [12].

Overall, how to utilize Wi-Fi signals to get a new location system that can function with better accuracy is an active and important area of research.

1.2 Objectives of this Thesis

The first objective is to investigate the technologies that are currently available and that can be used for creating an indoor positioning system. This is done by investigating the background of radio frequency technology and the techniques used to obtain the estimation of a user's position.

The primary objective of this project is to design a system that would be able to localize the assisted person in an indoor environment with high accuracy, but not require expensive hardware and software and work with the existing infrastructure.

1.3 Thesis Outline

Chapter 2 reviews the literature on indoor positioning systems, discusses the basic technique used for indoor positioning systems and explains the advantages and disadvantages of those techniques. Chapter 2 also discusses work done by other researchers to solve the indoor positioning problem. Proposed positioning systems have been developed using methods of Fingerprint and Propagation, and will be discussed in detail.

Chapter 3 describes the theory and design of the proposed positioning system. The algorithm design and logic is explained in detail, including different design features of the algorithm which motivated the methods used to solve the positioning problem.

Chapter 4 discusses the environment, equipment and experimental setup. The experiments are designed for off the shelf, low cost, easily available equipment and work on the existing infrastructure. Chapter 4 also discusses an analysis of empirical results and investigates how different design features of the algorithm will impact the accuracy of an indoor positioning system.

Chapter 5 identifies major conclusions from the research that leads to the contribution to knowledge and explores the suggestions for future work.

1.4 Contribution of this Thesis

The main contributions of this thesis are:

1. Identification of techniques suitable for localization and discovery of their limitations.
2. Localization methods and existing positioning systems are analyzed. This provides a broad overview on the topic of indoor localization in general and the various technologies that are currently available.
3. Implementation of an indoor positioning system WiFiPoz (Wi-Fi positioning system based on zoning method) based on widely used techniques.
    - Discover a good balance between accuracy and preparation effort. The software and equipment used is inexpensive and available off the shelf. The positioning system is easy to deploy.
    - Reduce the effort of maintaining the system due to changes in the environment.
    - Demonstrate the accuracy achieved using this method as comparable to other existing techniques. It is capable of determining the position of mobile devices within 2 meters.

Chapter 2: Literature Review

As mentioned in Chapter 1, estimating the location of a mobile device user inside a building is a very useful function for many different applications. This can be achieved by using various sensing technologies, such as infrared (IR), ultrasound and radio frequency (RF) signals [1]. Positioning systems using infrared and ultrasound are needed for line-of-sight, because IR and ultrasound can do not penetrate objects very well. The indoor environment is packed with obstructions, such as walls and furniture that make IR and ultrasound unsuitable for indoor use [13]. As a result, RF radio signal technology has received the most attention from researchers and is the most used technology for indoor location estimation [14].

This chapter provides a comprehensive review of the RF positioning system for indoor applications. All mentioned technologies are evaluated against their specific advantages and disadvantages for the implementation into the indoor positioning system.

2.1 Overview of RF Positioning Systems

Radio positioning systems involve the interaction between at least two communicating devices, transmitters and receivers. In all systems (Figure 2-1), a signal is transmitted from the transmitter and propagated through the air to the receiver. The receiver (mobile station) can estimate or calculate its location based on the properties of received signal.

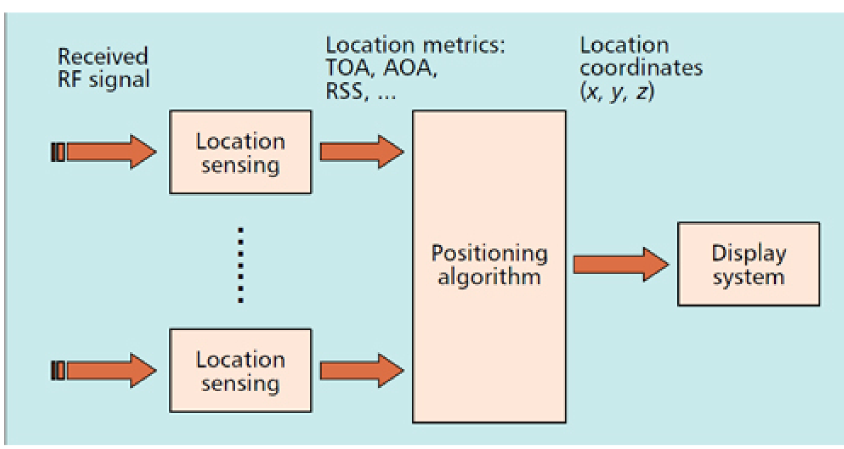


Figure 2-1: A Functional Block Diagram of Positioning System [15]

The first step of a positioning system is to read the received RF signal. Different kinds of RF sensors have advantages and disadvantages which are discussed below. The sensed signal provides information on Time of Arrival (ToA), Angle of Arrival (AoA) or received signal strength (RSS). ToA and RSS can be used for estimation of distance. The estimated distance is then deployed for the location estimation algorithm, which is the second major part of a location estimation system. The last part of a positioning system is the display of the x, y coordinates for a two or three dimensional area to a display system for the user [15]. The display will not be discussed in this thesis.

2.2 Major advantages and Disadvantages of RF Based Indoor Positioning Systems

The main advantage of using on RF based positioning system is the fact that radio signals generally have the ability to travel through objects. Although having a clear line-of-sight (LoS) will have better results on the signal quality, there is no absolute need for line-of-sight between a transmitter and receiver to get an acceptable result [16]. All these RF signals are subject to reflection, diffraction, absorption and multipath and these effects vary at different frequencies.

2.2.1 The Indoor Environment

LoS is not usually available over any long distance in the indoor environment. Walls are made of different materials. Furniture made of wood and steel is very common in offices. The density and movement of people inside a building changes rapidly. Humidity and temperature are more controllable inside a building. All of these factors cause the RF signal to reflect, diffract, scatter and absorb differently in various indoor environments.

2.2.2 Multipath Problem

The existence of multipath propagation is a major disadvantage of RF based systems.

Multipath propagation is the phenomenon that results in radio signal's reaching the receiving entity by two or more paths. It changes the characteristics of the radio signals sent by the transmitter along the way to the receiver [17].

As the example in Figure 2-2 shows, the direct path penetrates 3 walls but there is also another indirect path penetrating only one wall. Since the signal on the direct path is greatly attenuated when crossing the walls, the signal on the indirect path contributes to most of the received signal. So if we only consider the direct path, the predicted signal power at the receiver will be much lower than the real measurement [17].

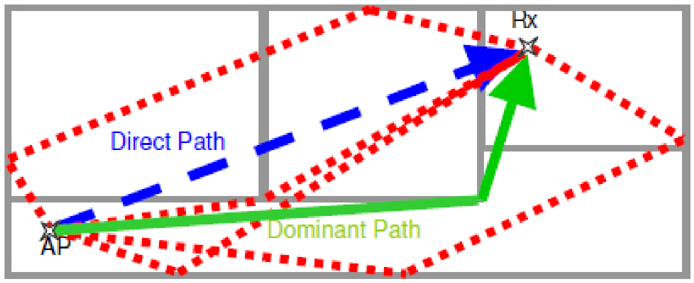


Figure 2-2: Illustration of the Multipath Model [17]

Multipath propagation makes it hard to derive the proper characteristics from the radio signal. Any objects like walls, furniture or people can cause multipath propagation. [12] [15] [18]

2.3 Categorization of Indoor Positioning Systems

Many people in academia and industry are currently working on the research and development of RF based indoor positioning systems using various technologies, but the existing systems are still limited. In this section, widely used technologies for indoor positioning systems will be introduced. They are categorized based on the three aspects of RF based indoor positioning systems: sensors, measurement methods and positioning algorithms.

2.3.1 Classification of Indoor Positioning Systems According to RF Sensors

This section classifies general approaches of different position systems relying on radio wave technology. The major sensing technologies, according to Tauber, are Wi-Fi, GPS, GMS, blue-tooth and RFID [19]. While there may be other technologies used for positioning systems, they are not included because they are outside the scope of my research, which is limited to wireless, signal based technologies.

Different wireless sensors can be deployed to produce the raw data for an indoor positioning system. The RF signals are subject to reflection, diffraction, absorption and multipath and the effects vary with different frequencies. However, each RF sensor technology has its own advantages and disadvantages based on the signal characteristics. Propagation speed, available bandwidth, cost, safety, power constraints, hardware requirements and regulatory constraints all play a role in influencing the selection of sensors for the design of positioning systems. [19]

2.3.1.1 Wi-Fi based Position Systems

The world is encountering a technological boom in the wireless industry and related applications. An unprecedented growth in WLAN (Wireless Local Area Networks) technology has been witnessed over the last few years. 'Infonetics Research' found that nearly all the latest laptops, personal digital assistants (PDAs) and mobile phones now come equipped with a wireless network interface card. This demonstrates that there is a growing demand for WLAN, which is not only restricted to businesses but equally in demand at universities, offices and homes [13].

"Wi-Fi" is one of the Institute of Electrical and Electronics Engineers or IEEE standard for WLAN. Wi-Fi stands for "Wireless Fidelity" and it is truly a freedom to make a connection to the Internet without the hassle of old- fashioned network cables. Wi-Fi or 802.11b is one of the 802.11 specifications family and is also the first dominating standard in the market [20]. This innovative technology is evolving technically and practically, leading WLAN to be a common sight at universities, airports, coffee shops, offices and organizations. The latest amendment of the original standard 802.11 is 802.11n, which was ratified in late 2009 [21]. IEEE 802.11ac is currently under development, which will provide a high throughput WLAN. However, older standards such as 802.11a/b/g are still commonly used.

In summary, Wi-Fi networks are qualified for indoor positioning for two main reasons. The first reason is Wi-Fi networks are used in public locations. The second reason is most mobile devices like laptops and smart phones support the wireless IEEE standard 802.11. Wi-Fi is a ubiquitous technology that is broadly accepted by the users and represents a cost efficient and reliable technique that indoor positioning services can employ.

2.3.1.2 Global Positioning System (GPS)

The Global Positioning System (GPS) (Figure 2-3) which focuses on providing positioning capability in open outdoor environments has been proven to work accurately [5]. It uses Time-of-Arrival technology between the satellite sending its signal and the moment that the signal was received. The receiver also calculates the position of the satellite based on information periodically sent in the same signal. By comparing the position and range of multiple satellites through a triangulation process, the receiver can calculate its own location down to meter-level accuracy [22] in outdoor environments.

Figure 2-3: Global Positioning System (GPS)

The accuracy, price and size of GPS receivers make this technology suitable for outdoor navigation. GPS, however, has never worked as accurately in the indoor environment which has a more restricted space. Due to walls and other obstructions, a GPS signal is usually difficult to receive indoors. This makes it difficult to calculate distances between the reference points and the location device [23].

In the future, new GPS receivers which have a better tracking sensitivity can be expected [24]. These can open doors for indoor positioning and navigation, but it still depends on the conditions of the building where it is used. For example, better receiver might work on the top floor, but the signal will still be too weak for the other floors of a building.

2.3.1.3 GSM based Systems: The Global System for Mobile Communication

Overall, cellular networks provide much better coverage than Wi-Fi, but for a long time they were not considered for indoor localization due to their low accuracy.

This research relies on the use of wide fingerprints which can provide good accuracy, of about 2.5-5.4m [25]. However, it requires special hardware (programmable GSM modem and CDMA scanners). With narrow fingerprints which could be acquired with conventional hardware, the localization accuracy is low, more than 10 meters [25], which is not acceptable for the indoor environment.

2.3.1.4 Bluetooth Based Systems

A large number of handheld devices now have Bluetooth communication functionality, which makes it a technology readily available for various functions such as positioning and navigation. Many papers have been published about using Bluetooth for positioning [26] [27].

The most important problem when using Bluetooth for positioning is the uncertainty of response time. The discovery of devices that are near to the user takes from 2 to 11 seconds. In a mobile environment, this can have severe consequences on the position accuracy. Making a Bluetooth device easier to discover will increase the percentage of time spent in the so called 'inquiry state', which requires more energy. Lack of a predictable response time is still a big drawback of a Bluetooth based positioning systems. Many papers have described the response time of detecting Bluetooth devices in range, but none of them solve the problem completely [27] [26] [28].

Another problem of Bluetooth positioning systems is that there are not many devices that support the 'inquiry with RSSI' command. Without this option, it always takes time to connect to a Bluetooth device to read out RSSI values. As this cannot be done during the detection phase, it will further increase the time to get RSSI for more precise location estimation.

2.3.1.5 Radio Frequency Identification (RFID)

Radio Frequency Identification (RFID) is an automatic identification method, relying on storing remotely retrieved data. An RFID tag is required for this technology, which is an object that can be attached to or incorporated into a product, animal or person for the purpose of identification using radio waves [29]. There are two types of tags. Passive tags require no internal power source and therefore have a detectable range of only a few centimeters, so it is not practical for an indoor positioning system. Active tags require a power source and can be detected over several meters, but the price is much higher than passive tags.

Active RFID tag systems support reading out an RSSI value of the received signal. There are many positioning systems based on active RFID tags [30] [31]. However, the disadvantages of using RFID tags is that there are only very little active RFID receivers available to be used in handheld devices and the devices that are available are very expensive (about $1000 per receiver).

2.3.2 Classification of Positioning Systems according to Measurement Methods (deterministic techniques)

2.3.2.1 Received Signal Strength Indicator

The Received Signal Strength Indicator (RSSI) indicates how strong a radio signal has been received and is mostly measured in dBm (1 dBm = 1.3milliwatt). The closer a receiver is to a transmitter, the stronger the signal will be, because the signal's power attenuates as it propagates through the air and the attenuation is proportional to the distance [32].

A major drawback of this technique is the fact that the signal distribution of networks is affected by many external parameters, resulting in irregular distribution patterns rather than the ideal circular ones [32]. However, there are many reasons for choosing RSSI as a metric. RSS indicating-capable equipment is widely available and provides a cost-effective means of a positioning system. RSSI-based techniques are eligible provided they have satisfactory accuracy and require little or off- line calibration and prior knowledge of the environment [32].

In Chapter 3, 4 and 5, RSSI, RSS, SS, all indicate the measurement of received signal strength.

2.3.2.2 Time of Arrival (ToA) and Time Difference of Arrival (TDoA)

Measuring the Time of Arrival (ToA) of a transmitted signal is an accurate but far more complex technology to determine the position of a mobile device. There are many research studies based on this technology [33] [34]. The ToA technique measures the signal propagation time between a transmitter and a signal receiver [35]. The ToA technique requires the base stations and receiver to have precisely synchronized clocks. Each positioning signal sent by the base station is accumulated with a precise timestamp to determine when the signal was sent [34]. In the step of location estimation, a straightforward geometric method is used to compute the intersection point of the circles of ToA. This intersection indicates the location of a mobile user.

There are two major disadvantages of the ToA technique [36]. The first disadvantage is it requires additional very specific hardware for measurements. The measurement accuracy of ToA has a significant impact on the accuracy of the location estimation. The speed of radio signal propagation through the air is around 3×10^8 almost the same as the speed of light, An error in the measured ToA of 1 micro second leads to an error in distance estimation of around 300 meters. The second disadvantage is that all transmitters and receivers in the system have to be precisely synchronized. Time stamps must be labeled in the transmitting signal in order for the measuring unit to discern the distance that the signal has traveled.

Time Difference of Arrival (TDoA) is a variation of ToA, but it requires only the fixed stations to have synchronized internal clocks. It involves multiple stationary receivers that have synchronized clocks which collaborate to find the location of the signal's source [37]. A signal is transmitted from the mobile device to the synchronized receivers. The other important difference between TDoA and ToA is the fact that for a ToA the transmitters are at a fixed position and the mobile receivers calculate their position with the received time stamps. For the TDoA system, the transmitters are the mobile devices and there are fixed receivers that pick up the time signals from the mobile station (beacons), with which they can calculate the position of mobile devices (transmitters).

Because of the similarities with ToA and TDoA, they have the same disadvantages.

2.3.2.3 Angle of Arrival (AoA)

Angle of Arrival (AoA) describes a positioning technique where the angle of received signal is measured by Base Stations (i.e. Access Points). To measure the received angle, the Base Stations (BSs) of systems applying AoA are equipped with a direction aware antenna, which is usually composed of an array of antenna elements that are able to divide their directivity lobes equivalently among different directions [15]. If at least two base stations are able to determine this angle, the intersection of virtual lines drawn from the base station decides the position of the mobile device, as the Figure 2-4 shows. For this method, a mobile device normally plays the role of transmitter and the fixed stations play the role of the receiver, because it is easy to install the direction aware antennas to the fixed station and calibrate it properly.

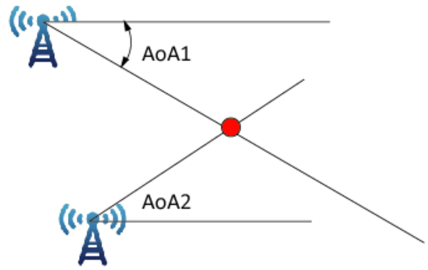


Figure 2-4: Angle of Arrival (AoA)

The AoA technique is vulnerable to certain effects of signal distribution, in particular to multipath effects as they cause the signal to reach the base stations via different paths from different angles. As a consequence, Direct Line of Sight (DLoS) is a prerequisite for flawless functionality of the AoA technique. As a result, it is impractical for scenarios vulnerable to multipath effects like indoor scenarios.

An advantage of this system is that no complex or synchronization is necessary. But the need for direction aware antennas is very costly.

2.3.3 Classification of Indoor Positioning system Based on Positioning Algorithm

Location estimation is a critical step for positioning systems. This section describes applicable methods, techniques and algorithms for radio frequency based positioning that are currently widely used.

2.3.3.1 Cell of Origin

The Cell of Origin technique (COO) determines the position of a device by assuming the device is located closest to the position of the base station which provides the best quality of service. It is also called nearest base station [16] [38]. The position of the base station (transmitter) emitting the signal with the strongest signal is assumed to equal the unknown position of the mobile device (Figure 2-5).

This technique is rather easy to deploy. The major drawback is its accuracy which depends directly on the size of the cells of the network. As network cells cover a larger area, the accuracy will be very low (e.g. approx. 100 to 1000 meters for Wi-Fi; 1 to 1000 km2 for GSM). However, in a dense network of stations, this simple technique can be very useful.

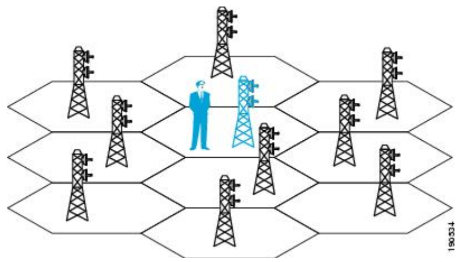


Figure 2-5: Cell of Origin [38]

2.3.3.2 Triangulation

Triangulation uses the geometric properties of triangles to compute the crossing point based on distance or angle. There are two subcategories. One is Trilateration, which uses distance measurements. The other is Angulation, using primarily angle or bearing measurements [39].

2.3.3.2.1 Lateration

Lateration or Trilateration describes techniques to determine the position of an object by the distance to three or more reference points, whose location coordinates have to be known to the system. The coordinates of all base stations in the network need to be stored in a database. As Figure 2-6 [39] shows, the distances to at least three known positions have to be known in order to locate a device on a 2-dimensional space. The distances can be calculated based on RSSI, ToA or TDoA information [39].

The drawbacks are that it is very difficult to get a good estimation of the actual distance to a base station, due to the noise factors in an indoor environment [40].

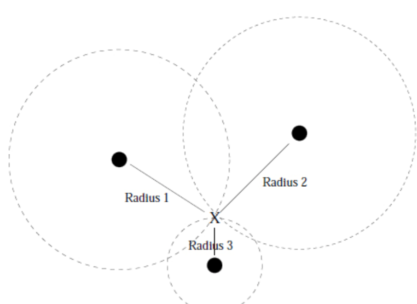


Figure 2-6: Determing 2D position using Lateration

2.3.3.2.2 Angulation

Angulation is based on the AoA technology, which requires an extra device. In general, the angle measurements at two base stations and coordinates of these two base stations are required. Simple trigonometry can then be applied to determine the actual position of mobile device, as Figure 2-7 [41] shows below.

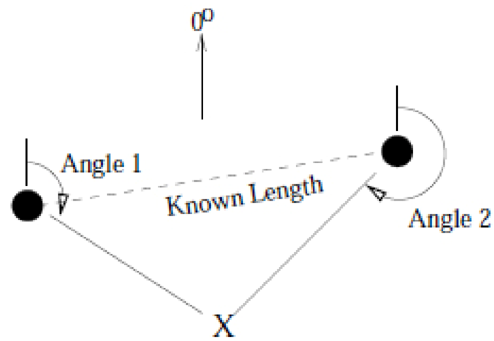


Figure 2-7: Determing 2D Position using Angulation

2.3.3.3 Fingerprint method

The Fingerprint technique is based on the fact that characteristics of the propagation signal are different at each location. This means each location has a unique 'signature'. Therefore, signal strength patterns of a wireless device are collected at different locations away from Access Points. Many researchers also call it an *"empirical model"* of location estimation or a "Pattern Recognition" technique.

The use of RSSI in the Fingerprint technique is considered to be a more accurate method for location estimation [12].

2.3.3.3.1 Terminology

Reference points are discrete locations spread over the covered area, which sometimes look like a grid.

The patterns / fingerprints / signatures describe the RSSI pattern from Access Points at certain positions which then are recorded into the database. They indicate the strength of the reception of each network at each Reference Point.

Radio map describes the signal distribution for all wireless networks receivable in the area to be covered by the positioning system. It is captured and stored in a database.

2.3.3.3.2 Algorithms

Fingerprint systems have two main phases: a **training phase (offline phase)** and a **real-time phase (online phase).**

In a training phase, patterns from different Reference Points are captured and accumulated into a Radio Map. In the later real-time phase, a new pattern will be received describing the signal characteristics at the yet unknown position. This new measured signal pattern will then be checked against the prerecorded fingerprints in the Radio Map. The most similar one is assumed to be the nearest reference point, and the coordinate of nearest Reference Point will be considered as the estimation of the unknown position (Figure 2-8).

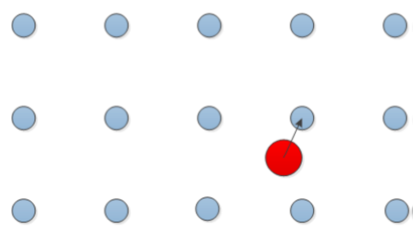


Figure 2-8: Fingerprint System [16]

Several projects were conducted based on the matching of Wi-Fi signal strength patterns [14].

One problem is the impact of reference points. The placement and number of reference points will affect the performance of algorithms. Another problem is that updates to the database will take a long time and require a lot of effort.

2.3.3.4 Propagation Method

The Propagation model is based on the fact that as a radio wave travels through an environment it loses signal strength. The loss of signal strength can then be modeled by using known radio propagation and path loss theories. Propagation model use these theories. The distance from a wireless device to an Access Point can then be calculated based on the given RSSI. The Triangulation method can then be applied to determine the location of the device [42].

For Propagation method, a no signal strength database is required. The accuracy of an estimated location could be decreased by the multipath effect. Thus, the Propagation methods are less accurate when compared to the Fingerprint techniques [14].

2.4 Research Projects

2.4.1 First research project using Fingerprint and Propagation methods

RADAR [14] is an indoor RF-based user location and tracking system, developed by Microsoft Research in 2000. This project was the first research project that put efforts into using Wi-Fi signal strengths for indoor localization. Later projects were started heavily relying on the methodologies it has represented [43].

Two different algorithms were implemented for the indoor location estimate. One of them employed the Fingerprint technique. Wi-Fi signal strength readings were measured on various points of the test site. The readings were recorded to a database. When a location query was submitted by the user, the system compared the user's current signal strength values with the signal strength values in the database. To determine the unknown user's location, the k-nearest neighbor search algorithm was proposed. The k-nearest neighbor search algorithm is a simple algorithm that stores all available examples and searches for a supplied entry the k number of the examples that have the highest similarity measures. Further explanation of the k-nearest neighbor algorithm will be explained in Chapter 3. The other method described in RADAR is Propagation method using the measured signal strength values. The proposed system was accurate at 3 meters for 50% of the time when using the Fingerprint approach and at 4.3 meters (50% of the time) when using the Propagation method.

2.4.2 Algorithms Applied to Improve the Performance of Fingerprint methods

Based on this initial work in the Fingerprint technique, this line of research was adopted and several different algorithms were applied to improve the performance. Raman Kumar, et al., [44] proposed two mechanisms using received signal strength fingerprints to improve the accuracy of a RF based location determination system that to identify location. First, they applied the well-known Kalman filter to generate filtered samples for the training step, as well as for the querying step. Kalman filtering has been used earlier for estimates of a changing location of a mobile device [45]. Next, they used a technique of "multiple observers" in location estimation. One "observer" is the mobile station. It scans the signal strength arrived from the Access Points. The mobile station estimates its location based on the information it gets. The additional "observers" are the Access Points, which support the measurement of the mobile station's RSSI. Merging the estimates from "multiple observers" can improve the accuracy.

Accuracy with a filtered training sample with a precision of 3 meters jumped from 80% with the existing method to 90%. Accuracy with a precision of 2.3 meters jumped to 90% with multiple observers. 99% accuracy was achieved by the filtering method only at about a 13 meter precision, while this accuracy was achieved with multiple observers at about 5 meters.

Three different pattern matching algorithms employed in the Fingerprint technique have been studied by Gilles Wassi et al [12]. They are the multi-layer perceptron (MLP) neural networks, the Generalized Radia Neural Network (GRNN) and the k-nearest neighbor (KNN). Their performances in terms of localization accuracy are compared on both training and testing data. The results as Table 2-1 shows, k-nearest neighbor give the best accuracy. The experimental results also show that the localization accuracy improves when the grid spacing (distance between Reference Points) is decreased.

Table 2-1: Compare Distance Error among MLP, GRNN and KNN [12]

Algorithm	**25^{th} (meter)**	**50^{th} (meter)**
KNN	1.26	2.4
MLP	1.39	3.09
GRNN	1.37	2.94
RADAR [14]	1.92	2.9

There is also a great deal of research [46] [24] [43] [32]based on the Propagation method. This approach mainly consists of two phases. The first phase is a distance estimation phase, which is the calculation of distance from RSSI values from the access points based on the formula. The second phase is the Location Estimation phase, which is the determination of the most probable location of wireless device using coordinates of various known Access Points and the distance gained at the first phase. There is a weighted center of mass based trilateration approach is proposed for the Location Estimation phase [46].

2.4.3 Zone Division

Dik Lun Lee and Qiuxia Chen from Hong Kong University [20] presented a new approach which continuously records the signal strengths received at the user client, and verify the user's location by backtracking to the user's previous locations that are not likely to be reachable from the previous locations. This method requires a Geographical location model (zoning approach) to represent the possible locations and path within the indoor environment. The zoning model integrates the map information which can deduct the estimation variance by deleting some unrealistic cases like going through walls. The zoning approach is also used in other research [32] [43].

Tussanai Parthornratt and Kittiphan Techakittiroj proposed a concept of applying radio Propagation method into separate regions/zones based on Geographical Information [43]. Transmitter – Receiver distance between Mobile Station and each Base Station can be used to find out the location of mobile station using lateration (Triangulation) method. They demonstrate this method in a small area and report this method can decrease error distance from 8.95 meters to 5.05 meters compared to the using Propagation method without applying zoning approach. But this paper just presented a very general idea about zoning approach. In this paper, I developed this concept and implemented it in a testing area with 7 zones. Details will be described in chapter 3 and 4.

2.5 Conclusion

After reviewing the different methods and techniques for an indoor positioning system, my approach will be based on existing Wi-Fi infrastructure using RSS. Plus, indoor positioning algorithms, the Fingerprint method and Propagation method combined with lateration will be tested. The basic requirements for this positioning system are:

- Should provide at least room-level accuracy, at a distance of 5 m.
- Does not require the installation of special hardware.

Table 2-2 provides a summary of all the technologies.

Table 2-2: Evaluation of Existing Technologies for Indoor Positioning System

	Indoor Error Distance <=5m	No extra cost	note
Classification by RF Sensors			
Wi - Fi	√	√	
GPS		√	
GSM	√		To get ED < 5m, need a hardware
Bluetooth	√	√	Long response time
RFID		√	
Classification by measurement methods			
RSSI	√	√	
TOA / TDOA	√		
AOA	√		
Classification by Positioning Algorithm			
Cell of Origin		√	
Lateration (Triangulation)	√	√	
Fingerprint method	√	√	
Propagation method	√	√	

Considering different RF Sensors, GPS performs very well in an outdoor environment, but the GPS signal is usually difficult to receive indoors, and will end up with a high error rate. GSM based systems require special hardware. Otherwise the localization accuracy will be rather low. A large number of handheld devices now have Bluetooth functionality, but it always takes a long time to read signal values, which is not practical for a mobile positioning system. Active RFID receivers are very expensive. Wi-Fi is a ubiquitous technology that is broadly accepted by users and represents a cost efficient and reliable technique that indoor positioning services can employ.

Analysis of the current signal measurement methods included, the angle of arrival (AoA), received signal strength (RSS), time of arrival (ToA) and time difference of arrival (TDoA). The techniques of AoA, ToA TDoA require a degree of time synchronization that is difficult to achieve using inexpensive off-the-shelf WLAN hardware. However, RSS indicating – capable equipment is widely available in Wi-Fi devices. Utilizing existing WLAN infrastructure by reading RSS is a cost effective solution for this thesis.

For the positioning algorithms, Cell of Origin is not accurate. Existing non-trainable algorithms such as Propagation method do not work very well when compared against trainable algorithms (Fingerprint method). Plus, triangulation (Lateration) usually works well with the Propagation method. There are trade - offs with all of these approaches, but the trainable Fingerprint method has the lowest error distant rate. It seems to be the most promising approach to determine the location of a mobile device.

In summary, this thesis is developing a cost effective solution for an indoor positioning system with high accuracy, based on an existing Wi-Fi infrastructure and using RSSI. It will take advantage of the Fingerprint and other positioning algorithms such as Propagation method. The details of the proposed indoor positioning system will be introduced in Chapter 3 & 4.

.

Chapter 3: Positioning System Design and Implementation

3.1 Over view

This chapter proposes a positioning system named **WiFiPoz** (Wi-Fi positioning system based on the zoning method). The design and algorithms will be explained in detail. The details of implementation will be introduced in Chapter 4. In order to explore another possible solution, this chapter also introduces an indoor positioning system named **GIS** (Geometric Information based positioning System, which is based on a proposed concept model by another researcher [43].) The analysis of results produced by **WiFiPoz** and **the GIS** system will be covered in Chapter 4.

The general idea of WiFiPoz and GIS is to divide an area into several zones based on the geographic features and then combine the traditional Fingerprint method and the Propagation method to determine the position of the mobile device. Generally speaking, they are an improvement of the traditional Fingerprint method. Both methods are decomposed into two main phases. The first one is the **training phase**, and the second one is the **real-time phase**. These two phases are sometimes also referred to as the offline and online phases. In the training phase, WiFiPoz constructs an enhanced Radio Map based on a trained Radio Map using zone information and the Propagation method.

The programming language used in this research is C#. C# programs run on the Microsoft .NET framework. The .NET Framework provides a comprehensive and consistent programming model for building applications. It includes a set of standard class libraries which contain a large number of functions ready for use.

Definitions:

In order to describe the proposed algorithms, the following terms are principally defined. A **Mobile Station** refers to an unknown location of a mobile device user. A **Training Point** is a location in the testing area. In the offline phase RSS from Access Points at this point will be read, and coordinates at this point will be measured. A **Reference Point** is a point that responds its RSSI to assist locating the Mobile Station. **A Fingerprint** is the measurements (coordinates and RSS from Access Points) of a Reference Point. All fingerprints make a **Radio Map**.

ZID = an identification of a pre-defined zone.

BID = a pre-defined identification of a testing point, which is a Mobile Station.

RID = a pre-defined identification of a reference point.

$K = N_{neighbors}$ the number of reference nodes which close to blind node, and used for estimating the location of the blind node.

3.2 Flow chart of Indoor Positioning System - WiFiPoz

Figure 3-1 is the work flow chart of WiFiPoz. In the ***Offline phase (training phase),*** the signal strength of some predefined training points is measured and stored into a database, which is a Radio Map. Then parameters of predefined zones are calculated and stored into the database. Fingerprints on the grid of each zone can be computed using the Propagation method based on zone parameters and the original Radio Map. In this stage, an enhanced Radio Map is constructed.

In the ***online phase (real-time phase),*** the implementation is very similar to the traditional Fingerprint method with k-nearest neighbor. The first step is to measure the signal strength of all known networks in range. An adapted version of the Euclidean distance algorithm is employed to determine which reference points are the k-nearest neighbor in the signal. The positions of these reference points are then used to determine the mobile device and its geo-coordinates

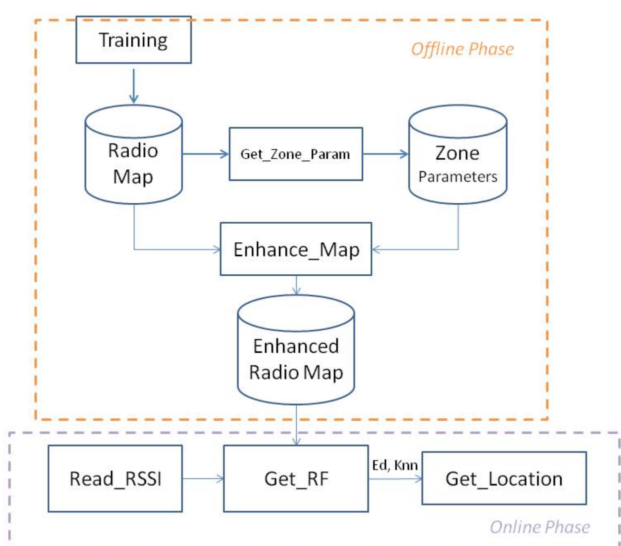


Figure 3-1: Work Flow of WiFiPoz

3.3 Training Phase

3.3.1 Table of Access Point Information

In the **WiFiPoz** positioning system, it is necessary to have a list of available Access Point. Access Point information can be saved in a table with the schema AP (ap_id, MAC, ap_x, ap_y, SSID). ap_id is a primary key (pk) in the AP table (Figure 3-2), it is associated with the unique MAC address. ap_x and ap_y, indicates the location of this Access Point. SSID refers to the name of this network.

AP
ap_id (pk)
MAC
ap_x
ap_y
SSID

Figure 3-2: AP Table

A primary key (pk) in a table uniquely identifies each record in the table. A foreign key (fk) is a referential constraint between two tables.

3.3.2 Fingerprint Capturing

3.3.2.1 Signal Strength Reading

The first crucial part in the training phase is to access the wireless hardware and obtain access point information. What's needed is a way to retrieve a list of access points within range of the mobile device and read the signal strength.

To capture the signal strength, the Microsoft dynamic link library (DLL) WlanAPI.dll is used. WlanAPI.dll contains pre-compiled implementation code for Native Wi-Fi API functions. An API (Application Programming Interface) is source code based specification intended to be used as an interface by software components to communicate with each other, so that programmers can build applications consistent with the operating environment [47]. Native Wi-Fi API has functions, structures, and enumerations that support wireless network connectivity and wireless profile management [48].

In Windows 7, wlanapi.dll is in the system's folder. The functions and enumerations in wlanapi.dll are defined and managed through platform invocation services (p/invoke) interop ([DllImport ("wlanapi.dll")]) . P/Invoke allows managed code written in c# to call unmanaged functions that are implemented in DLL. A managed wrapper WlanApi.cs is built on the result from P/Invoke and contains functions in Native Wi-Fi API.

The Native API provides a lot of functions for querying various parameters of wireless networks. *WlanBssEntry* is the structure that contains information about a basic service set. *dot11Ssid* represent the SSID of the access point. The service set identifier SSID in most cases is a human readable name for the Wi-Fi. The SSID is not unique - it just an easier way for persons to identify a network from a list.

dot11Bssid is the media access (MAC) address of the access point for a network. It is represented by an 8 bytes identifier, which is assumed to be unique. *rssi* is the received signal strength in dBm. (A signal strength of s Watts is equivalent to $10*\log_{10}(s/0.001)$). A program built on this structure reads the RSSI from visible access point with corresponding MAC address.

With the presence of noise, it has been observed that (1) The RSSI value of the access points measured at the mobile station is not stable. However, it is distributed within a small interval. (2) The number of access points covering a location may vary with time. As a result, single measurements at each reference point are not sufficient. Using a mean of several measurements captured at one Reference Point can reduce the effect of environment impacts [49].

The orientation of the device also has a significant impact on the RSSI received by the device. Measurements for different orientations at each Reference Point are able to improve the accuracy of the system.

3.3.2.2 Store Information of Training Points

The union of all training phase fingerprints and their corresponding location resembles the Radio Map which is stored by a database. Fingerprints describe the signal characteristics of every Access Point receivable at a specific Training Point. All of the RSSI measurements collected during the offline phase are stored into a single, unified table named *t_measurement* containing tuples of the form: t_measurement (ap_id, P_id, y, mean_ss, median_ss, geo_d).

Where:

- ap_id: The ID of certain Access Point
- p_id: ID of this training point (p_id)
- mean_ss: average RSSI of this Access Point at this training point
- median_ss: median RSSI of this Access Point at this training point
- geo_d: geometry distance between the training point and Access Point

The location information for each training point is stored in a table with the schema t_point (p_id, p_x, p_y, zone_id), p_x and p_y indicates the coordinate of the training point. zone_id refers to the zone for this training point. As **Error! Reference source not found.Error! Reference source not found.Error! Reference source not found.**Figure 3-3 shows, the table *t_measurement* and *t_point* is linked by the p_id.

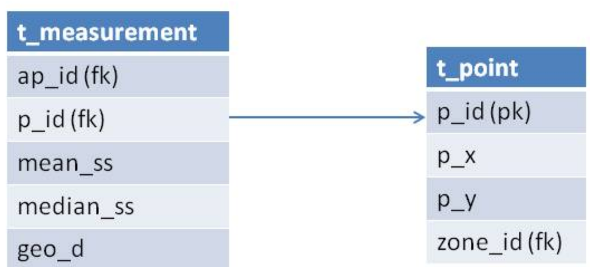


Figure 3-3: Radio Map

3.3.3 Zoning Approach

3.3.3.1 Radio Propagation Model and the Wall Attenuation Factor

The model represents a common indoor characteristic encountered by a positioning system. Attenuation due to an intervening wall or partition is described by Equation 1:

Equation 1: Radio Propagation Model

$$\mathbf{P(d)[dBm] = P(d_0)[dBm] - 10\ n\log\left(\frac{d}{d_0}\right) - WAF} \qquad \mathbf{[14]}$$

Where n is a path loss exponent and *WAF* is a wall attenuation factor. The path loss exponent (n) indicates the rate at which the path loss increases with distance (d). Overall, *WAF* is the attenuation loss from intervening walls and partitions. In buildings, obstruction varies from place to place.

$P(d_0)$ is the RSSI at some reference distance d_0. We usually pick up the reference at 1 meter where RSSI is -45dBm. This result is gained by measurement in the experiment. Then, from Equation 1: **Radio Propagation Model**

$$\mathbf{P(d)[dBm] = P(d0)[dBm]} - 10\, n\log\left(\frac{d}{d0}\right) - \mathbf{WAF}$$

The received signal power P(d) in a Training Points that the Propagation model and can be described as follows:

$$\mathbf{P(d) = -10n\log d - (WAF + 45);}$$

Assume in a zone that all Training Points in it have the same n and *WAF*.

Make:

$$\mathbf{b_0 = -10n, \text{ and } b1 = -(WAF + 45);}$$

Then we get:

Equation 2: Simplified Radio Propagation model

$$\mathbf{P(d) = b1 * \log d + b0}$$

The equation shows that the relationship between P(d) and logd is linear. It is possible to get b_1 and b_0 using linear regression.

3.3.3.2 Zone Divided Method

Although the Propagation model takes the environment into account, the indoor environment is very complicated. On one hand, the Path Loss Exponent (n) and Wall Attenuation Factor (WAF) depend on the building layout and construction material. It is very difficult for us to get the correct measurements for them. On the other hand, because of the Multipath Effect, measuring a direct line between transmitter and receiver will lead to an underestimation in many cases.

The zoning approach is proposed by D. L. Lee and Q. Chen [20]. Each room or corridor is considered to be a zone in the area. A corridor can be represented by more than one zone according to the need of applications. As Figure 3-4: Zoning Approach shows, each zone looks like a bubble, which means there is no wall or other building inside of a zone. For all the points inside a zone, they have very similar *n* and *WAF* for each base station (Access Point). With the same parameters in a zone, the location estimation has higher accuracy when applying the Propagation model, even though there is a very complicated floor plan with diverse building materials.

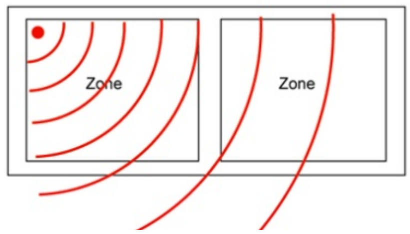


Figure 3-4: Zoning Approach

The boundaries of each zone should be saved into the zone table in the form:

(zone_id, min_x, min_y, max_x, max_y).

3.3.3.3 Zone parameter Estimates using Linear Regression

From 3.2.2.1, we get the linear equation for a Base Station in a zone:

$$\mathbf{P(d) = b1 * \log d + b_0}$$

d represents the distance between a Training Point and an Access Point. P(d) represents the RSSI at that Training Point.

Simplified, the equation above becomes:

$$\mathbf{y = b0 + b1 * x;}$$

Where, x = logd and y = P(d);

To set several training points in a zone, for each Training Point we measure the RSSI which is y_i, and logd which is x_i. With several pairs of x_i and y_i, according to the Liner regression, b_0 and b_1 can be calculated using Equation 3:

Equation 3: Linear Regression Formula [50]

$$b_1 = \frac{\sum xy - n\bar{x}\,\bar{y}}{\sum x^2 - n\bar{x}^2} \quad \text{and} \quad b_0 = \bar{y} - b_1\bar{x}$$

Where:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$\sum xy = \sum_{i=1}^{n} x_i y_i \qquad \sum x^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

Combining with the zone boundary, the schema for the zone table is (zone_id, b0, b1, min_x, min_y, max_x, max_y)

3.3.4 Enhanced Radio Map

3.3.4.1 Reasons for the Enhanced Radio Map

The number of Reference Points is an important factor for a location estimate; it has a direct impact on the accuracy. The more Reference Points added into a certain area, the better the result the system will get. But that would increase the effort on the fingerprint capture in the training phase. Another disadvantage is the fact that the database needs to be updated every time when something in the network or environment change, For example, removal of an Access Point or rearrangement of the physical environment would make previously captured reference points inaccurate or completely useless. If there is a big number of Reference Points, the maintenance work will be very difficult.

The general concept of the WiFiPoz positioning system is illustrated in Figure 3-5: Enhanced Radio Map. The black dots indicate the Training Points where we collect data and record the RSSI values. The triangle represents the location of a Mobile Station, in the real-time phase. If we estimate the location of Mobile Station based on just two training points the result will be prone to low accuracy. But if we project the fingerprints onto a grid, then we get a larger number of reference points and the result will be much better. Plus, additional fingerprints are computed by the system based on the Propagation model, so no additional human effort is required.

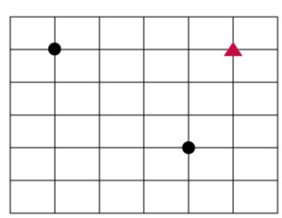


Figure 3-5: Enhanced Radio Map (The black dots indicate the Training Point and triangle indicate the Mobile Station)

3.3.4.2 Constructing the Enhanced Radio Map

The first step is to convert a zone into an (m * n) grid. The grid spacing defines the distance between a pair of points next to each other. A 0.5-1 meter grid spacing will be sufficient. This distance was selected because, if two Reference Points are too close to each other, their fingerprints will not differ enough to distinguish between them.

The coordinates of a point in the intersection of the grid can be recorded as RID(x, y). The distance from this RID to a base station is marked as d. From the Equation 2,

$$\mathbf{P(d) = b_1 * \log d + b_0}$$

b_0 and b_1 are the pre-stored zone parameter, and logd can also be gained easily. Fingerprint in this point (RSS_1, RSS_2, …,RSS_n) can be obtained by calculation.

Repeating the steps above, a new Radio Map is constructed with two tables *en_measurements (ap_id, p_id, mean_ss)* and *t_point (p_id, p_x, p_y, zone_id)*. It has more Reference Points within a zone, as shown in Figure 3-6.

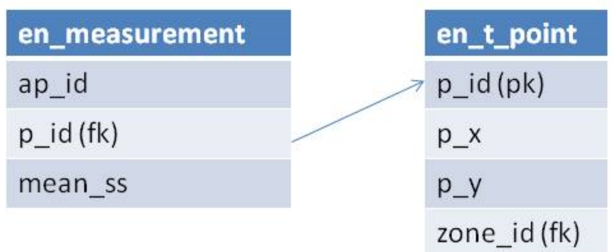


Figure 3-6: Enhanced Radio Map

3.4 Real-time Phase

The unknown location of a Mobile Station can be estimated based on the Enhanced Radio Map from 3.3.4.2 using the adapted version of the Euclidean distance algorithm and k-nearest neighbor algorithm.

3.4.1 Mathematic Background

3.4.1.1 The Distance Algorithm

The Euclidean distance algorithm measures the simple distance between points in an n-dimensional space given by their respective coordinates. It is calculated by summing up the absolute distance for each dimension. The Euclidean distance between two points x and y is given by:

$d_E (x, y) =$

$=$

A variation of the Euclidean Distance is used to determine a distance value between the mobile station and a reference point. Equation 4 is for the Euclidean distance and interprets each network as a unique dimension, resulting in an N dimension space captured in the scan.

The distance is calculated by summing up all absolute distance values where the distance values are calculated for the network.

Equation 4: Euclidean Distance for the Network

$d_E (RID, BID) =$

d_E (RID, BID) indicates the Euclidean distance between a Mobile Station and an Reference Point. n is the ID number of the Access Point. RID_SS_i represents the RSSI from base station i, at the reference point. BID_SS_i represents the RSSI for Access Point i, at the mobile station.

3.4.1.2 K-nearest Neighbor Algorithm

In pattern recognition, the k-nearest neighbor algorithm (KNN) is a method for classifying objects based on the closest training examples. An object is classified by a majority vote of its neighbors which are found within a certain calculated distance. Figure 3-7 is an example from Wikipedia [47].

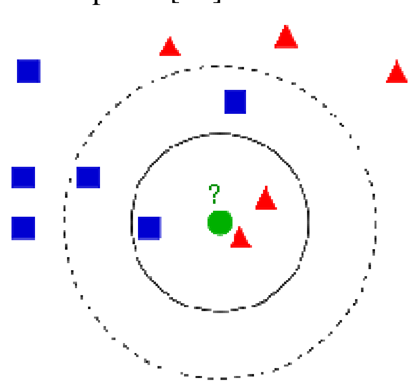


Figure 3-7: Defination of kNN algorithm

"The test sample circle should be classified either to the first class of squares or to the second class of triangles. If k = 3 it is assigned to the second class because

there are 2 triangles and only 1 square inside the inner circle. If k = 5 it is assigned to the first class (3 squares vs. 2 triangles inside the outer circle)".

3.4.2 Mobile Station Location

The adapted version of the Euclidean Distance algorithm is employed to get a list of Euclidean distances between the Mobile Station and Reference Points in an Enhanced Radio Map.

Searching through the list of Euclidean distances between the Mobile Stations and Reference Points, we can find the k closest Reference Points, those Reference Points are the k-nearest neighbor. We mark them as $N_{neighbor.}$ There are a fair amount of search algorithms to complete this task. We use the simple way to find out the $N_{neighbor}$ which does not warrant the complexity. This thesis research is focused on the analysis rather than on developing an optimal closest match search algorithm.

The adapted k-nearest neighbor algorithm is used to estimate the possible location of an unknown Mobile Station position. We use k-nearest neighbor, because, sometimes there might be several Reference Points that have the same distance to the Mobile Station. As Figure 3-8 shows, RP1, RP2 and RP3 are the k-nearest neighbor to the Mobile Station. If we simply assume the location of RP1, which has the shortest distance to Mobile Station, to be the location of the Mobile Station, then the error distance will increase. Averaging over the location of RP1, RP2 and RP3 we can get a better estimation of Mobile Station's location.

In summary, we find the k references from the Enhanced map and average their geo-coordinates, then consider the result to be the location of the Mobile Station.

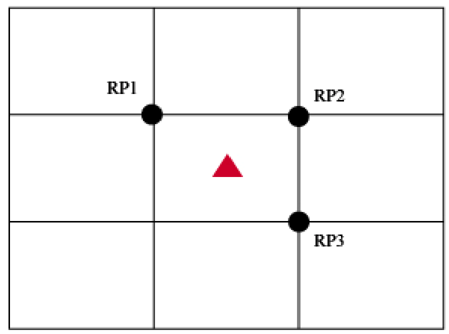


Figure 3-8: Benefit of k-nearest neighbor algorithm

3.5 Geometric Information Based Positioning System (GID)

3.5.1 Overview

Based on the concept proposed by T. Parthornratt and K. Techakittiroj [43], this thesis implemented Geometric Information based Positioning System (GIS). Their lateration method adopted to locate the Mobile Station. There are two reasons to develop and implement the GIS method. First, I wanted to explore other solutions based on the zoning approach combined with the Propagation model. Second, comparison with the IS method by experimentation will demonstrate that **WiFiPoz** is the most efficient solution to improve accuracy.

Figure 3-9 is the work flow chart of GIS. In the **training phase**, the signal strength of some predefined training points is measured and constructs a Radio Map. Then, parameters of predefined zones are calculated and stored into the database. These two steps are similar to WiFiPoz.

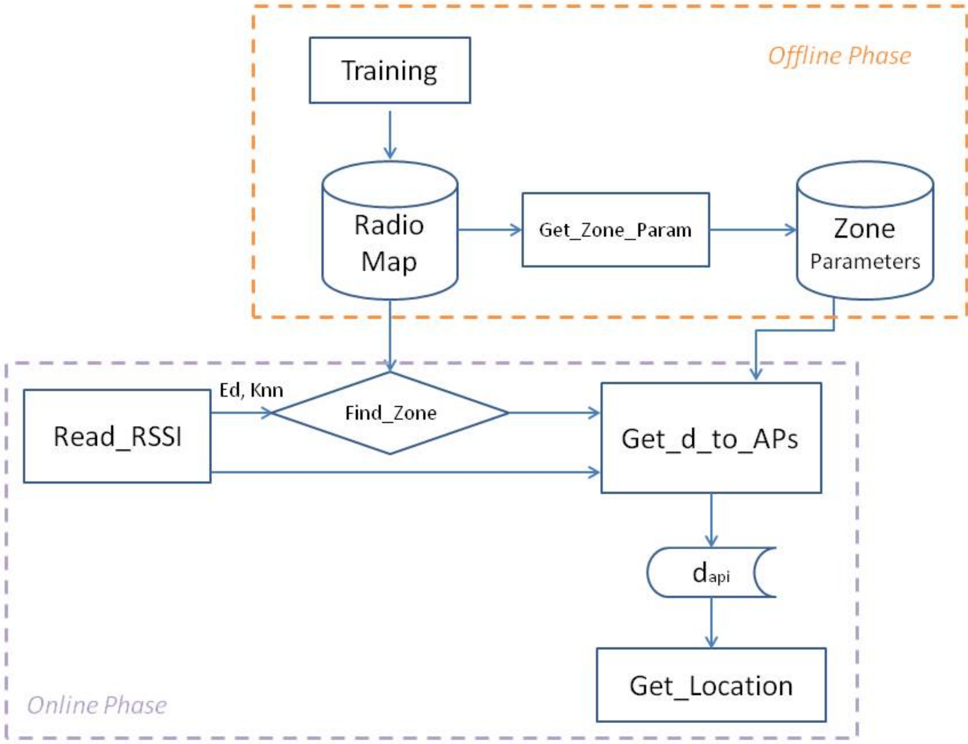


Figure 3-9: Work Flow of GIS

In the **real-time phase**, there are 3 steps to determine the unknown location of the mobile station:

1. Euclidean distance is combined with the k-nearest neighbor algorithm to determine that the Mobile Station belongs to which zone.
2. Using the Propagation model associated with the information in the *zone* table to estimate the geometric distance between an Mobile Station and each Access Point (d_1, d_2, …, d_n).
3. Using the Lateration algorithm to find out the position of this Mobile Station based on the distances calculated in step 2.

3.5.1.1 Zone Determination

The Euclidean Distance algorithm was implemented to get a list of distances (d_list) between the mobile station and references. Then an adapted k-nearest neighbor algorithm was used to find the matched ZID.

In each zone, n fingerprints are stored in the database.

Let, $1 \leq k \leq n$

Then, select k reference points which correspond to the k shortest distance values in the d_list.

Query the zone table to get the ZIDs for those k reference points.

We assume the mobile station is in the ZID with highest vote from k reference points.

3.5.1.2 Distance Estimate

From the Equation 2 introduced in 3.2.2.1:

$$P(d) = b_1 * \log d + b_0$$

We get:

$$d = 10^{(P(d) - b0)/b1}$$

p(d) is the detected RSSI. b_0 and b_1 can be read from the zone table and are usually negative numbers. The value of RSSI is related to the distance between the base station and mobile station. A lower RSSI indicates a longer distance.

3.5.1.3 Location Estimate

We got the estimated values of distance from the Mobile Station to each base station as stated in the last section. There is the fact that the distribution of radio signals is not ideal, and it is affected by a larger number of environmental variables, the temperature, furniture and people. As a result, the values of distance we get have some level of error and might not be very precise. It has some level of error. The problem for positioning systems based on triangulation (lateration) of mobile devices is that the accuracy of this distance value is a key component for the overall accuracy of systems implementing triangulation; so the technique is prone to be inaccurate.

Figure 3-10 shows the ideal condition, which means the distance from Mobile Station (x,y) to each Access Point(x_i, y_i) is very precise, i is the ap_id. Then we can get equations as below.

$$(x - x_1)^2 + (y - y_1)^2 - r_1 = 0$$

$$(x - x_2)^2 + (y - y_2)^2 - r_2 = 0$$

…

$(x - x_n)^2 + (y - y_n)^2 - r_{n} = 0$

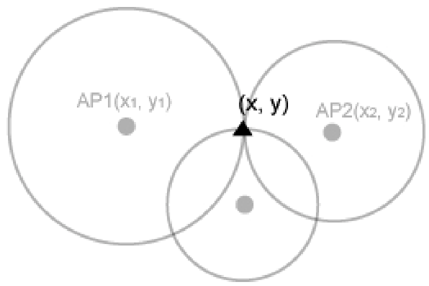


Figure 3-10: Ideal Model for Triangulation

Where x_1 and y_1 are x-y coordinates of AP_1, while x_2 and y_2 are x-y coordinates of AP_2, r1 and r2 are the values of the distance calculated from 3.5.1.2.

But the estimated values of distance always has error, as Figure 3-11 shows, there are many possible conditions as mentioned previously that introduce errors into the distance measurements.

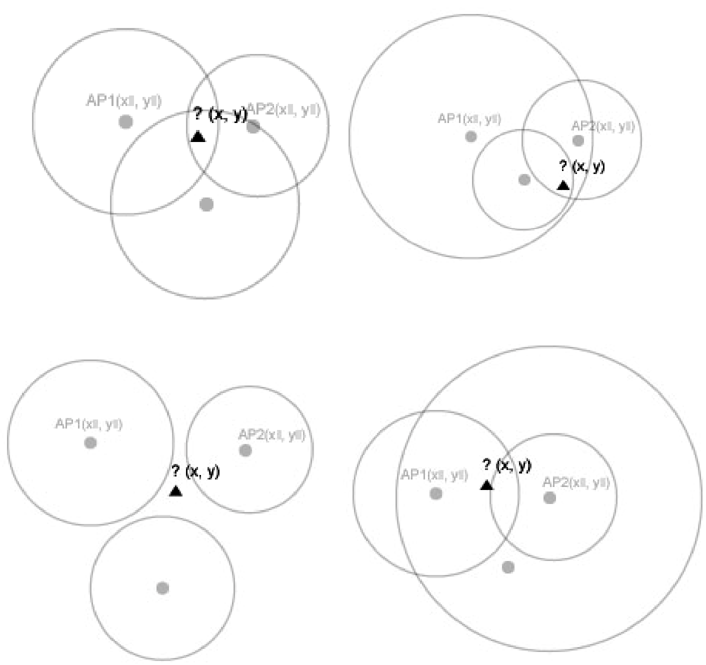


Figure 3-11: Distance Estimation with Error

$$f_n(x, y) = (x - x_n)^2 + (y - y_n)^2 - r_n$$

$$\|\bar{f}(x, y)\| = \sqrt{f_1(x,y)^2 + f_2(x,y)^2 + \ldots + f_n(x,y)^2} \quad [21]$$

$\|\bar{f}(x, y)\|$ represents the error matrix due to inaccurate value of distances. Zone boundaries for both horizontal and vertical dimension are taken to be the constraints for finding the min$\|\bar{f}(x, y)\|$.

Finally we got a point with geo-coordinate (x, y) which has the min $\|\bar{f}(x, y)\|$, this point is consider to be the location of Mobile Station.

3.6 Summary

This chapter provides a framework for theoretical explanation and insights about the proposed indoor positioning system using existing Wi-Fi infrastructure. The original positioning system model proposed by the author is WiFiPoz, which is developed based on the traditional Fingerprint method. The basic concept of WiFiPoz is to divide the indoor environment into several zones. In the training phase, the signal strength is read from each Access Point at several training points, all the information is saved into a database which is a Radio Map. The widely used Propagation model is then used to construct an Enhanced Radio Map. On the real-time phase, the characteristic of signal strength reading is compared to the Enhanced Radio Map to find out the location of the Mobile Station.

To explore an alternative solution, based on the fingerprinting and zoning approach, another positioning system GIS is also developed. The difference between WiFiPoz and GIS is this: instead of constructing an Enhanced Radio Map, a lateration method is applied to calculate the location of the Mobile Station.

The purpose of developing WiFiPoz and GIS is to improve the accuracy of an indoor positioning system without increasing the number of training points.

Chapter 4: Experiment and Results

4.1 Test Area

The location selected for the experiments is the south part of the third floor of the Computing and Engineering Building at Eastern Washington University (Figure 4-1). The testing area is viewed as a two-dimensional model. The testing area is approximately 43 meters by 17 meters.

Figure 4-1: Third floor of the Computer Science Building with AP's Locations

Reference points or testing points inside the testing area are presented by their coordinates on a local two-dimensional Cartesian coordinate system. The coordinate system used for this experiment is set up as in Figure 4-1 (numbers are in centimeters), with the origination point (0, 0) in the top left corner. The introduction of a new coordinate system allows new reference points or testing points to be easily added without the need to calculate the actual global geo-coordinates for them in the system. Positions of locations are described by their x and y coordinates relative to the origination point.

The experiment was conducted during the weekend and late afternoon, so the human traffic within this area was not very dense. In this area, there are many labs and offices.

The way finding requirements more often take place in the walkway, so the research focuses on the corridor.

4.2 Equipment and Tools

The equipment and tools, including hardware and software, selected for recording and measuring is off the shelf and affordable.

4.2.1 Access Points

This research is about using existing infrastructure. There are 4 Access Points installed in the testing area, as Figure 4-2 shows. The testing area has good Wi-Fi signal coverage. Figure 4-1 shows the layout of 4 Access Points in the testing area. Access Points are attached on the ceiling above the cable tray.

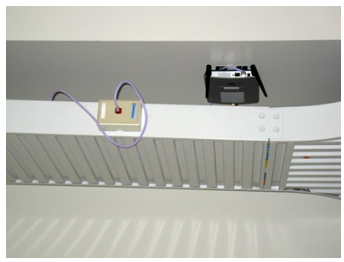


Figure 4-2: Location of AP2

4.2.2 Laptop and Wireless Card

The laptop used for the experiment is Lenovo ThinkPad X200 with the Intel Wi-Fi Link 5300 (802.11 a/b/g/n) Ethernet Card. The pre-installed Operating system is Win 7. (Figure 4-3)

Figure 4-3: Laptop used for Experiment (Lenovo ThinkPad X200)

4.2.3 Measurement Tool

A Bosch DLR 130 K digital laser measuring tool was used (Figure 4-4), which is accurate to within 1/16 inch, 130- feet range. Distance measured by this tool is easy and fast with fairly high accuracy. The coordinates of each fingerprint and test points were measured by this digital laser measurement tool.

Figure 4-4: Bosch DLR 130K Digital Laser Measuring Tool

4.2.4 Software

The software used for this project includes:

- Windows 7
- Visual studio 2010
- SQLite
- Windows Office 2007

4.3 Training Phrase

4.3.1 Access Points used for this Experiment

As Figure 4-1 shows, there are 4 Access Points detected in the testing area. The MAC address, SSID and the location of each Access Point are stored in a table (Table 4-1), ready for later use.

Table 4-1: Profiles about Access Points used for the Experiment

ap_id	MAC Address	Position(x, y)	SSID
1	00-13-19-31-64-00	(70, 140)	Wi-EWUSouth
2	00-13-19-31-5D-30	(1655, 1217)	Wi-EWUSouth
3	00-13-19-31-5E-60	(2936, 140)	Wi-EWUSouth
4	00-13-19-31-5D-20	(4090, 1132)	Wi-EWUSouth

4.3.2 Zone Divided

The testing area is divided into 7 zones (Figure 4-5). Three basic rules are used in dividing up the space into zones:

- There is no wall or building construction obstacles between any two points within the zone.
- Each zone can receive a very strong signal from at least one Access Point. "Strong" means the RSSI is higher than -50dBm.
- Zones are of equal size.

The boundaries of each zone are recorded in a table, ready for later use.
Zone_boundaries (zone_id, min_x, min_y, max_x, max_y). The list of zone boundaries is included in Appendix A

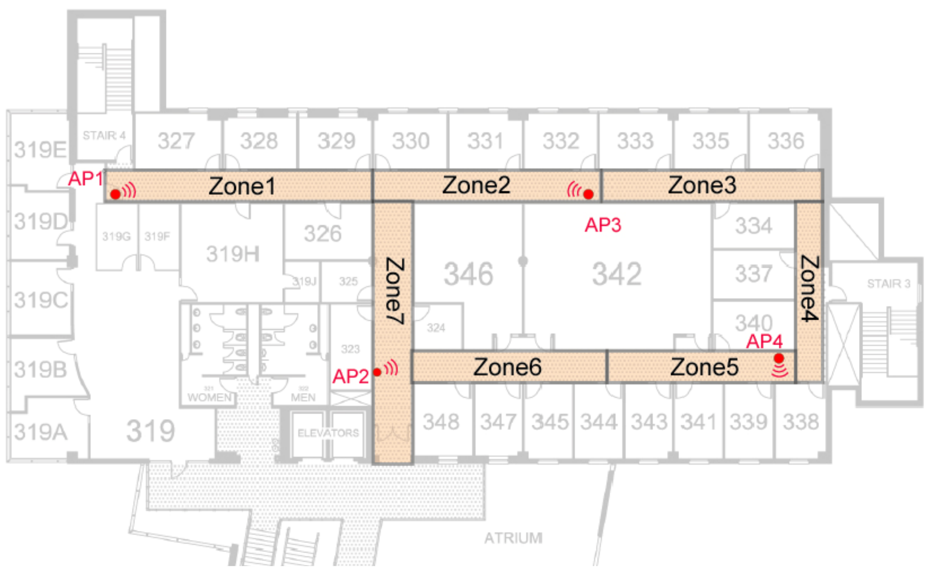


Figure 4-5: Zone Divided

4.3.3 Training Points Setup

A total of 35 training points are set up for this experiment. Each zone has 5 testing points. These 5 points are randomly distributed and cover the whole zone. The distance between two points is about 2.5 – 3.5 meters. The training points are numbered from 1 through 35 (Figure 4-6). The coordinate measurement of 35 points is listed in Appendix B. Generally, the training points are numbered as the user walks along the hallway. Other settings for this research will be explained later in this chapter.

Figure 4-6: The Distribution of 35 Training Points

4.3.4 Received Signal Strength Capture

4.3.4.1 Overview

For the offline phase, two applications are developed for capturing data and refining the data records into the Radio Map. To capture the Received Signal Strength (RSS) from each Access Point, an application called *Training* is developed in C#. It is used to capture data describing the signal strength of 4 nearby Access Points.

The *Training* application supports the following steps:

- Scan for wireless networks in range
- Display a list of all found Access Points to be used for position estimate, if any of the network is missing, display "scan failed" and scan for wireless networks again
- Run periodic scans where data is collected every second
- Save the captured data to the database.

4.3.4.2 Native Wi-Fi API

Microsoft provides a Native Wi-Fi API (Application Programming Interface) for developing wireless applications. Native Wi-Fi API has functions, structures and enumerations that support wireless network connectivity and wireless profile management. In Win 7, Wlanapi.dll, which contains Wi-Fi APIs, is located in the system folder. The class *Wlan* defines the Native Wi-Fi API in C# through P/Invokes. It manages codes in *wlanp.dll*.

The class *WlanClient* is the entry point to native Wi-Fi management. It is built on top of the class *Wlan*. The class *WlanClient* is a wrapper that integrates all enumerations structures and function into the Native Wifi API that manages the Wifi setting.

In the *Training* class, a *WlanClient* instance is created. The function *DoScan*() invokes the functions of native Wi-Fi management, scans the available networks and prints the information.

Within *DoScan(),* the function *GetNetworkBssList()* retrieves the basic service set (BBS) list of all available Access Points. The structure named WLAN_BSS_ENTRY contains information about a BSS. Some members in this structure can be used for fingerprint capturing, including:

- dot11Ssid: It indicates the SSID of the Access Point. SSID is short for Service Set Identifier, which is referred to as a network name. The SSID differentiates one WLAN form another, so in the specific WLAN all access points use the same name. It returns a byte array, with a length up to 32. SSID can't be used to identify each Access Point.
- dot11Bssid: The media access control (MAC) address of the Access Point. Each Access Point has a unique MAC address, which can be used to identify Access Points.
- rssi: The received signal strength indicator value

The function DoScan() scans the visible networks at the current position of device. MAC addresses from all available Access Points are compared to the list of pre-selected Access Points (AP_1, AP_2, AP_3, and AP_4) for this experiment. The raw data describing the signal characteristics at pre-selected Access Points were printed and recorded.

4.3.4.3 Number of Samples

An important finding regarding characteristics of the signal strength is that the network signal is not constant over time, even over a short time interval. In the worst case, the signal strength is not receivable in some of the scan responses. The fluctuations of the signal strength is due to the fact that the signal is weakened through various effects inflicted by the environment as absorption or refraction, weakening the signal to a level at which it cannot be clearly distinguished from the signal noise. According to information gathered from the various test runs conducted during this thesis project, signals of networks with poor reception, such as Access Points that are located far away from the mobile station, were significantly different between each scan. The averaged RSS from several scans minimized the effects of these fluctuations, resulting in much more stable signal characteristics, as compared to one scan or the median of several scans.

As pointed out in Chapter 2, the orientation of the device has a significant impact on the signal strength distribution of Wi-Fi signals due to the blocking effect of the human body. In test runs, the orientation of the device did affect the result, but not very significantly. Taking the user orientation into consideration will make the fingerprint database very complicated. Simply including at least four orientations in the scans and averaging them seems to provide an acceptable result.

It is quite impractical to require a device to measure the receivable signal strength over several minutes during the real-time phase at a fixed position. Only a few scans may be conducted for a fingerprint training or positioning estimation, because the person carrying the device could move after a short time period.

Figure 4-7 is the chart for averaged signal strength at training point No. 5, varied by scanning times. The time between each scan is one second; each set of scans cover at least 4 orientations (north, south, east, and west). The averaged result from 200 scans is not much different from the averaged result from 100 scans, 50 scans, 20 scans and 10 scans. The result from 5 scans decreases the accuracy dramatically (more than 5 dBm). Repeating the same testing on other training points, we can have very similar findings.

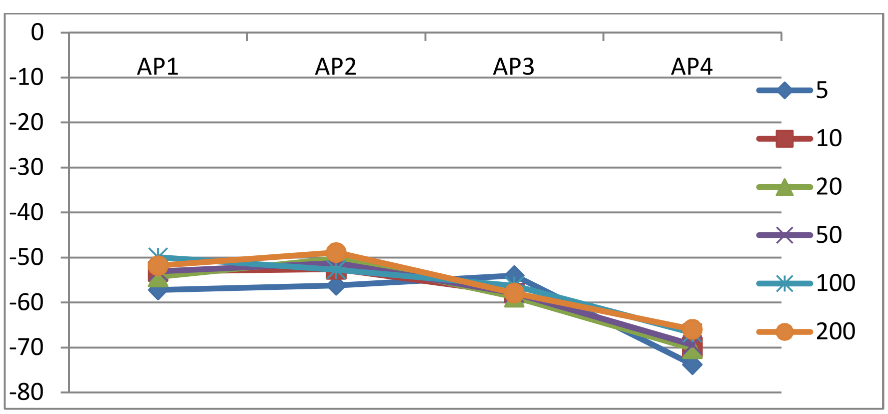


Figure 4-7: Averaged Signal Strength at Training Point No.5

As a result, in the RSS measurement at each of the 4 Access Points, used for location estimation at each training point, is the mean of 10 scans from different user orientations at that location. The time between 2 scans is one second.

4.3.4.4 Interface of Training Application

Figure 4-8 shows the interface of the scanner application *Training*. When RSSI for all 4 Access Points are captured, it is counted as one scan and displayed on the list box. If signal dropout occurs, which means missing RSSI reading from one or more Access Points used for this experiment, it displays “scan fail” and re-scan. After 10 successful runs, the mean, median, min and max will be calculated and printed on the list box. Training point number and coordinate (x, y), requires manual input. At the end, the characteristics of this physical training point are logged into a database, which will be explained later in this section.

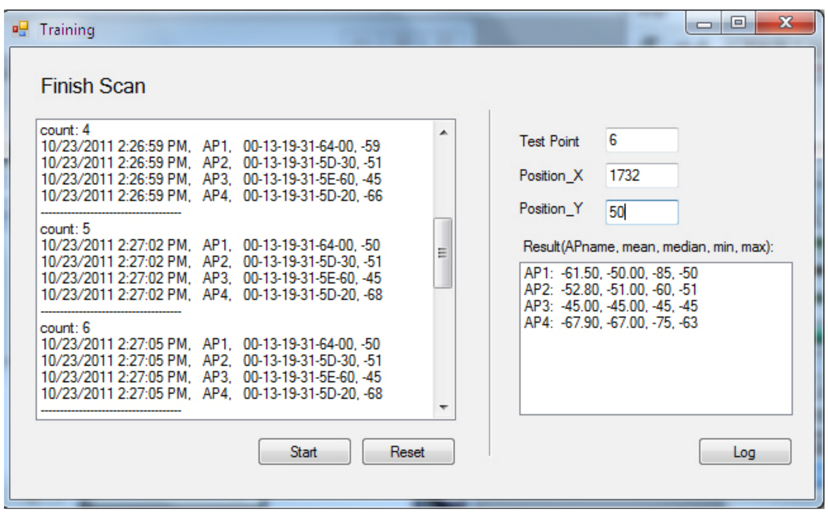


Figure 4-8: Training Application Interface

4.3.5 Geometric Distance between Training Points and each Access Point

The geometric distance between a given Training Point TP (x_{tp}, y_{tp}) and an Access Point AP (x_{ap}, y_{ap}) is calculated using the following formula:

$$d = \sqrt{\left(x_{tp} - x_{tp}\right)^2 + (y_{ry} - y_{rp})^2}$$

The distance from the 35 training points to each Access Point is shown in Figure 4-9. They are stored for later use. The method GetGeoDistance() implements this function and saves result into the table *t_point*.

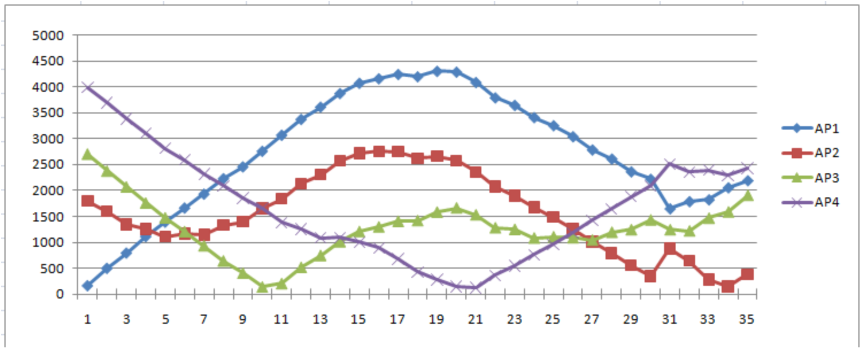


Figure 4-9: Distances from Training Point to Each Access Point

4.3.6 Zone Parameter

The application called *processing* is developed to compute and store the zone parameters.

Inside the testing area, the building structure is complicated as the material used for walls varies. On one hand, simply applying the Propagation model to the whole area would end up with an unacceptable result. Figure 4-10 shows the relationship between the actual distance and signal strength for AP2. For many training points, they have nearly identical distance measurements from AP2, but the signal strength varies a great deal from -50 dBm to -75 dBm. On the other hand, finding out the WAF and n for each wall is not practical.

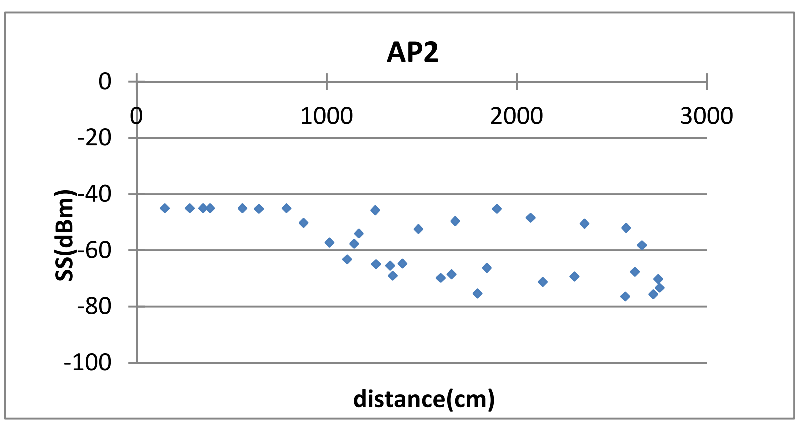


Figure 4-10:Distance vs RSSI of AP2, for 35 Training Points

As we discussed in 3.2.2.1, applying the Propagation model (Equation 1) to each zone may be a better solution. Within a zone, the wall attenuation factor (WAF) and path loss exponent (n) show little difference between two training points.

Simplifying the Propagation model into Equation 2, there should be a linear relationship between logd and p(d). Figure 4-11 shows the relationship between signal strength and logd at the five training points in Zone 2 for all four Access Points. The x_axis indicates the logd, where d here is the distance between the mobile device and the access point. The y-axis indicates the signal strength in dBm. The lines represent calculated results by applying a linear regression for each Access Point. R_2 is the coefficient of determination. AP_i (b0, b1, R_2) , where $i \in \{1, 2, 3, 4\}$, then will be saved as s parameters for this zone.

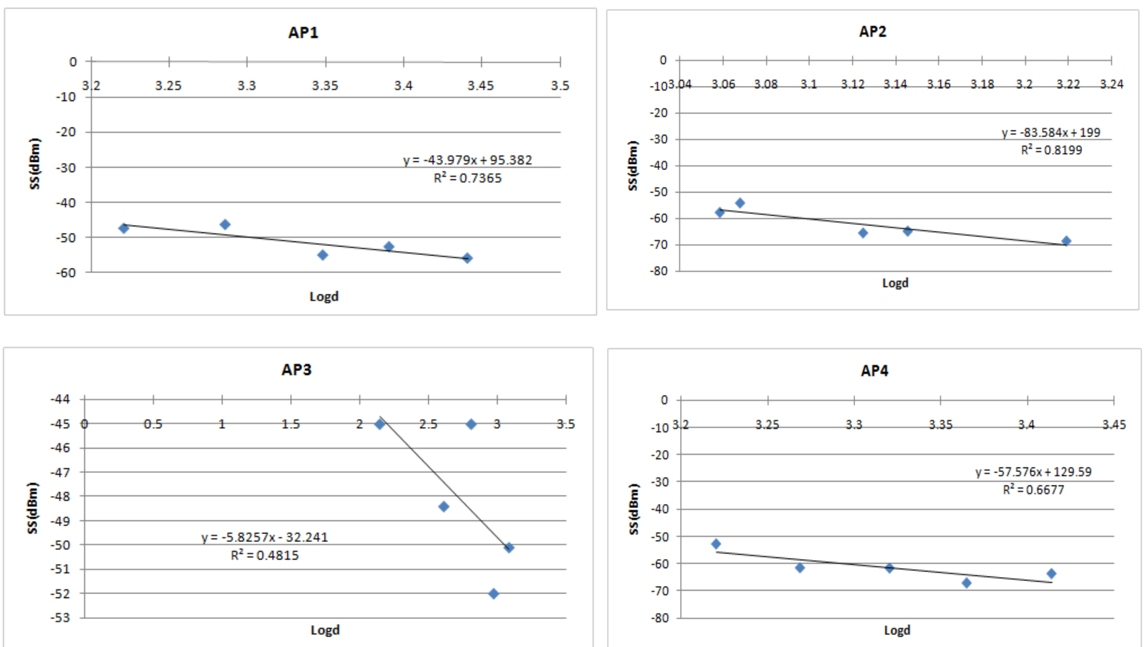


Figure 4-11: logd vs Received Signal Strength of each Access Point at Zone 2

The function GetZoneParameters() calculates and store parameters for each zone. A list of all the parameters for each zone is listed at appendix C. 50% of R_2 are larger than 0.6.

4.3.7 Radio Map

The fingerprints for all training points are saved in SQLite. SQLite is a lightweight database management system based on text files and is very easy to plug into a C# application. Plus, it is integrated into many mobile OSs.

For this application, 5 tables have been defined within the SQLite database system (Figure 4-12):

AP (ap_id, MAC, ap_x, ap_y, SSID)

t_point (p_id, p_x, p_y, zone_id)

t_measurement (ap_id, p_id, mean_ss, median_ss, geo_d)

Zone_boundary (zone_id, min_x, min_y, max_x, max_y)

Zone_param(ap_id, zone_id, b0, b1, r2)

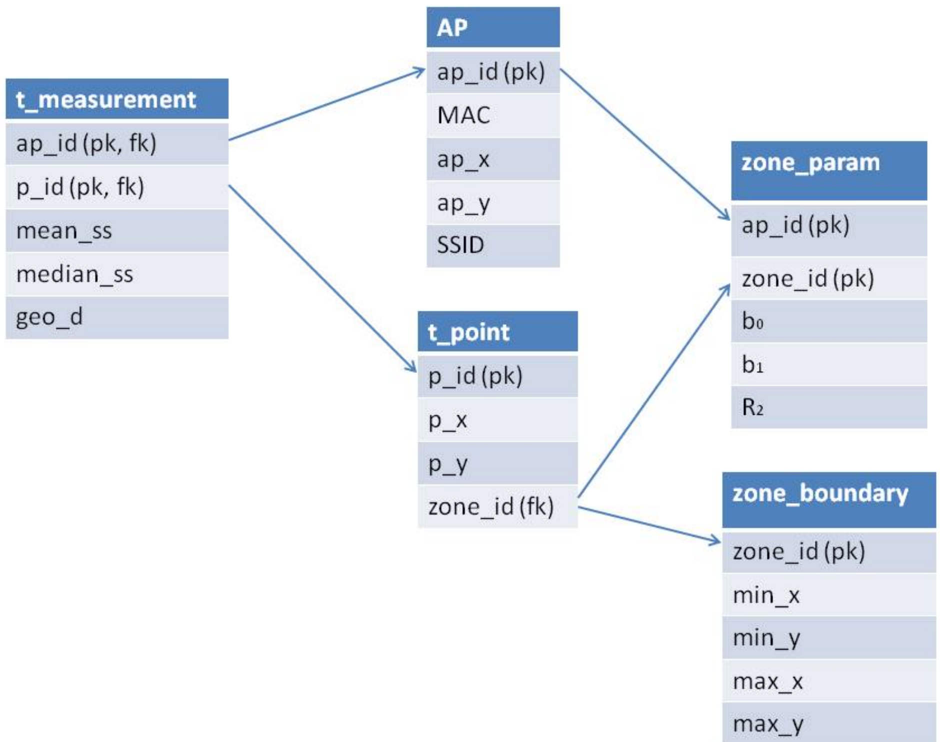


Figure 4-12:Tables Used for WiFiPoz and GIS

The table *AP* stores the Access Point information, including the id for each Access Point (ap_id), the MAC address, Access Point coordinates (ap_x, ap_y) and the network name (SSID). The table *t_point* stores a list of all training points with their respective local coordinates. p_id is the id number of a training point. (p_x, p_y) are the coordinates of a training point. zone_id is the id number of the zone in which the training point resides. Table *t_measurement* stores all the fingerprint measurements for all training points with the primary key (ap_id, p_id). A primary key (pk) in a table uniquely identifies each record in the table. A foreign key (fk) is a referential constraint between two tables. It matches the primary key in another table. ap_id is foreign key matches the primary key in table *AP*. p_id is a foreign key that matches the primary key of table *t_point*. *zone* table stores the zone id and boundaries for each zone (min_x, min_y, max_x, max_y). The zone parameters (b0, b1, R2) are computed and stored in the table *zone*. Foreign key *zone_id* in *t_point* links table *zone*.

For WiFiPoz, the Radio Map is updated using zone parameters. The Enhanced Radio Map consists of two tables *en_measurement* and *en_t_point*, which will be used in the real-time phase.

en_measurement (ap_id, p_id, mean_ss)

en_t_point (p_id, p_x, p_y)

The reference points are extended from training points. Each point with p_id is the intersection of a grid within a zone. The mean_ss is computed by Equation 2:

$P(d) = b_1 * \log d + b_0$. P(d) is stored as mean_ss and then stored in the *en_measurement* table.

4.4 Real-time Phase

4.4.1 Testing Points

28 testing points are set up in the experimental area, as Figure 4-13 shows. There are 4 testing points within a zone. WiFiPoz and GIS are applied to estimate the position of each testing point.

Figure 4-13: 28 Testing Points

4.4.2 Real-time Phase of GIS

For GIS, there are 3 steps to estimate the position of the mobile station.

1. Determine the zone in which the Mobile Station resides.
2. Get the geometric distance between the Mobile Station and each Access Point.
3. Estimate the most probable position of the Mobile Station according to the coordinates.

4.4.2.1 Zone Determination

The k-nearest neighbor algorithm associated with the Euclidean distance algorithm is implemented to determine which zone the Mobile Station belongs to.

First, find out the k nearest reference points from the Radio Map, in terms of the Euclidean distance. Voting by the k reference points, the one with highest tickets is considered to be the zone that the Mobile Station belongs to. To avoid a tie situation, k should be an odd number. It is necessary to have k smaller than the number of reference points in a zone.

The zone determination results at k =1, k=3, and k=5 for 28 testing points, are shown as below:

K = 1:

0 0 0 0 1 1 1 1 <u>1</u> 2 2 2 3 3 3 3 4 4 4 4 5 <u>3</u> 5 5 6 6 6 6

K = 3:

0 0 0 0 1 1 1 1 2 2 2 2 3 3 3 3 4 4 4 4 5 <u>4</u> 5 5 6 6 6 6

K = 5:

0 0 0 0 1 1 <u>2 2</u> 2 2 2 2 3 3 3 3 4 4 4 4 5 <u>4 4</u> 5 6 <u>5</u> 6 6

The ones with underscores indicates an incorrect zone estimations.

The result were summarized in Figure 4-14. For those 28 testing points, even for k = 1 the zone accuracy rate is better than 90%. When k = 3, we had the best result of the zone determination. As a result, in this experiment we made k = 3. For this adopted k-nearest neighbor algorithm, it is necessary to have k be an odd number. Which made the decision easier. We can also assume that in order to get a better result, k should be smaller than the number of Training Points within a zone.

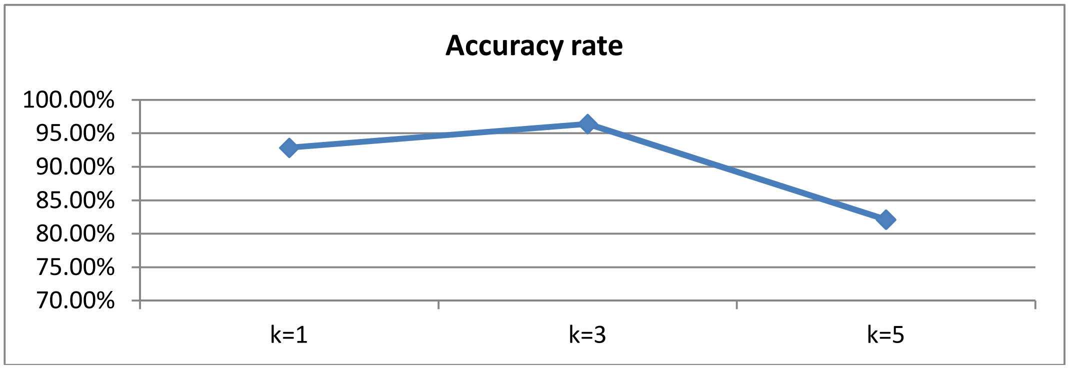


Figure 4-14: Accuracy Rate for the Zone Estimation based on Different k.

4.4.2.2 Location Estimation

From 3.3.3.2, the distance between the Mobile Station and an Access Point can be calculated using the

Formula: $d = 10^{(P(d) - b0)/b1}$.

d1, d2, d3, d4 indicate the distance between the Mobile Station to each Access Point. The mobile station can be anywhere in the zone. The position inside the zone boundaries with min $\|\bar{f}(x,y)\|$ indicates the best estimate of mobile station position.

4.4.3 Position Estimate Result from WiFiPoz

In the offline phase, an Enhanced Radio Map is constructed based on the fingerprints captured for each of the training point.

In the real-time phase, the positioning system found the locations which have k closest RSSI characteristics to the Mobile Station, in terms of Euclidean distance, which is used to identify the Mobile Station location.

Figure 4-15 shows the location estimation using WiFiPoz. It is generated in Abobe Illustrator, with a 1:1 scale. The small squares indicate the actual 28 testing points. A small circle represents a location estimated by the positioning system. The results are very encouraging: WiFiPoz is able to estimate a user’s location to within 2 meters of the actual location.

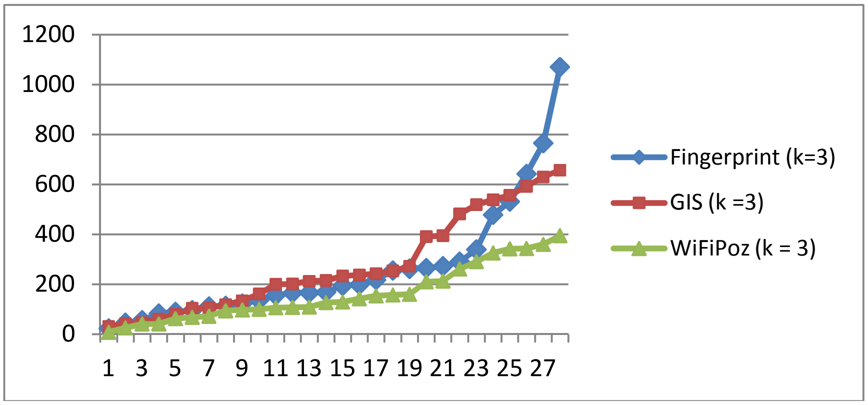
Figure 4-15: Location Estimation Results for WiFiPoz

In the RADAR project [14], the system is reported to have achieved 50% accuracy within about 2.13 meters error distance in the Fingerprint method with k= 3 nearest neighbors. We implement this mechanism, but are able to achieve better accuracy at the 50th percentile is about 1.68 meters. This is possibly due to more Access Points used and a less “noisy” experiment setup. In my experiment, I use 4 Access Points; this can get much better results compared to 3 Access Points. Also, for this experiment, I run the same experiment several times. When I set up the routers in different locations and run the same program, I get a worse result. Using existing infrastructure can dramatically improve the result. I also find that running the experiment in the evening can get a much better result than in the daytime. This research is run in a better test bed compared to other research.

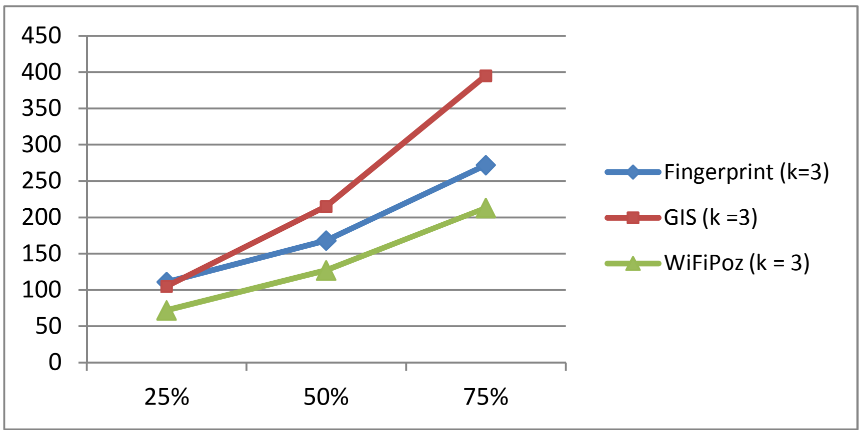
The distance between the actual MS’ position and location determined by the system is measured as error distance. In Figure 4-16 (a), error distances for 28 testing points are sorted and compared for 3 algorithms (Traditional Fingerprint method with k-nearest neighbor algorithm, WiFiPoz and GIS). Figure 4-16 summarizes the information in Figure 4-16 (a) in terms of the 25th, 50th (median), and 75th percentile values of the error distance for each method.

As Figure 4-16 shows, WiFiPoz have the best estimation for most of the Training Points. The 50% percentile (median) is 1.27 meters. The 75% percentile of the error distance is 2.13 meters. GIS had the lowest accuracy. With a 50% percentile of 2.15 meters and a

75% percentile of the error distance is 3.95 meters. Generally, the error distance is GIS > FP > WiFiPoz. The positioning system with lower error distance is considered to be a better one.



(a)



(b)

Figure 4-16: Error Distance(cm) for 35 Trainning Points

4.4.4 Test sets

There is a disadvantage to systems built based on the Fingerprint model. The systems will always require a considerable amount of manual efforts for Radio Map construction when they are used in a new environment. Reducing the training points can speed up the Radio Map construction processes. In this research, I want to demonstrate that WiFiPoz and GIS can get better results with fewer training points.

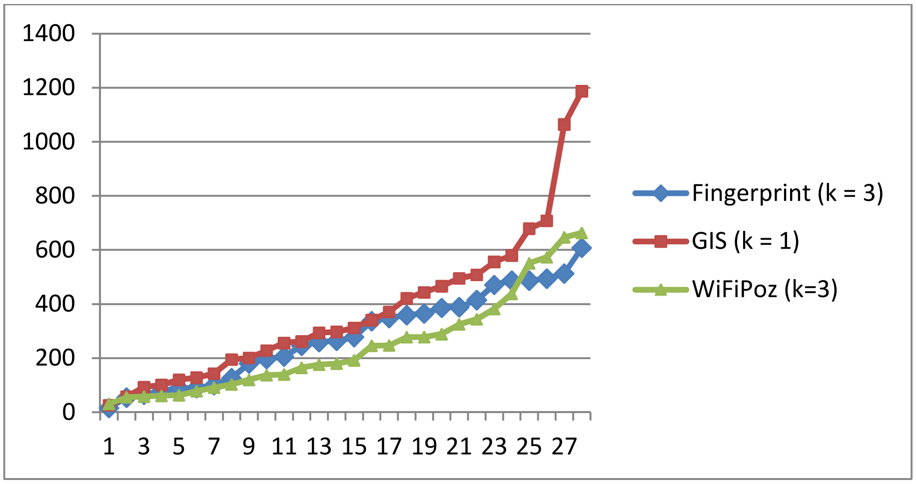
There is no change to the 28 testing points setting in the real-time phase. In the training phase, there are the three setups used for both WiFiPoz and GIS. 1) 5 training points per zone with a total of 35 in the testing area. The result for this setting is discussed in section 4.3. 2) 3 training points per zone, with a total of 21 in the area (Figure 4-17). 3) 2 training points per zone, with a total of 14 training points in the area (Figure 4-18).

Figure 4-17: 21 Training Points Setting

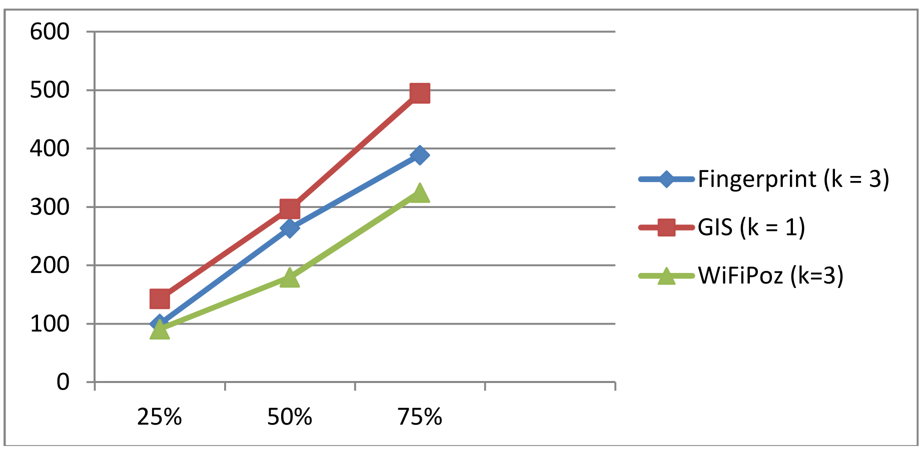
Figure 4-18: 14 Training Points Setting

4.4.4.1 *Setting with 21 Training Points*

We repeat the analysis of the previous sections, with a smaller data set of training points-3 training points in each zone, totaling 21 training points in the testing area. In Figure 4-19, it can be clearly observed that WiFiPoz still has very good accuracy.



(a)



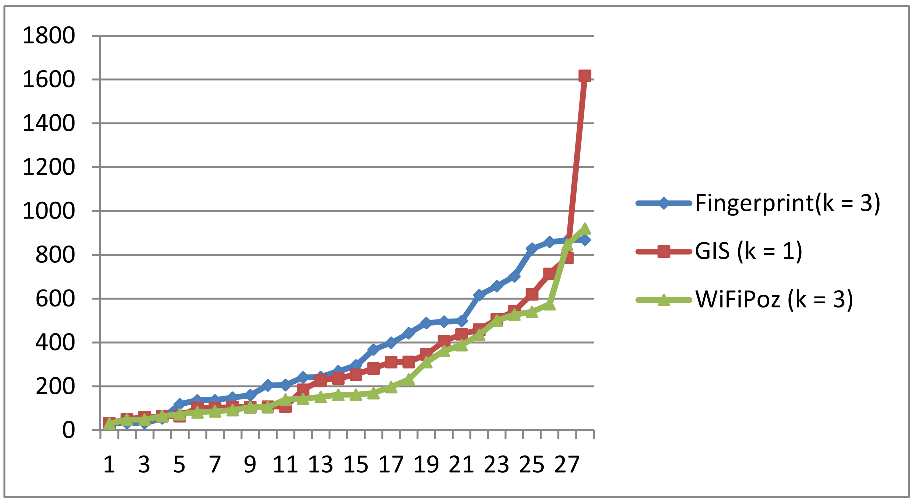
(b)

Figure 4-19: ErrorDdistance(cm) for 21 Training Points

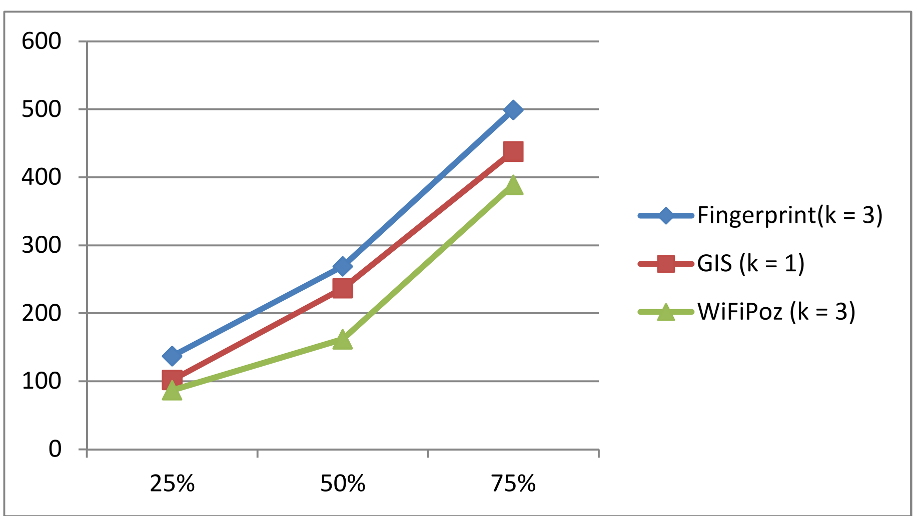
Considering 50% percentile, for instance, the error distance of WiFiPoz is still less than 2 meters. The results are not too different from the 35 training points' results with the error distance for each algorithm increased by about 30%. We now discuss reducing the training points to 14.

4.4.4.2 Setting with 14 Training Points

As Figure 4-20 shows, we repeat the analysis as before except that we had 14 training points in the testing area, with 2 points in each zone. This time the results are different from before. The error distance is FP > GIS > WiFiPoz. It can be clearly observed that WiFiPoz performed extremely well. The median error distance of WiFiPoz is 1.62 meters compare to the traditional Fingerprint which is 2.69 meters. This represents a 40% improvement.



(a)



(b)

Figure 4-20: Error Distance(cm) for 14 Training Points

4.5 Conclusion

Figure 4-21summaries the results from the three sets of experiments for the three algorithms. In order to get a better comparison, we only look at the 50th percentile value of the error distance, which is the median.

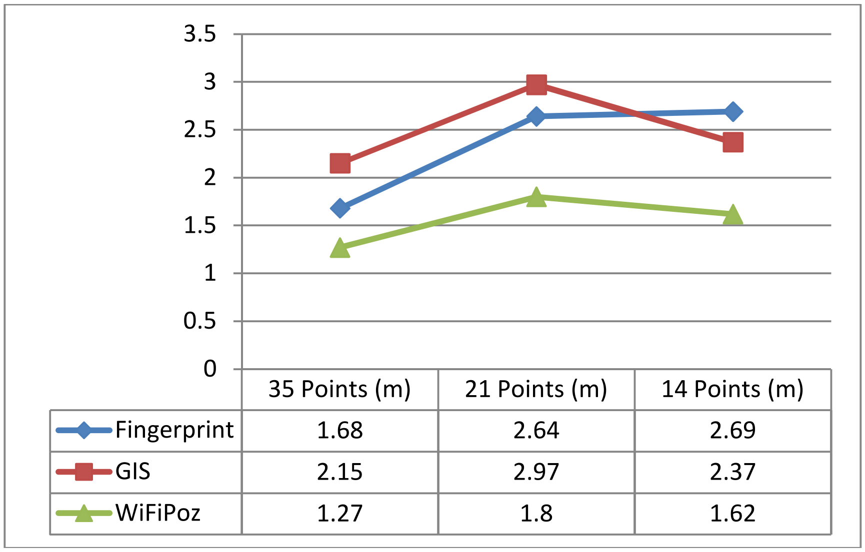


Figure 4-21: Error Distance at 50thPercentle for 3 Algorithms at 3 Settings

After recording the results, we make a couple of observations. First, for the traditional Fingerprint method, when the training points were reduced, the error distance increased dramatically, from 1.68 meters to 2.69 meters, or about 38%. The increase is not linear. The error distance for WiFiPoz and GIS were also affected by the number of training points in the test area. For the 35 training points set, the error distance for WiFiPoz is 1.27 meters. For the 14 training points set, the error distance is 1.62 meters. For GIS, the 35 training points set error distance is 2.15 meters and the 14 Points set has an error distance of 2.37 meters. While the number of training points does not have overwhelming impacts on the accuracy of WiFiPoz and GIS, the effort of measuring 14 training points is much less than measuring 35 training points.

We also observe that WiFiPoz significantly increases the accuracy of position estimation, compared to the traditional Fingerprint algorithm. For instance, at 35 training points, WiFiPoz reduces the error distance to 1.27 meters from 1.68 meters. At 21 training points, WiFiPoz reduces the error distance from 2.64 meters to 1.8 meters. This is about

50% improvement. With 14 training points, WiFiPoz reduces the error distance from 2.69 meters to 1.62 meters. Plus, WiFiPoz improves the accuracy in most cases. In summary, I propose WiFiPoz to be a better algorithm for an indoor positioning system.

WiFiPoz is highly efficient both in accuracy and costs. Experimental results show that with the traditional setting for the Fingerprint method, the accuracy is increased. Yet, when reducing the training points dramatically, WiFiPoz sill has acceptable error distance, comparable to the traditional Fingerprint method with its laborious training phase. The results achieved from the algorithm proposed in this thesis are superior and reliable.

Chapter 5: Conclusion and Future Work

The performance metric of an indoor positioning system is defined in terms of positional error accuracy and reliability. This paper proposes 2 algorithms for indoor positioning systems: WiFiPoz and GIS. The overall mean error calculated for all reported location estimation for WiFiPoz is less than 2 meters, which is much better than existing methods. Mean error distance for GIS is about 2.7 meters, which is comparable to existing methods. Moreover, the equipment and tools used are inexpensive and available off the shelf. We compare our methods to a Fingerprint method. Another widely used algorithm for positioning systems is the Propagation model, which has much worse accuracy (4 -10 m). There are also many other technologies for creating a positioning system, but none of them can work on existing infrastructure and using off the shelf tools with accuracy within 5 m.

The existing training based position systems such as the Fingerprint method also has implementation limitations. A large effort is required in the training process, including data correction and measurement. My experimental results demonstrated that the proposed methods, especially WiFiPoz, work very well with a minimum of effort in the data collection process.

5.1 Future Work

- Since the environment has an effect on the indoor position estimate, it would be worthwhile to carry out further experiments in different environments to see how well WiFiPoz and GIS work compared to a traditional Fingerprint method.
- The scalability of the system can be investigated by increasing the test area. It would be interesting to investigate how well the methods work in a large space like a mall or underground area.
- In this thesis, I focus primarily on the positioning system design and analysis. The database system and algorithms used in this thesis is for a small amount of data. Some issues such as searching time will be important considerations, if we want to apply this system to a larger area.

- An application (end product) can be developed based on research presented in this thesis. It will require some automation of data collection and an interface between the front end and the back end.
- Explore the implementation in other mobile devices, for example mobile phones or mobile computers with different wireless cards.
- The location reporting system can be extended to three dimensions. Floors and multi-stories could be included, although doing this will create new issues.
- These local coordinates have to be mapped to geo-coordinates using the WGS 84 geodesic reference system [49] in order to be used by other applications (e. g. Google maps) to provide the global location information.

References

[1] J. Hightower and G. Borriello, "A survey and taxonomy of location systems for ubiquitous computing," *IEEE COmputer,* vol. 34, no. 8, pp. 57-66, 2001.

[2] M. Satyanarayanan, "Pervasive Computing: Vision and Challenges," IEEE Personal Communications, Jan, 2001.

[3] A.H. Sayed, A. Tarighat and n. Khajehnouri, "Network-based Wireless Location: Challenges Faced in Developing Techniques for Accurate Wireless Location Information," *Signal Processing Magazine, IEEE,* vol. 22, no. 4, pp. 24-40, 2005.

[4] K. Yedavalli, "Location Determination using IEEE 802.11b," Indian Institute of Technology, Madras, India, 2002.

[5] "How gPS works: accuracy," [Online]. Available: http://www.kowoma.de/en/gps/accuracy.htm.

[6] A. Popleteev, "Indoor Positioning Using FM Radio Signals," International Doctorate school in Information and Communication Technologies, Trento, 2011.

[7] H. Liu, et al., "Survey of Wireless Indoor Positioning Techniques and Systems," *Systems, Man, and Cybernetics, Part C: Applications and Reviews,* vol. 37, no. 6, pp. 1067 - 1080, 2007.

[8] S. Holm, "Hybrid ultrasound-RFID indoor positioning: Combining the best of both worlds," *RFID, 2009 IEEE International Conference,* pp. 155 - 162, 2009.

[9] A. LaMarca, et al., "Place Lab: Device Positioning Using Radio Beacons in the Wild," Pervasive, 2005.

[10] P. Bolliger, "Redpin - Adaptive, Xero-Configuration Indoor Localization through User Collaboration," Insitute for Pervasive Computing, ETH Zurich, Switzerland, 2008.

[11] P. A. Zandbergen, "Accuracy of iPhone Locations: A Comparison of Assisted GPS, WiFi and Cellular Positioning," *Transactions in GIS,* p. 5–26, 2009.

[12] G.I Wassi, C. Despins, D. Grenier, C. Nerguizian, "Indoor Location Using Received Signal Strength of IEEE 802.11b Access Point," in *Proceedings of IEEE Canadian Conference on Electrial and Computer Engineer, [[/ 1367-1370*, 2005.

[13] S. Ali, "Indoor Geolocation for Wireless Networks," cranfiled University, 2007.

[14] P, Bahl, and V.N Padmanabhan, "RADAR: An inbuilding RF-based user and tracking system," *Proceedings of IEEE infocom,* vol. 0, no. 5, pp. 775-784, 2000.

[15] K. Pahlavan and X. li, "Indoor Geolocation Science and Technology," *IEEE Communications Magazine,* 2002.

[16] S. Gondor, "Building a Location Sesing and Positioning Service to Enable Indoor Location Based Services," RWTH Aachen University, 2010.

[17] H. Wang et al., "Enhancing the map usage for indoor location-aware systems," pp. 151-160, 2007.

[18] K. Pahlavan et al., "indoor Geolocation in the absence of Direct Path," *IEEE wireless Communications Magazine,* vol. 13, no. 6, pp. 50-58, 2006.

[19] J. A. Tauber et al., "Indoor Location Systems for Pervasive Computing," may 2007. [Online]. Available: http://theory.lcs.mit.edu/~josh/papers/location.pdf.

[20] D. L. Lee and Q. Chen, "A Model-based WiFi Locallization Method," in *In Proceedings of Infoscale*, 2007.

[21] [Online]. Available: http://standards.ieee.org/findstds/standard/802.11n-2009.html.

[22] P. Enge and P. Misra, "Special issue on GPS: the Global Positionging Aystem," in *In Proceedings of the IEEE*, 1999.

[23] J. Shih, "Wireless LAN Location System," The University of Queensland, Indooroopilly, 2003.

[24] A. Bose and C. H. Foh, "A Practical Path Loss Model for Indoor WiFi Positioning Enhancement," in *Comunications and Signal Processing, 2007 6th International Conference*, 2007.

[25] V. Otsason, et al., "Accurate GSM indoor localization," University of Washington, Seattle, 2006.

[26] U. Bandare, et al., "Design and Implementation of a Bluetooth Signal Strength Based Location Sensing System," *Radio and Wireless Conference, 2004 IEEE ,* pp. 319-322, 2004.

[27] B.S Peterson, et al., "A Specification Compatible Bluetooth Inquiry Simplification," *System Sciences, 2004. Proceedings of the 37th Annual Hawaii International*

Conference, p. 9, 2004.

[28] E. Welsh, et al., "Improving Connection Times for Bluetooth Devices in Mobile environments," Rice University, Houston, TX, 2002.

[29] O. Aluf, "Active RFID TAGs system analysis of energy consumption as excitable linear bifurcation system," *Microwaves, Communications, Antennas and Electronics Systems, 2009. COMCAS 2009. IEEE International Conference on ,* pp. 1-9, 2009.

[30] A. Bekkali and M. Matsumoto, "RFID Indoor Tracking based on Inter-tags Distance Measurement," *Wireless Telecommunications Symposium,* p. 1, 2007.

[31] T. Hori, et al., "A Multi-Sensing-Range Method for Position Estimation of Passive RFID Tags," *Networking and Communications, 2008. WIMOB '08. IEEE International Conference on Wireless and Mobile Computing, ,* pp. 208-213, 2008.

[32] S. Hara, et al., "Propagation Characteristics of IEEE 802.15.4 Radio Signal and Their Applicaion for Location Estimation," *Vehicular Technology Conference,* vol. 1, pp. 97-101, 2005.

[33] A. Gunther and C. Hoene, "Measuring Round Trip Times to Determine the Distance Between WLAN Nodes," *Processing of Networking '05,* pp. 768-779, 2005.

[34] M. Liombart, M. Ciurana and F. Barcelo-Arroyo, "On the Scalability of a Novel WLAN Positioning System Based on Time of Arrival Measurements," *Positioning, Navigation and Communication, 2008. WPNC 2008. 5th Workshop,* pp. 15-21, 2008.

[35] L. A. N. Laboratory, "Time-of-arrival location technique," 29 July 1999. [Online]. Available: http://www.fas.org/sgp/othergov/doe/lanl/pubs/00416648.pdf.

[36] N. A. Alsindi, "Performance of TOA Estimation Algorithms in," *Worcester Polytechnic Institute,* 2004.

[37] "TOA model," [Online]. Available: http://www.ciscosky.org/network/wireless/ShenWangLaoZhongDeMoXianDin.

[38] cisco, "Location Tracking Approaches," [Online]. Available: http://www.cisco.com/en/US/docs/solutions/Enterprise/Mobility/wifich2.html.

[39] J. Hightower and G. Borriello, "Location Sensing Techniques," 8 August 2001. [Online]. Available: http://portolano.cs.washington.edu/papers/UW-CSE-01-07-01.pdf.

[40] D. Cotroneo and S. Russo, "Implementing Positioning Services over an Ubiquitous Infrastructure," *Software Technologies for Future Embedded and Ubiquitous Systems,* pp. 14-18, 2004.

[41] M. v. d. Goor, "Indoor Localization with wireless Sensor Networks," 2009.

[42] N. K. Sharma, "A Weighted Center of Mass Based Trilateration Approach for locating wireless devices in indoor environment," *MobiWac '06 Proceedings of the 4th ACM international workshop on Mobility management and wireless access ,* 2006.

[43] T. Parthornratt and K. Techakittiroj, "Improving Accuracy of WiFi Positioning System by Using Geographical Information System (GIS)," *Wireless Telecommunications Symposium, 2006. WTS '06 ,* pp. 1 - 6 , 2006.

[44] K. Raman Kumar, V. Apte and Y.A Powar, "Improving the Accuracy of Wireless LAN based Location Determination Systems using Kalman Filter and Multiple Observers," *Wireless Communications and Networking Conference,* pp. 463 - 468, 2006.

[45] I. Guvenc, et al.,, "Enhancements to RSS Based Indoor Tracking Systems Using Kalman Filters," *GSPx and Intrnational Signal Processing Conference,* March 2003.

[46] N. K. Sharma, "A Weighted Center of Mass Based Trilateration Approach," *Proceedings of the Ingternational Workshop on Mobility Management and Wireless Acess,* pp. 112-115, 2006.

[47] "Application Programming Interface," [Online]. Available: http://en.wikipedia.org/wiki/Application_programming_interface.

[48] MSFT, "about Native Wifi," [Online]. Available: http://msdn.microsoft.com/en-us/library/windows/desktop/ms705945%28v=vs.85%29.aspx.

[49] M. Youssef, A. Agrawala, and U. Shankar, "WLAN Location Determination via Clustering and Probability Distributions," in *IEEE PerCom 2003*, 2003.

[50] R. Jain, "Simple Linear Regression Models," Washington University in Saint Louis, 2008.

[51] "GPS & Selective Availability Q & A," [Online]. Available: http://ngs.woc.noaa.gov/FGCS/info/sans_SA/docs/GPS_SA_Event_QAs.pdf.

[52] [Online]. Available: www.stanford.edu/~kcobb/hrp259/lecture13.ppt.

Appendix A: Table of zone boundary (cm)

	Zone1	Zone2	Zone3	Zone4	Zone5	Zone6	Zone7
X	0~1620	1620~3000	3000~4330	4170~4330	3000~4170	1850~3000	1620~ 1850
Y	0~170	0 ~ 170	0~ 170	170~ 1280	1100~ 1280	1100 ~ 1280	170~1760

Appendix B: The Coordinate Measurement of 35 Training Points (cm)

Zone1	(232, 126)	(561, 50)	(864, 126)	(1180, 50)	(1462, 126)
Zone2	(1732, 50)	(2002, 126)	(2300, 50)	(2530, 126)	(2830, 50)
Zone3	(3140, 126)	(3445, 50)	(3684, 126)	(3945, 50)	(4145, 126)
Zone4	(4230, 246)	(4300, 486)	(4230, 725)	(4300, 945)	(4230, 1189)
Zone5	(4011, 1231)	(3726, 1150)	(3550, 1231)	(3330, 1150)	(3137, 1231)
Zone6	(2910, 1231)	(2667, 1150)	(2443, 1231)	(2207, 1150)	(2004, 1231)
Zone7	(1706, 340)	(1826, 590)	(1706, 942)	(1826, 1250)	(1706, 1653)

Appendix C: zone parameters for 35 Training Points Setting

The list of R^2

	Zone 1	Zone 2	Zone 3	Zone 4	Zone 5	Zone 6	Zone 7
AP 1	0.4459	0.7365	0.3092	0.8387	0.7312	0.1205	0.3317
AP 2	0.9157	0.8199	0.8448	0.6786	0.1038	0.1741	0.4496
AP 3	0.1455	0.4815	0.2858	0.881	0.3731	0.9222	0.856
AP 4	0.5382	0.6677	0.6716	0.628	0.26	0.5378	0.6697

The list of b_1 co-efficient

	Zone1	Zone2	Zone3	Zone4	Zone5	Zone6	Zone7
AP1	-2.19	-43.97	43.82	-706.72	-93.74	-21.09	-49.69
AP2	-54.32	-83.58	-58.91	-584.84	11.02	-10.22	-5.49
AP3	-12.66	-5.82	-4.12	-14.52	-47.00	-56.77	-39.44
AP4	-32.27	-57.57	-90.97	-5.54	0.28	34.00	-355.90

The list of b_0 offset

	Zone1	Zone2	Zone3	Zone4	Zone5	Zone6	Zone7
AP1	-40.11	95.38	-210.55	2489.77	256.35	-2.45	96.39
AP2	102.46	199.00	126.21	1939.61	-85.28	-18.36	-31.70
AP3	-10.84	-32.24	-34.91	-9.14	94.86	122.43	67.94
AP4	38.53	129.59	224.80	-37.34	-48.38	-160.61	1137.66

VITA

Author: Xiaoyi Ye

Place of Birth: China

Undergraduate School Attended: Donghua University, China

Graduate School Attended: Georgia Institute of Technology

Degrees Awarded: Bachelor of Science, 2003, Donghua Unibersity
Master of Industrial Design, 2008, Georgia Institute of Technology

Honors and Awards: Graduate Assistantship, Center for Assistive Technology and Environmental Access, 2006-2008, Georgia Institute of Technology

Graduate Assistantship, Computer Science Department, 2011-2012, Eastern Washington University