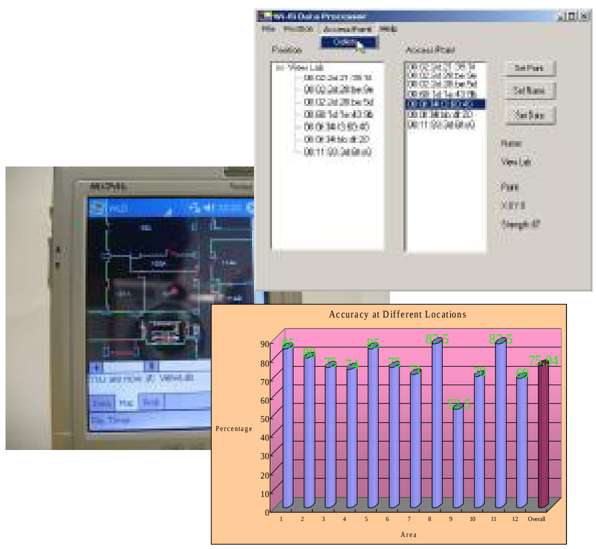
LYU0401 Location-Based Multimedia Mobile Service

2003/2004 Final Year Project

First Term



Supervisor

Professor Michael R. Lyu

Fung Ka Kin

Ma Tilen

Abstract

This report is written to summarize all the work of our final year project entitled Location-Based Multimedia Mobile Service in the first semester. During this semester, we mainly focus on the problem of localization. Therefore, base on our knowledge and developed tools in localization, we are able to further develop a location-based service.

In this semester, we have chosen the 1st floor of the Ho Sin-Hang Engineering Building, the Chinese University of Hong Kong to study the problem of localization. Our goal is to locate a person when he/she is walking around on the floor. To achieve the goal, we have encountered a lot of problems. For example, how to collect the signals from the Wireless LAN access points, how to process the data, how to choose and apply algorithms in order to use the collected signals and so on. Every problem mentioned will affect the accuracy in localization. Therefore, in this report, we are going to introduce how we can solve these problems.

We first give an introduction about the current issue of localization in recent years. As we have chosen the Wireless LAN (WLAN) network for localization, the topics in Wireless LAN fundamentals will help you to understand some terms and standard in WLAN as well as our project. In the next section, we introduce different approaches and algorithms in localization, so that you have a brief idea on how to achieve localization. After that, we will give a detail description on how we apply the Area-Based Probability algorithm and solve the problems we face. We will also list out the result of the experiments that we have done on the accuracy of our system and give some analysis of them.

We have implemented our first application Wi-Fi Location System (WLS). It is a tool for developer to implement Location-based System. WLS consists of 3 components, namely, Wi-Fi Signal Scanner (WSS), Wi-Fi Data Processor (WDP) and Wi-Fi Location Detector (WLD). WSS is used to collect signal data from Wireless LAN access points at target place. The data is then processed by WDP. The output data of WDP is used by WLD to detect the current position. So, they have a close relationship between each other in WLS. The detail explanation of WLS will be given later in this report.

At the end, we will give a brief summary and a listing of the problem that we

have faced together with our solution. Then we will describe our contribution to the project. The last section is our project progress and future work in next semester.

1. Introduction

Localization is necessary for many higher level sensor network functions such as tracking, monitoring and geometric-based routing. There are many applications that provide services based on the location of the user, such as telephone follow me, which forwards phone calls to the user's current location, everywhere printing, which chooses the nearest printer for mobile users, and intelligent tourist, which offers help information based on a tourist's location.

Many positioning systems designed to determine or track a user's location have been proposed in recent years. Those systems fall into three categories: global location systems, wide-area location systems based on cellular networks, and indoor location systems.

A typical global location system is the Global Positioning System (GPS), which receives signals from multiple satellites and employs a triangulation process to determine physical locations with an accuracy of approximately 10 m. However, GPS is inefficient for indoor use or in urban areas where high buildings shield the satellite signals.

Several cellular-network-based wide-area location systems have been proposed in recent years. The technological methods of location determination involve measuring the signal strength, the angle of signal arrival, and/or the time difference of signal arrival. However, the accuracy of wide-area location systems is highly limited by the cell size. Moreover, the effectiveness of systems for an indoor environment is also limited by the multiple reflections suffered by the radio frequency (RF) signal.

For an indoor environment, several systems based on various technologies such as infrared (IR), ultrasound, video surveillance, and radio signal are emerging. Among these systems, radio-signal-based approaches—more specifically, the wireless local-area network (WLAN) (IEEE 802.11b, also named Wi-Fi) radio-signal-based positioning system—have drawn great attention in recent years.

A WLAN-based positioning system has distinct advantages over all other systems. First, it is an economical solution because the WLAN network usually exists already as part of the communications infrastructure. For a notebook computer, personal digital assistant (PDA), or other mobile devices equipped with WLAN capability, the positioning system can be implemented simply in software—generally in middleware or at the application level. This software based location system significantly reduces cost with respect to dedicated architectures.

Second, the WLAN-based positioning system covers a large area compared with other types of indoor positioning systems. The WLAN-based positioning system may work in a large building or even across many buildings. Third, it is a stable system owing to its robust RF signal propagation. Video- or IR-based location systems are subject to restrictions, such as line-of-sight limitations or poor performance with fluorescent lighting or in direct sunlight.

In this innovative WLAN-based indoor positioning technology, the signal strength distribution of WLAN access points is collected to train a position-determination model. The training phase is followed by the working phase, during which the mobile device observes the WLAN signals and applies the position-determination model to calculate a position. To reduce the complexity of the training phase, the position-determination model is only trained from a limited amount of collected samples. To improve the accuracy of the location system, a localization algorithm is introduced in which the position determination relies on both collected signal strength and knowledge of space topology.

We have set up the WLAN-based positioning system on the first floor of the Ho Sin-Hang Engineering Building. The system is installed in Personal Digital Assistant (PDA) with Window CE as its platform. We have used experiments to evaluate the performance of our system. The results of these experiments indicate that our system achieves an accuracy of 76% to 99.5% depending on the sample size of the testing set.

2. Introduction to Wireless Network

2.1 Access Point (AP)

It is a hardware device or a computer's software that acts as a communication link for user to access the wired Local Area Network (LAN). Each Access Point has a unique Network Interface Card (NIC). User can communicate the Access Point through this interface.

2.2 Wireless Terminology

Media Access Control address (MAC Address)

It is a unique hardware address that identifies each node in the network. Referencing to OSI 7 Layer Model, the Data Link Control (DLC) layer in the IEEE 802 network is divided into two sub layers: the Logical Link Control (LLC) layer and the Media Access Control (MAC) layer. And each different type of network medium requires a different MAC layer.

Those networks that do not confirm to IEEE 802 standards, the nodes in the network are called Data Link Control (DLC) address.

Service set identifier (SSID)

It is a 32 character. It adds to the header of packets that acts as a password when mobile device sends the packet through Wireless Local Area Network (WLAN). The SSID in one WLAN is different from another. All access points and mobile devices connecting to a specific WLAN must use the same SSID. They can be permitted to join the WLAN provided that they can provide the unique SSID. However, SSID can be sniffed in WLAN which easily get from a packet. It does not supply any security to WLAN.

Receive Signal Strength Indicator (RSSI)

It is the signal strength received by the mobile device in the WLAN. Its unit is in dBm. The smaller values of RSSI, the greater strength received by the mobile device. RSSI is a negative integer number.

Wireless Fidelity (Wi-Fi)

The Terms “Wireless Fidelity” is published by Wi-Fi Alliance. Formerly, the term “Wi-Fi” was only referring to IEEE 802.11b. Now, “Wi-Fi” generally means the all type of wireless network, including IEEE 802.11a, IEEE 802.11b, etc. Any networking products are tested by “Wi-Fi Alliance”. It is certificated as “Wi-Fi Certified”. This guarantee “Wi-Fi Certified” product can communicate with other “Wi-Fi Certified” product without problem.

Wired Equivalent Privacy (WEP)

It is a security protocol for wireless Local Area Network (WLAN). It aims to be protected from unauthorized access. It protects the WLAN by encrypting the data over the radio channel. However, it found that WEP can be cracked by advanced technique. WEP is used at the two lowest layers of the Open System Interconnection (OSI) model - the data link and physical layers; it therefore does not offer end-to-end security.

Wi-Fi Protected Access (WPA)

It is designed to improve upon the security features of WEP. It can work with existing Wi-Fi products that have been enabled with WEP. It improves data encryption through temporal key integrity protocol (TKIP). And it adds authentication through the extensible authentication protocol (EAP). EAP is a more secure encryption system ensuring only authorized users can access the WLAN.

2.3 Wireless Standard

IEEE 802.11

It is a family of wireless Local Area Network (WLAN) specifications developed by Institute of Electrical and Electronics Engineers (IEEE).

Currently, there are four specifications in IEEE 802.11 family:

a. IEEE 802.11
b. IEEE 802.11a
c. IEEE 802.11b
d. IEEE 802.11g

802.11e and 802.11i will be approved in 2004

IEEE 802.11

It operates on the 2.4 GHz band using either direct sequence spread spectrum (DSSS) or frequency hopping spread spectrum (FHSS). It provides up to 2 Mbps transmission rate.

IEEE 802.11a

It operates on 5 GHz band using orthogonal frequency division multiplexing encoding scheme (OFDM). It provides up to 54 Mbps transmission rate. It is not interoperable with IEEE 802.11b. Only eight channels are available.

IEEE 802.11b

It is often called “Wi-Fi”. It operates on 2.4 GHz using direct-sequence spread spectrum (DSSS) with complementary code keying (CCK). This allows high access to data at up to 300 feet from base station. It provides up to 11 Mbps transmission rate. Fourteen channels are available for this standard. Only eleven channels can be used in the United States due to Federal Communications Commission (FCC) regulations. Comparing to IEEE 802.11a, it requires fewer access points to cover large areas.

IEEE 802.11g

It operates on 2.4 GHz band. It provides up to 54 Mbps transmission rate. It uses orthogonal frequency division multiplexing encoding scheme (OFDM) when transmission rate is above 20 Mbps. Otherwise, it uses direct-sequence spread spectrum (DSSS) with complementary code keying (CCK). It improves security enhancements over 802.11. It is compatible with IEEE 802.11b. Fourteen channels are available in the 2.4 GHz band.

IEEE 802.11e

It is the first standard that targets on home and business environment. It has Quality-of-Service (QoS) in this standard. It adds multimedia support to the existing IEEE 802.11b and IEEE 802.11a wireless standards such that it has a full compatibility with previous standard. It can implement in following kinds of application, e.g. Video on demand, audio on demand, voice over IP (VoIP) and high-speed Internet access, etc.

IEEE 802.11i

It adds the Advanced Encryption Standard (AES) security protocol. This protocol is much stringer than the Wi-Fi Protected Access security standard (WFA) which implemented in the previous standard.

2.4 Open System Interconnection (OSI) 7 Layer Model

The model defines a networking framework in seven layers. It is divided into 7 Layer. The protocol is implemented according to each layer requirement in the model. Control in each layer is passed from one layer to next layer , starting at the top layer(application layer) to the bottom layer(physical layer) over the channel to the next station and back up the layer from the bottom layer to top layer.

The following is the OSI 7 Layer Model:

Application
Presentation
Session
Transport
Network
Data Link
Physical

Figure 2.1

Application Layer (Layer 7)

Layer 7 is the Application Layer. This layer defines the end-user and end-application protocols, such as telnet, http, and ftp. Everything is application-specific must be defined or implemented in this layer.

Presentation Layer (Layer 6)

Layer 6 is the Presentation Layer. This layer defines the data representation. It formats the data into specific format before transmit to the outside network. Protocol conversions, encryption/ decryption and graphics expansion are all implemented in this layer.

Session Layer (Layer 5)
Layer 5 is the Session Layer. It establishes, manages and terminates connections between applications in the station. This layer handling sets up, coordinates, and terminates conversations, exchanges, and dialogues between the applications at station.

Transport (Layer 4)
Layer4 is the Transport Layer. This layer handles the transferring data between end systems. It provides the end-to-end error recovery and flow control. It ensures the data can transfer correctly. Transmission Control Protocol (TCP) and User Datagram Protocol (UDP) is the example that implement in this layer.

Network (Layer 3)
Layer3 is the Network Layer. This layer provides communication between the end system by switching and routing technologies. It also handles addressing, internetworking, error handling, congestion control and packet sequencing. Internet Protocol (IP) is the protocol implementing in this layer.

Data Link (Layer 2)
Layer2 is the Data Link Layer. It is divided into two sub layers: The Media Access Control (MAC) layer and the Logical Link Control (LLC) layer. The MAC layer controls how computer sending and receiving the data on the network. The LLC layer controls frame synchronization, flow control and error checking.

Physical (Layer 1)
Layer1 is the Physical Layer. It defines the physical and electrical signal of the network. The Network Interface Card (NIC) in the computer and the interfaces in the network equipments all run at this level.

3. Algorithms in Localization

3.1 Terms and Definitions

In order to understand the algorithm used in localization in a more systematic and formal way, we introduce some terms and definitions which are commonly used to describe Localization algorithm in wireless LAN.

A wireless LAN access point is represented by AP. Suppose there are n Access Points on the floor, they are represented by AP1, AP2, …, APn. Also, we assume that there are total m areas to be located, that is they can be represented by A_1, A_2, …, A_k. The offline measured signal strengths and locations an algorithm used is called the training set T_0. A training set (T_0) consists of a set of fingerprints (S_i) at m different areas A_i. More mathematically, T_0 can be represented by this equation, $T_0 = \{(A_i, S_i)\}$, i = 1 … m.

So, what is a fingerprint S? S_i is the set of expected average signal strengths measured from APs at a particular area A_i. As there are totally n APs, so a fingerprint S_i should have n elements and each of them is from one AP. More formally, $S_i = (s_{i1}, …, s_{in})$, where s_{ij} is the expected average signal strength from AP_j.

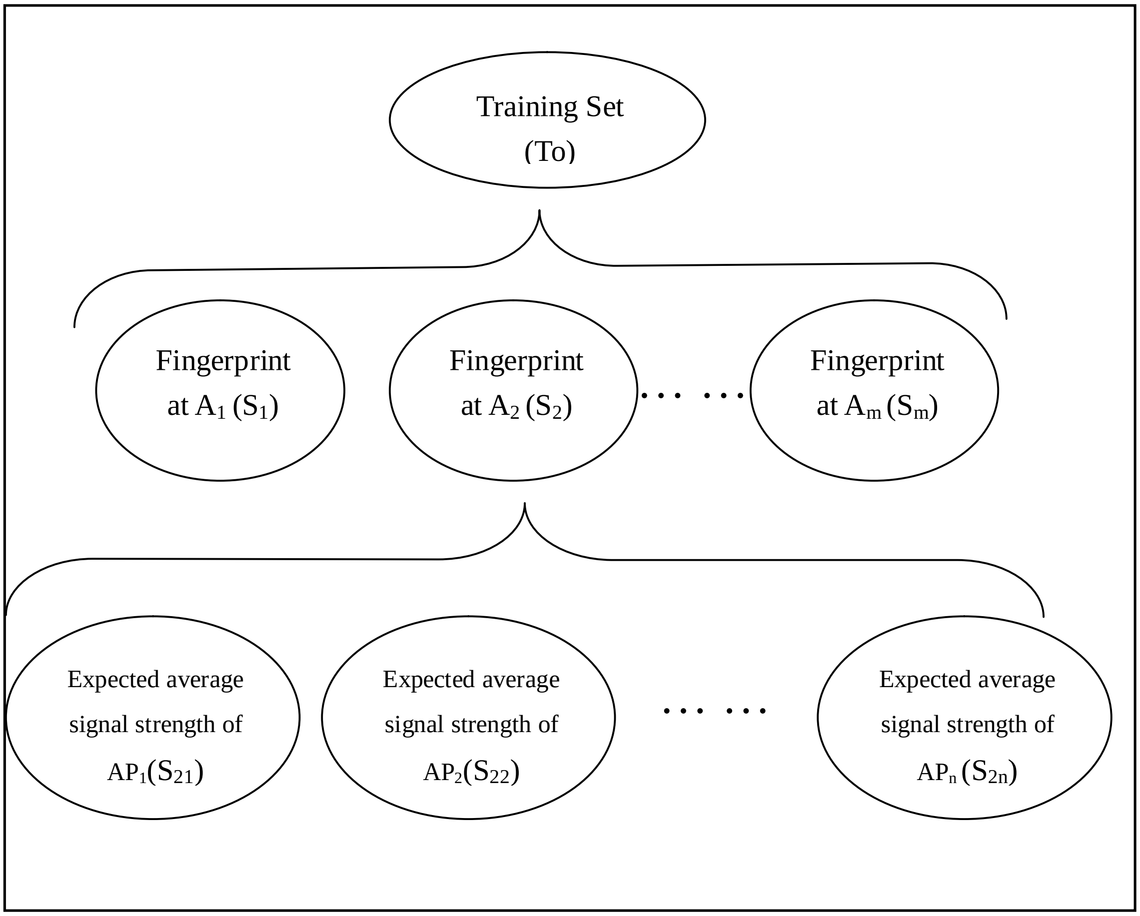


Figure 3.1 summarizes the structure of a training set

Received signal strength (RRS) refers to the signal strengths received from APs by the mobile device when a user uses it for localization. RRS is similar to a fingerprint. It contains the set of signal strengths measured from APs at the current position. The number of elements in RRS depends on how many access points the mobile device can detect at the current position.

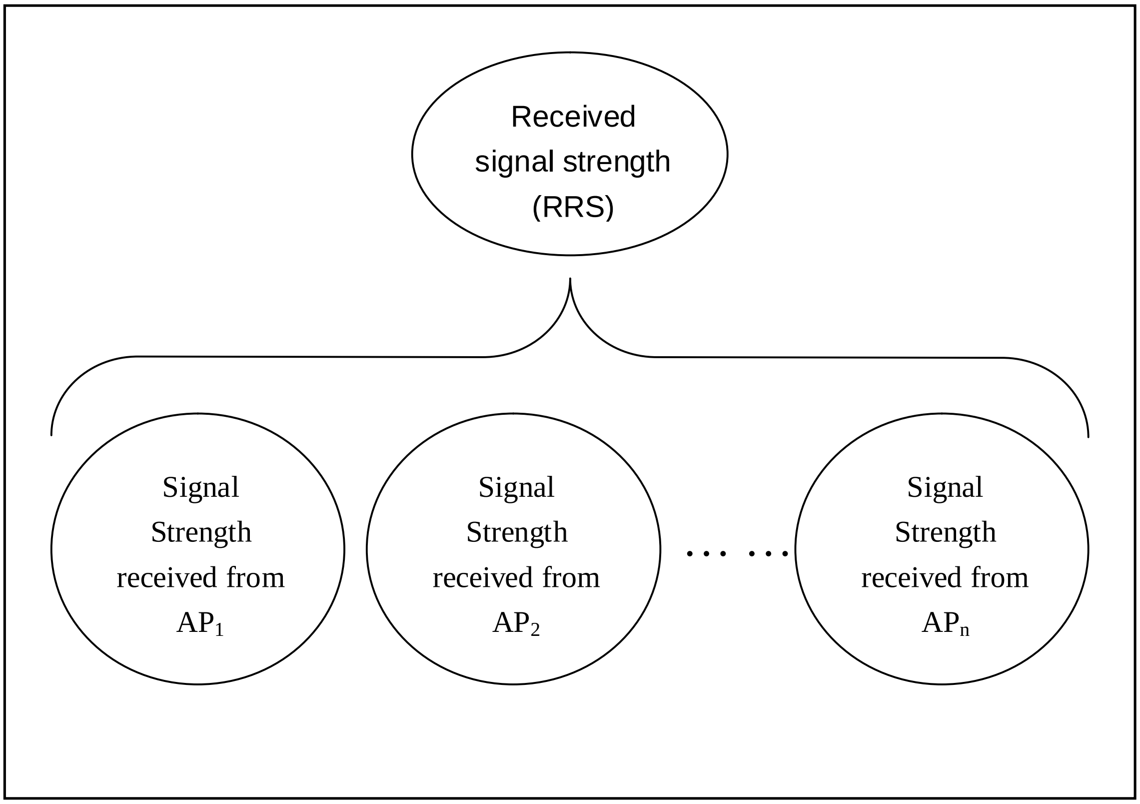


Figure 3.2 summarizes the structure of a RRS

3.2 Generating a Training Set

As mentioned above, a Training Set is the offline measured signal strengths and locations an algorithm used in localization. A training set is essential for localization because it records the characteristics of different areas to be located. A training set is later used by an algorithm to distinguish different locations and find out the best fit location when a testing set is applied. A testing set is measured at real time when a person uses a mobile device to locate himself/herself. Therefore, it is important to prepare a training set for localization. The following will show a common way to generate a training set step by step.

In order to get the expected average signal strength from a particular AP_j at A_i, we need to read a series of signal strengths (s_{ijk}) for this particular AP_j with a constant time between samples. After that, we calculate sij by averaging the series, {s_{ij1}, s_{ij2}…, s_{ijoij} }, where oij is the number of samples from AP_j at A_i.

Then, we can generate S_i by using the above method and measure n series of signal strengths (s_{ijk}) for all the n APs. Finally, we use the same method to generate the fingerprints S at all m areas to be located and hence we will generate the training set.

3.3 Distance Mapping Algorithm

This algorithm is a simple algorithm which does not require a training set. So, it can be implemented more quickly. We first find out the locations of all the wireless LAN access points. During the localization, we create a testing set by measuring the signal strengths received locally from different access points. Then, we use the assumption that the distance between the current location and a particular access point is in direct proportion with the signal strengths measured from the access point.

Therefore, the possible locations that we can predict from the signal strengths from one particular AP are in a circular shape illustrated. In order to find the current location, we need to minimize the set of possible locations. It can be done easily by overlapping each of the circular set of possible locations predicted by different APs. Figure 3.3 shows how we can apply this algorithm and find out the current location. With stronger signal strengths received from an AP, the distance from it is smaller and so the circle illustrated in the figure is smaller. On the other hand, if the signal strengths received are weaker, the distance from it is longer and hence a larger circle.

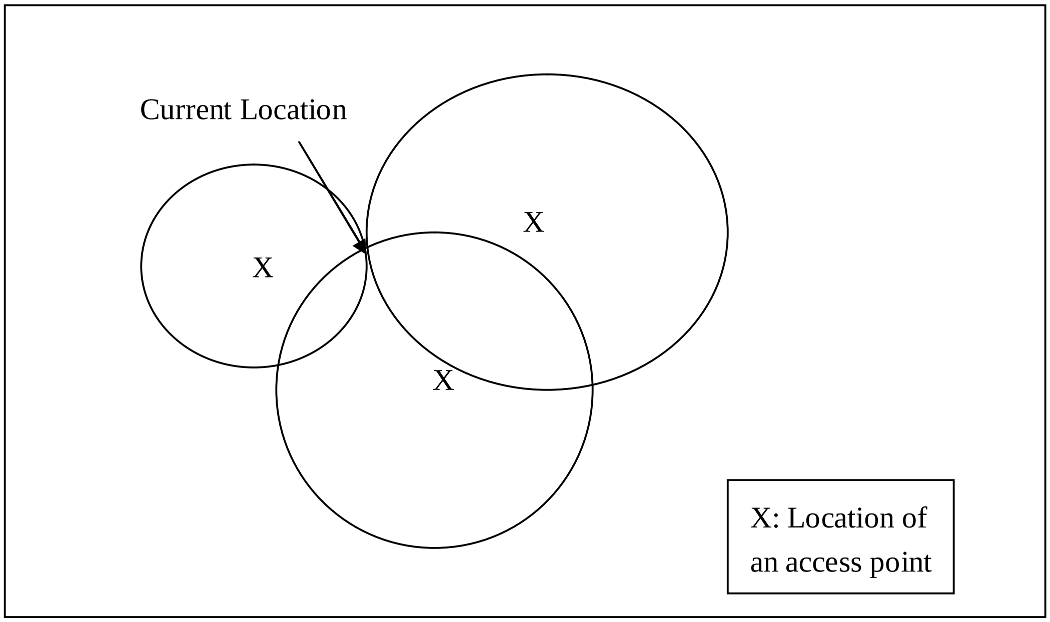


Figure 3.3

3.4 Point-Based Approach VS Area-Based Approach

There are mainly two approaches used in localization, namely, the point-based approach and the area-based approach. Previous WLAN localization work often uses the traditional point-based approach to localization. In these approaches, the localization goal is to return a single point of possible location for the object to be located.

The above Distance Mapping algorithm is a simple algorithm using the point-based approach. However, in our case for a WLAN-enabled mobile device, we have used a novel area-based approach for WLAN localization systems based on the advantages of using this approach. The goal in an area-based localization system is to return a possible location of the mobile object as an area rather than a single location. So, in our case, we will return area of rooms, the areas of corridor corners or the area of corridor in front of certain rooms as the output result.

The advantages of area-based approaches arise from the fundamental uncertainty resulting from probabilistic approaches used by WLAN localization systems. Area-based systems are better able to utilize and describe this uncertainty in a meaningful manner as compared to point-based approaches.

One advantage is that area-based approaches can direct the user in his search for an object in a more systematic manner as compared to point-based approaches. For example, when looking for lost keys using an area-based approach, the user can begin his search for the keys in the most likely area maybe at room 101. Then he continually expands his search to the next most likely area maybe to room 102. The user thus searches in a systematic manner in relation to the likelihood of the object's presence in the area.

A second advantage of area-based localization is that it presents the user an understanding of the system in a more natural and intuitive manner than a point-based approach. If our system returns Rm101 as the result, the user will have a better confidence for its location rather than returning a point with x and y coordinates.

The critical property that area-based systems exhibit in dealing with the uncertainty is their ability to trade accuracy for precision. Intuitively, localization accuracy is the error between the estimated position and the object's true position. For area-based systems, we define accuracy as the distance the object is from the returned area. Precision describes the size of the area. A point is infinitely precise, but may not be very accurate. On the other hand, the area containing the entire scope of the localization system (e.g. the whole building) would have a high accuracy but poor precision. In order to achieve our goal to apply localization on 1st floor of the engineering building, we are not required to have a result as precise as a point; however, we apply the area-based approach to increase the accuracy for the WLAN-based system to localize objects.

3.5 Simple Point Matching Algorithm

The strategy behind Simple Point Matching (SPM) is to find a set of areas(Ai) that fall within a threshold of the RSS for each AP independently, and then return the area(s) that forms the intersection of each AP's set. Figure 3.4 describes SPM in pseudo code.

input $Grid, \Sigma, FT$ **output** $Area$
$\forall j = 1 \ldots n, noise[j] = 0$
loop until $Area \neq \phi$
  $Area = FT$
  for $j = 1$ **to** n **begin**
    $noise[j] = noise[j] + \Sigma[j]$
    candidateTiles$[j]$ = findTiles$(\overline{s_{lj}} - noise[j], \overline{s_{lj}} + noise[j])$
    $Area = Area \bigcap$ candidateTiles$[j]$
  end for
end loop
return $Area$

The SPM algorithm.

Figure 3.4

The Grid, Σ= (ε1, . . . , εn), FT , and Area, correspond to the a vector of the expected signals' standard deviation received from each AP, a set of all the tiles on the floor, and the returned area, respectively.

More formally, SPM first finds n sets of areas, one for each AP_j , j = 1 . . . n, that "match" all fingerprints $S_l = (s_{l1}, \ldots, s_{ln})$. The matching areas for each AP_j are found by adding an expected "noise" level, q to s_{lj}, and then by returning all the floor areas that fall within the expected threshold, $s_{lj} \pm q$ (we may substituted a value of -92 dBm for missing signals). SPM then returns the area formed by intersecting all matched areas from the different AP area sets.

For increasing the precision of the algorithm, i.e., to find the fewest high probability areas, it starts from a very low q. However, it then runs into the risk of returning no areas when one of the area sets returned from a particular AP is empty. Thus, on an empty intersection, the algorithm additively increases q, i.e., it first tries q, 2q, 3q,… , until a non-zero set of tiles results. Therefore, even in the worst case, a non-empty intersection will result when q is large enough.

An important issue is the how to pick the q for each AP_j. An intuitive way is to pick the expected standard deviation of the signals received from AP_j. For simplicity, you may pick 10 as the standard deviation for all APs as it is commonly used by localization. However, the standard deviation can vary from an access point to another, depending on the strength of the signal. Therefore, different standard deviation can be used as the threshold q for each AP to increase the accuracy in localization.

3.6 Area Based Probability

The strategy used by the Area-Based Probability (ABP) algorithm is to return a set of areas bounded by a probability that the object is within the returned set. We call the probability, α, the confidence, and it is an adjustable parameter.

To find a result area set, ABP computes the likelihood of received signal strength (RSS) that matches the fingerprints of each area in the training set (T0). Then it normalizes these likelihoods given the prior conditions: (1) the object must be on floor, and (2) all areas are a-priori equally likely to be visited. ABP then returns the top probability areas whose sum matches the desired confidence.

The confidence controls the accuracy-precision tradeoff. ABP thus stands on a more formal mathematical foundation than SPM. Figure 3.5 summarizes the ABP algorithm.

---

input $L, \bar{S}_l, conf$ **output** $Area$
$\forall i = 1 \ldots L, H(i,1) = L_i, H(i,2) = P\left(\bar{S}_l | L_i\right)$
$c = \sum_{i=1}^{L} H(i,2)$
$\forall i = 1 \cdots L, H(i,2) = \frac{H(i,2)}{c}$
sort descending on second column(H)
$Area = \phi, k = 1, prob = 0$
loop until $prob \geq conf$
  $Area = Area \bigcup H(k,1), prob = prob + H(k,2)$
  $k = k + 1$
end loop
return $Area$

The ABP algorithm.

Figure 3.5

In ABP, signals received from different access points are assumed to be independent. For each AP_j , j = 1 . . . n, the sequence of received signal strengths s_{ijk}, k = 1 . . . o_{ij} , at each (x_i, y_i) in the original training set, T_0, is modeled as a Gaussian distribution. Although this assumption is not generally true, it significantly simplifies the computations with little performance degradation.

To apply the algorithm, we first generate a training set by the procedure mentioned in Section 3.2. Then, we compute the "mean" parameter of the distribution, s_{ij} for each fingerprint in the original training set T_0. For each AP_j , we assume that the standard deviation of the distribution at all the areas on the floor is constant, and equals to ε_j.

The algorithm use Bayes' rule to compute the probability of being at each location, L_i, on the floor given the testing set S_l, i.e., the received signal strength of the object to be located. The Bayes' rule is given below:

$$P\left(L_i|S_l\right) = \frac{P\left(S_l|L_i\right) \times P\left(L_i\right)}{P\left(S_l\right)} \qquad (1)$$

$P(S_l \mid L_i)$ is the probability of having the testing set S_l when having known the current location is at L_i. $P(S_l)$ is the probability of receiving this set of signal strengths S_l and $P(L_i)$ is the probability at the location L_i.

However, $P(S_l)$ is a constant c_1. Moreover, given we do not have prior knowledge about the exact object's location. We assume that the object to be localized is equally likely to be at any location on the floor. So, $P(L_i) = P(L_j)$, for all i and j and $P(L_i) = c_2$ for all i, where c_2 is another constant. Therefore, after we let $c= c_2/ c_1$, equation (1) is rewritten as equation (2) as follow:

$$P\left(L_i|\bar{S}_l\right) = c \times P\left(\bar{S}_l|L_i\right) \qquad (2)$$

By using equation (2), we can derive the following:

$$Max\{P(A_i \mid S_t) \} = Max\{c{*}P(S_t \mid A_i) \} = Max\{P(S_t \mid A_i) \} \qquad (3)$$

Without having to know the value c, we can just return the location (area A_i) with maximum of $P(S_l \mid A_i)$. In order to compute $P(S_l \mid A_i)$ for all areas, we use the whole set $S_l = (s_{l1}, \ldots, s_{ln})$ for computing the probability, i.e., signal strengths from all the access points. We also substitute a value of -92 dBm for missing signals, when the current location is not covered by the access point.

Let $P(s_{lj} \mid A_i)$ be the probability of having a signal strength of s_{lj} from AP_j given the location is at A_i. $P(s_{lj} \mid A_i)$ can be found by using our Gaussian distribution assumption. The follow is the equation of Gaussian distribution:

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)} \qquad (4)$$

So, we compute $P(S_l \mid A_i)$ by using the formula of independent event in probability:

$$P(S_l \mid A_i) = P(s_{l1} \mid A_i) \text{ X } P(s_{l2} \mid A_i) \text{ X } \ldots \text{ X } P(s_{ln} \mid A_i) \qquad (5)$$

Given that the object must be at exactly at one area, i.e., $\Sigma P(A_i \mid S_l) = 1$ for all I, ABP computes the actual probability density of the object for each area on the floor. Finally, ABP returns the top probability area(s) above its confidence, ε. This area(s) forms the result of localization.

4. Experiments and Implementation of Localization Algorithm

4.1 Our Choice of Algorithm

Distance Mapping Algorithm is simple and efficient in localization however it is not accurate and practical. It is because in real situation, the signal strength of an access point is not steady. It will fluctuate within a range. Therefore, the possible locations returned from one access point changes with time. By overlapping these area sets from all the access points, we can imagine that the resulting area is fluctuating. It indicates that the algorithm is very inaccurate.

Moreover, it is difficult to get a precise result from this algorithm. Sometimes, the overlapping area is an empty set (illustrated by Figure 4.1). And sometimes, a large area as the result set will be returned (illustrated by Figure 4.2).

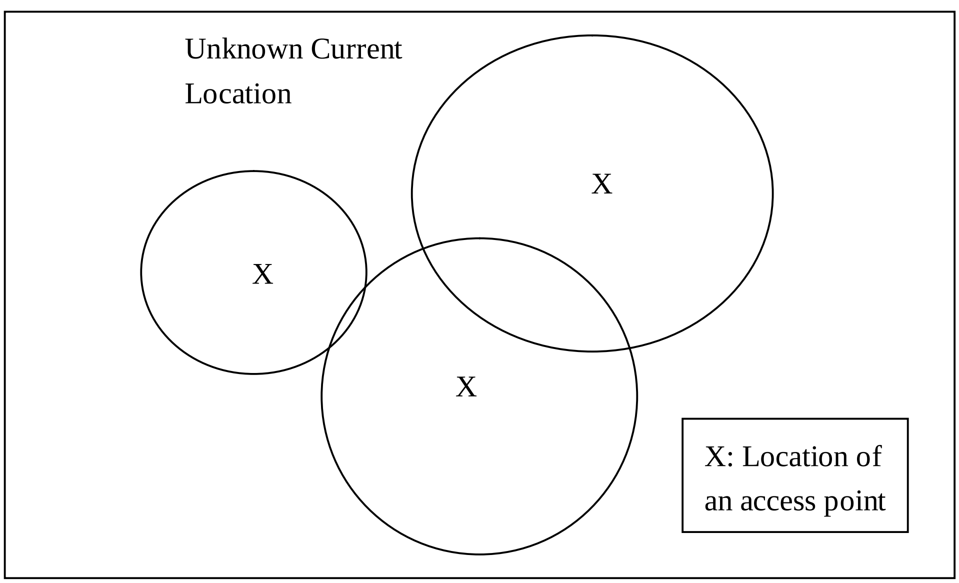


Figure 4.1

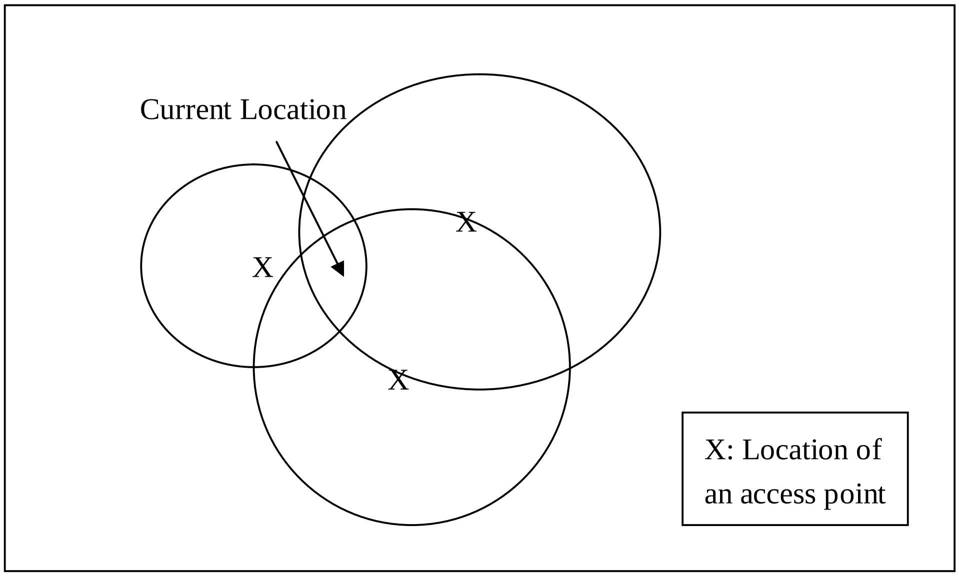


Figure 4.2

In comparing the Simple Point Matching and the Area-Based Probability algorithm, ABP is a more mathematically approach than SPM since it introduces a Gaussian distribution modal and the Bayles' rule to find out the probability. More importantly, ABP is more robust and accurate than SPM. In SPM, sets of areas are returned from each AP and they are equally important. The resulting area is formed by intersecting these sets. However, in ABP, if the signal strengths received from a particular access point is closer to its mean value, then it will return a higher probability. In other words, more accurate signals received from one AP can compensate for other inaccurate measurements from other APs. As a result, we have chosen ABP in our project of localization.

4.2 Applying the Area-Based Approach

In order to apply the area-based approach, we divide the whole 1st floor of the Ho Sin-Hang Engineering building into different areas. We count each room as a single area and we further divide the corridor into several areas. Figure 4.3 is the floor plan of the 1st floor of the engineering building.

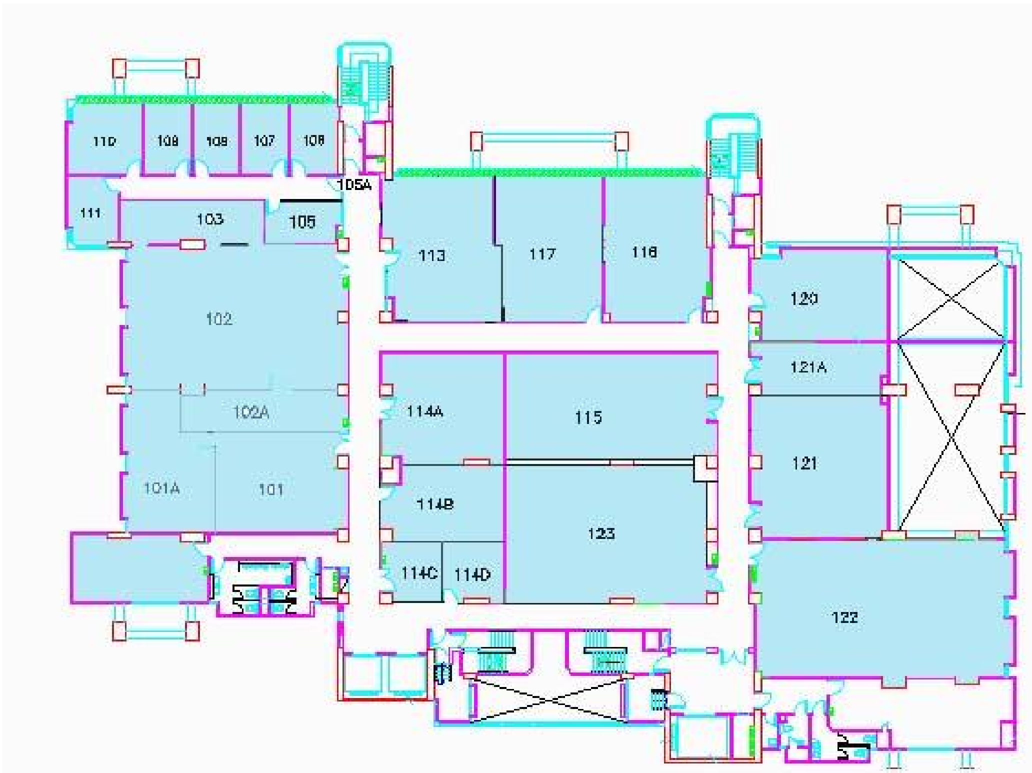


Figure 4.3

For simplicity, we only choose 12 different areas on the 1st floor to apply the localization. Area 1 is the one near the lifts, area 2 is the one near the toilet, area 3 is the one in front of Rm101, area 4 is at the North-West corner of the corridor, area 5 is near the North-West stairway on the first floor, area 6 is the one in front of Rm117, area 7 is at North-East corner of the corridor, area 8 is near the North-East stairway, area 9 is the one in front of Rm121, area 10 is at South-East corner of the corridor, area 11 is near the door of Rm123 and area 12 is at Rm101.

4.3 Generating Training Set

We have implemented a program called “Wi-Fi Location Detector”. It is a powerful tool to measure the signal received from the access points. The detail of this program is in section 6.7.

At each area chosen, we measure the signal strength from the access points for 1 minute. The access points detected depends on the location of measurement. Then, we average the samples so that we get the training set.

Figure 4.4 summarizes the fingerprints at the 12 areas. The missing data entry represent the access point is undetectable at the corresponding position.

Position	1	2	3	4	5	6	7	8	9	10	11	12
AP MAC address	Signal Strength (dBm)											
00:02:2d:28:be:9e	-70	-62	-58	-67	-73	-78	-83	-86	-84	-81	-78	-55
00:02:2d:28:be:5d	-67	-59	-60	-71	-76	-79	-81	-86	-81	-83	-79	-52
00:60:1d:1e:43:9b	-79	-87	-85	-84	-89	-80	-76	-77	-66	-63	-77	-90
00:0f:34:f3:60:40	-63	-69	-65	-74	-76	-72	-77	-84	-76	-74	-66	-79
00:02:2d:21:39:1f			-82	-78	-82	-59	-78	-73	-83	-85	-82	
00:11:93:3d:6f:c0				-90	-85	-86	-89	-88				
00:11:20:93:65:c0					-89	-89						-90
00:0f:34:bb:df:20				-89	-90	-82	-88	-88				
00:0c:ce:21:1b:9d						-87						
00:0c:85:35:33:d2					-88			-88				
00:11:20:93:63:90						-89						-88
00:0c:85:35:33:d4									-87			
00:04:76:a7:ab:a3												-90

Figure 4.4

After getting the samples, we find out the standard derivations of each access point. They are ranged from 7-10. There are totally 13 access points that can be detected on the floor, however, we do not use all of them for localization. Some of the Wireless LAN access points are located on another floor of outside the engineering building, so only very weak signals can be collected from them. We then ignore these access points because they have the least contribution to localization and also the computation time of the Area-Based Probability algorithm can be shortened. Finally, we have chosen

7 access points listed in Figure 4.5.

Name of Access Points	MAC Address
VIEWTECH AP1000 ONE A	00:02:2d:28:be:9e
VIEWTECH AP1000 ONE B	00:02:2d:28:be:5d
CSWaveLAN	00:60:1d:1e:43:9b
أأأأأ	00:0f:34:f3:60:40
CSWaveLAN	00:02:2d:21:39:1f
ERGWAVE	00:11:93:3d:6f:c0
ERGWAVE	00:0f:34:bb:df:20

Figure 4.5

For missing signal strengths, we input -92 dBm as entry. We consider -92 dBm because it is slightly lower than the weakest signal strength we have recorded. Therefore, these values can also be used by localization algorithms in calculations. After the data-processing, the data are shown in Figure 4.6.

Position	1	2	3	4	5	6	7	8	9	10	11	12
AP MAC address	Signal Strength (dBm)											
00:02:2d:28:be:9e	-70	-62	-58	-67	-73	-78	-83	-86	-84	-81	-78	-55
00:02:2d:28:be:5d	-67	-59	-60	-71	-76	-79	-81	-86	-81	-83	-79	-52
00:60:1d:1e:43:9b	-79	-87	-85	-84	-89	-80	-76	-77	-66	-63	-77	-90
00:0f:34:f3:60:40	-63	-69	-65	-74	-76	-72	-77	-84	-76	-74	-66	-79
00:02:2d:21:39:1f	-92	-92	-82	-78	-82	-59	-78	-73	-83	-85	-82	-92
00:11:93:3d:6f:c0	-92	-92	-92	-90	-85	-86	-89	-88	-92	-92	-92	-92
00:0f:34:bb:df:20	-92	-92	-92	-89	-90	-82	-88	-88	-92	-92	-92	-92

Figure 4.6

4.4 Method in Getting the Testing Set

The above data is the training set. Then the remaining step is to apply the Area-Based Probability algorithm to use the training set. Originally, we have implemented Wi-Fi Signal Scanner to collect the signal strengths manually and generate the training set. We then modify it to collect the signal strengths and generate a testing set from time to time. The following table is an example of a set of received signal strength (RRS) we get from Point 1, i.e., at the lift.

AP MAC address	Signal Strength(dBm)
00:02:2d:28:be:9e	-71
00:02:2d:28:be:5d	-72
00:60:1d:1e:43:9b	-89
00:0f:34:f3:60:40	-49

Figure 4.7

From Figure 4.7, we can see that only 4 wireless LAN access points are detected. In order to complete and generate a testing set, we need to substitute a value of -92dBm for missing data of the remaining 3 wireless LAN access points. Then the resulting testing set is shown in Figure 4.8.

AP MAC address	Signal Strength(dBm)
00:02:2d:28:be:9e	-71
00:02:2d:28:be:5d	-72
00:60:1d:1e:43:9b	-89
00:0f:34:f3:60:40	-49
00:02:2d:21:39:1f	-92
00:11:93:3d:6f:c0	-92
00:0f:34:bb:df:20	-92

Figure 4.8

In order to get a more general set of testing set, we will first collect 4 sets of signals before we apply the localization. Then, we generate the testing set by averaging the values of the 4 samples. Therefore, the default sample size of the testing set is 4.

We have collected 100 testing sets for each of the 12 locations and applied the localization. Therefore, there are totally 1200 test cases for our localization system. We will then use the result of these test cases to estimate the accuracy of our system.

We would also like to find out how the number of samples in a testing set will affect on the accuracy, so we need to get the testing sets with different sample sizes. In order to do so, we have modified the program in Wi-Fi Location Detector so that when we invoke the collection of signals once, the generated testing set will contain the average value of new and current signal strengths.

In other words, we are able to generate a testing set with larger number of samples which should be more accurate.

4.5 Finding Probabilities at Locations

4.5.1 Applying Bayles' Rule

With the testing set S_l, we need to calculate the probability of being at each location, A_i, on the floor given the testing set, i.e., $P(A_i | S_l)$. By using the Bayles' rule

$$P\left(L_i|S_l\right) = \frac{P\left(S_l|L_i\right) \times P\left(L_i\right)}{P\left(S_l\right)} \qquad (1)$$

and the equation derived in Section 3.6 talking about the Area-Based Probability Algorithm,

$$P\left(L_i|\bar{S}_l\right) = c \times P\left(\bar{S}_l|L_i\right) \qquad (2)$$

,we know that we only need to calculate the probability of having the testing set S_l when having known the current location is at L_i, i.e., $P(S_l | L_i)$ in the above equation (2) for all locations.

4.5.2 Finding the Probability $P(s_{lj} | A_i)$

$P(s_{lj} | A_i)$ is the probability of having a signal strength of s_{lj} from AP_j given the location is at A_i. In order to find these probabilities, we need to use the assumption in Area-Based Probability algorithm. In ABP, signals received from different access points are assumed to be independent. For each wireless LAN access point, the sequence of received signal strengths s_{ijk}, k = 1 . . . o_{ij} , at each location in the original training set, T_0, is modeled as a Gaussian distribution (or Normal distribution) .

The equation of Gaussian distribution is

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$$

,where μ is the mean and σ is the standard derivation.

In our application, we can take μ as the expected average signal strengths for the access point to be calculated. And in order to simplify the computation, we take σ as 8.5. x is the signal strength received from the access point in the testing set. Figure 4.9 shows the graph of an example with mean signal strength = -60dBm and standard derivation of 8.5.

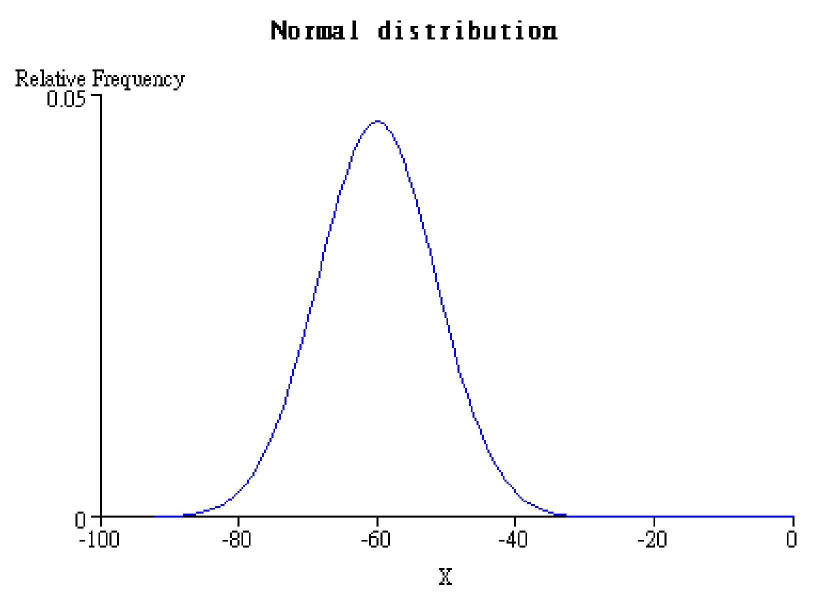


Figure 4.9

However, there is still a problem. The equation is a continuous function, but the signal strengths received are discrete values ranged from 0 to -92. Therefore, in order to use the equation, we need to define an interval. The most appropriate interval is 1 because each signal strength value should have an interval of 1 in the function of Gaussian distribution. Therefore, the probabilities can be found by the following equation:

$$\int_{x-0.5}^{x+0.5} P(x)\ dx = \int_{x-0.5}^{x+0.5} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}\, dx$$

The probabilities found by integrating the Gaussian distribution equation can be illustrated by the Figure 4.10

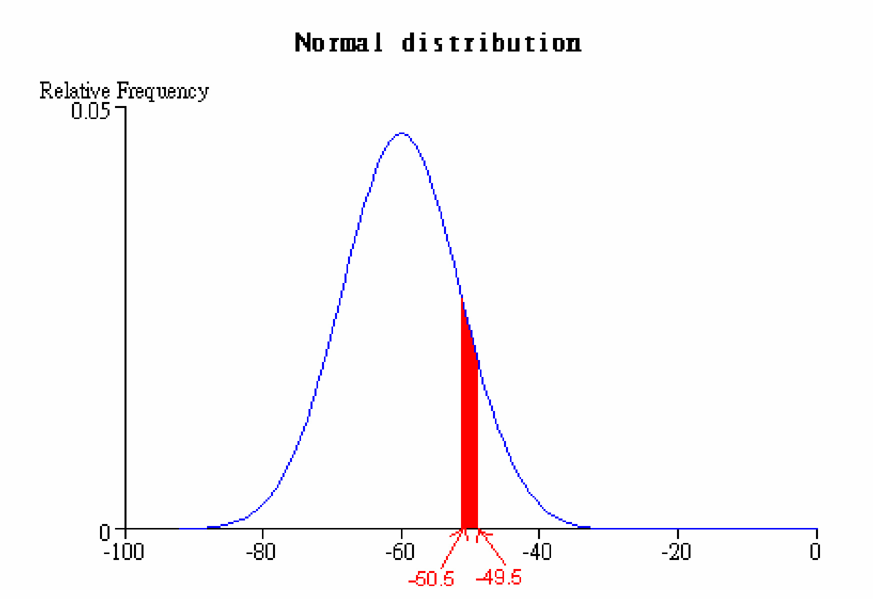


Figure 4.10

For example, when the signal strength in the testing set from a particular access point is -50dBm, the probability is represented by the red shaped area in Figure 4.10.

A standard normal distribution is a kind of normal distribution having μ=1 and σ=1. An arbitrary normal distribution can be converted to a standard normal distribution by changing variables. We let z = (x-μ) /σ and so we get the following equation.

$$P(x)\,dx = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\,dz.$$

4.5.3 Finding Integral of the Gaussian distribution function

After that, we should introduce the error function erf(z) in probability. erf(z) is the "error function" encountered in integrating the normal distribution (which is a normalized form of the Gaussian function). It is an entire function defined by

$$\operatorname{erf}(z) \equiv \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2}\,dt$$

We need to introduce the error function because it can calculate the integration of normal distribution function in terms of finite additions, subtractions and multiplications. To achieve this, we need another expression of erf(z) which can be defined as a Maclaurin series.

$$\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n z^{2n+1}}{n!(2n+1)}$$

The relationship between the normal distribution function and the error function can be summarized by the following equations:

$$\Phi(z) \equiv \frac{1}{\sqrt{2\pi}} \int_0^z e^{-x^2/2}\,dx = \tfrac{1}{2}\,\mathrm{erf}\left(\frac{z}{\sqrt{2}}\right)$$

Hence, we can compute the value of this Maclaurin series and find out the probabilities $P(s_{lj} \mid A_i)$ is the probability of having a signal strength of s_{lj} from AP_j given the location is at A_i for all i and j. We have found that if the value of signal strength is equal to the mean value of an access point, the probability calculated is approximately equal to 0.03.

4.5.4 Finding Probability $P(S_t \mid A_i)$

$P(S_t \mid A_i)$ is the probability of having the testing set S_t when having known the current location is at A_i. So, we compute $P(S_t \mid A_i)$ for all 12 locations by using the formula of independent event in probability:

$$P(S_t \mid A_j) = P(s_{t1} \mid A_j) \times P(s_{t2} \mid A_j) \times \ldots \times P(s_{tn} \mid A_j)$$

Then, we calculate the probability density based on the assumption that the mobile device must be at one of the 12 locations, i.e., $\Sigma P(A_i \mid S_t) = 1$ for all locations i. We simply sum all the probabilities of the 12 locations and then divide all of them by the calculated sum.

With this formula derived in the algorithm,

$$Max\{P(A_i \mid S_t)\} = Max\{c * P(S_t \mid A_j)\} = Max\{P(S_t \mid A_j)\}$$

We return the area or location that is with the highest probability to indicate the mobile device is currently at this area.

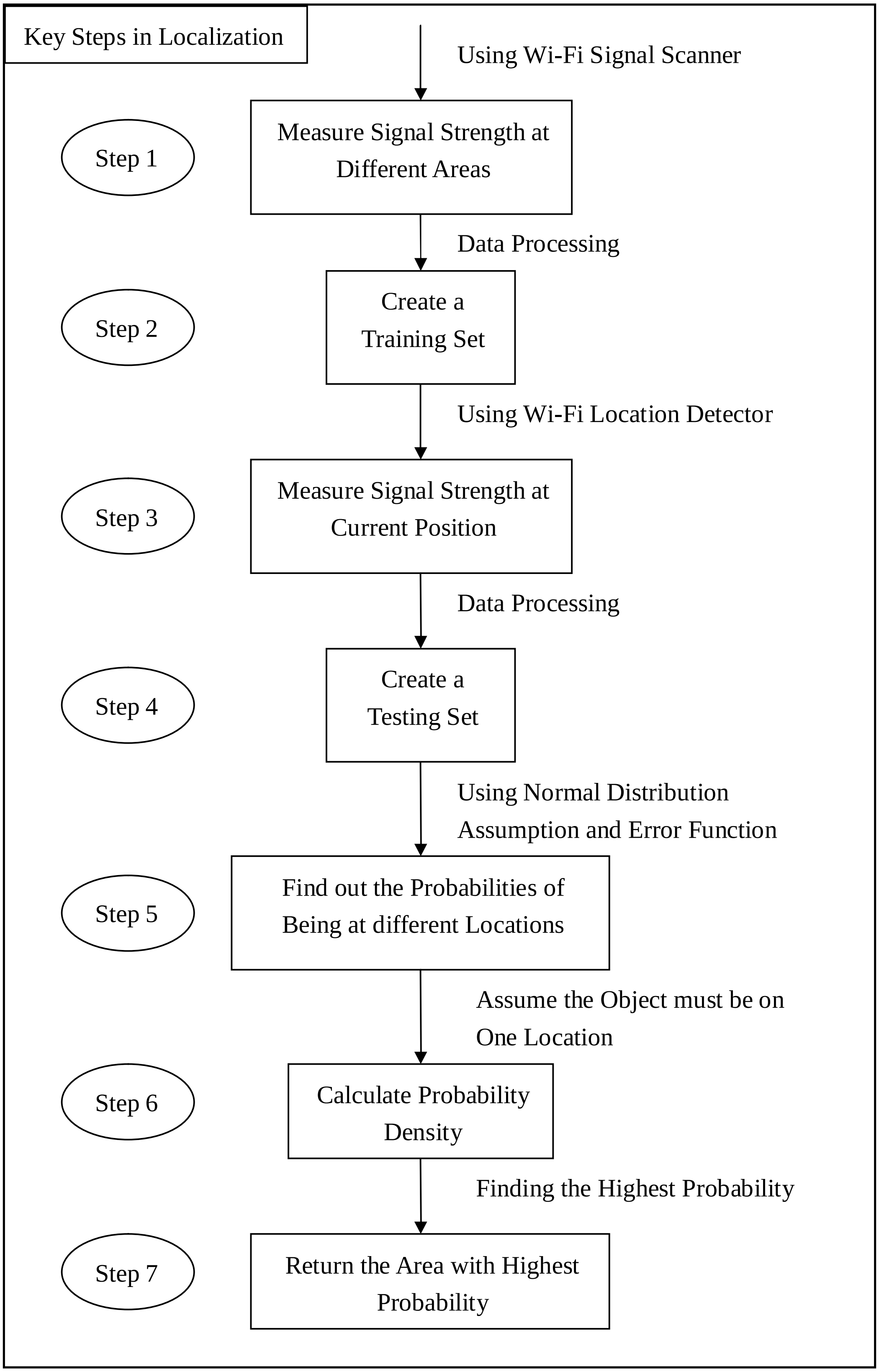


Figure 4.11 summarizes the key steps we use in localization

4.6 Experiment Result

The following 12 graphs show a general case that how the probabilities calculated by Area-Based Probability algorithm at the 12 areas change and achieve the goal of localization. Each of the graphs represents the Localization Graph at one area. There are totally 12 lines with different colors and each of them represent the probability of getting the testing set S_t given the location is at an area, i.e., $P(S_t \mid A_1)$, $P(S_t \mid A_2)$ … …$P(S_t \mid A_{12})$. The testing set is generated by averaging the signal strengths in the samples. The x-axis indicates the number of samples being used in the testing set. More samples are collected as time goes by so the x-axis can also be interpreted as a timeline. The y-axis indicates the value in terms of probability which is ranged from 0 to 1.

In ABP localization, the area with highest probability is returned, which indicates the mobile device is location at this area. In other words, within the same number of samples, the area with probability represented by a point in its line on the top is the returned area. Therefore, for example, for the Localization Graph at Area 1, the correct result should be the deep blue on the top. Otherwise, it means the result returned is incorrect.

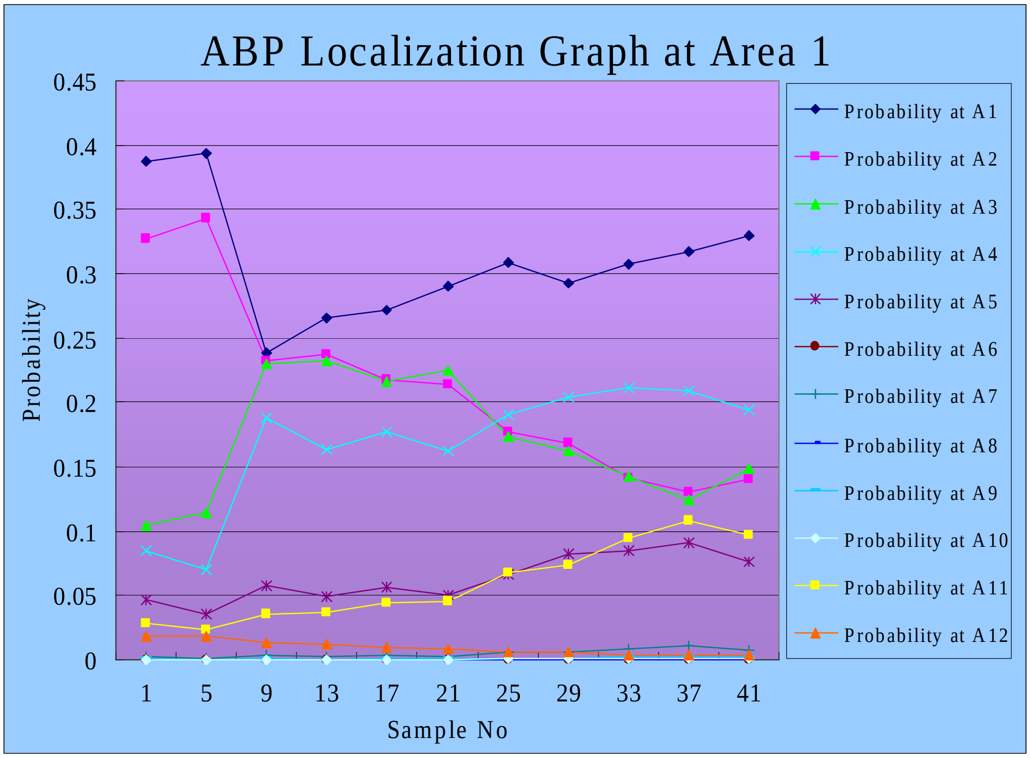


Figure 4.12

Area 1 is the one near the lift. From the graph, we can see that the deep blue line represents the probability of being at area 1. It shows the probability first starts at about 0.4 and then drops to 0.25 and finally it gradually increases to 0.33. The line is always on the top of other lines as the number of samples increases. Therefore, our system always indicates the object is at area 1 which is the correct result.

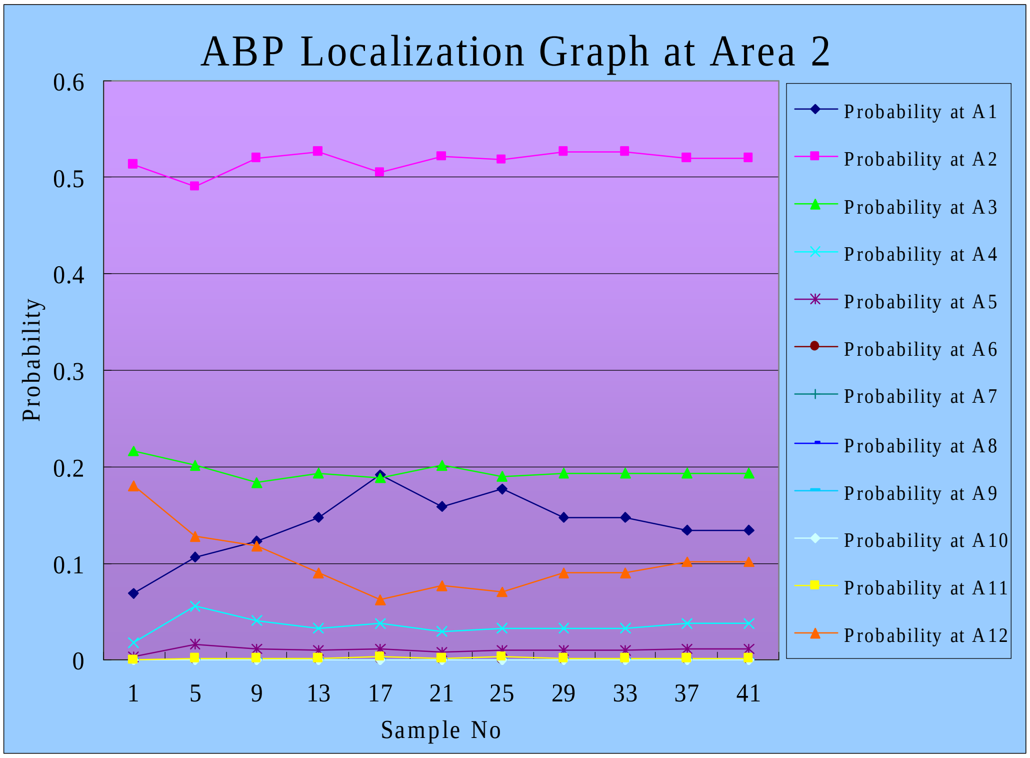


Figure 4.13

Area 2 is the one near the toilet. From the graph, we can see that the purple line representing the probability of being at area 2 is always on the top of other lines as the number of samples increases. The probability is stable and always stays around 0.5. For other probabilities, they are quite stable and always stay below a value of 0.2. Therefore, our system always indicates the object is at area 2 which is the correct result.

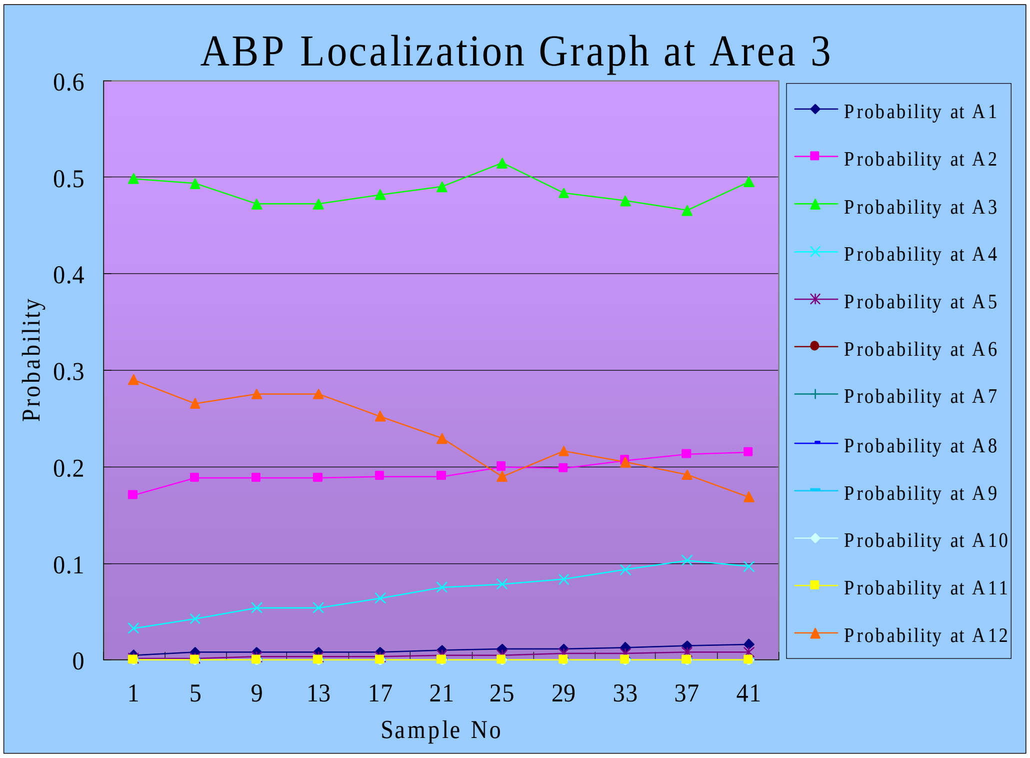


Figure 4.14

Area 3 is the corridor in front of Rm101. From the graph, we can see that the green line representing the probability of being at area 3 is always on the top of other lines as the number of samples increases. The probability is quite stable and is always around 0.5. For other probabilities, most of them are quite stable and all of them always stay below a value of 0.3. Therefore, our system always indicates the object is at area 3 which is the correct result.

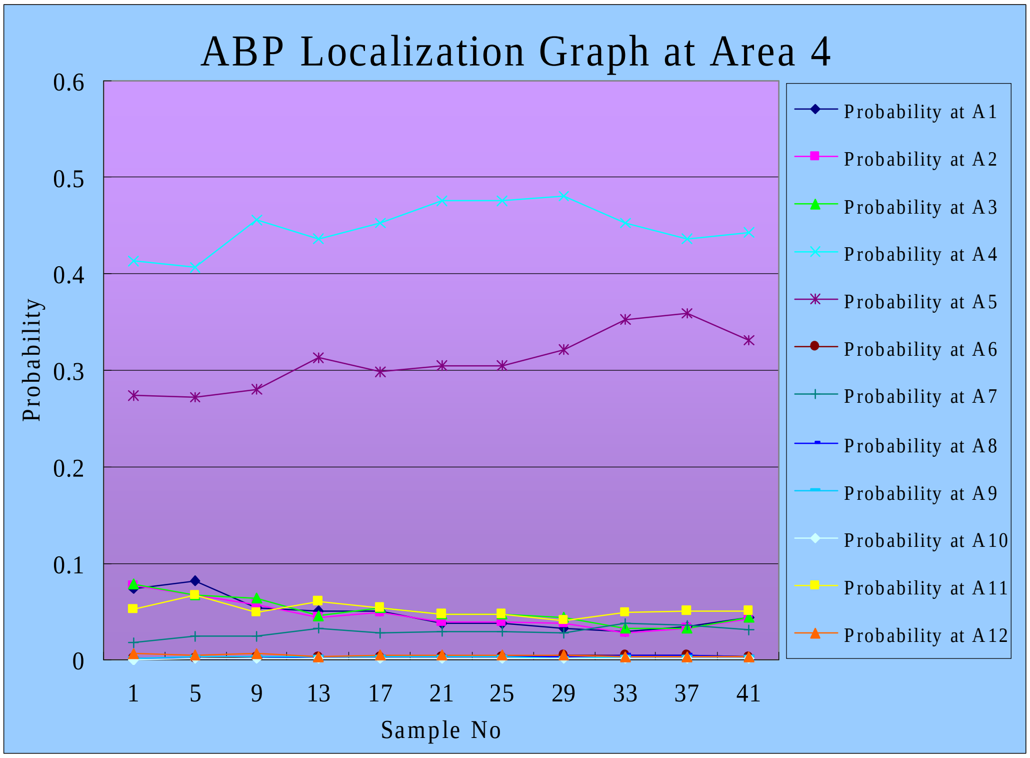


Figure 4.15

Area 4 is the North-West corner of the corridor. From the graph, we can see that the light blue line representing the probability of being at area 4 is always on the top of other lines as the number of samples increases. The probability is quite stable and is always over 0.4. The probability of being at area 5 remains at a comparatively high value around 0.3. For other probabilities, they are quite low, stable and always stay below a value of 0.1. Therefore, our system always indicates the object is at area 4 which is the correct result.

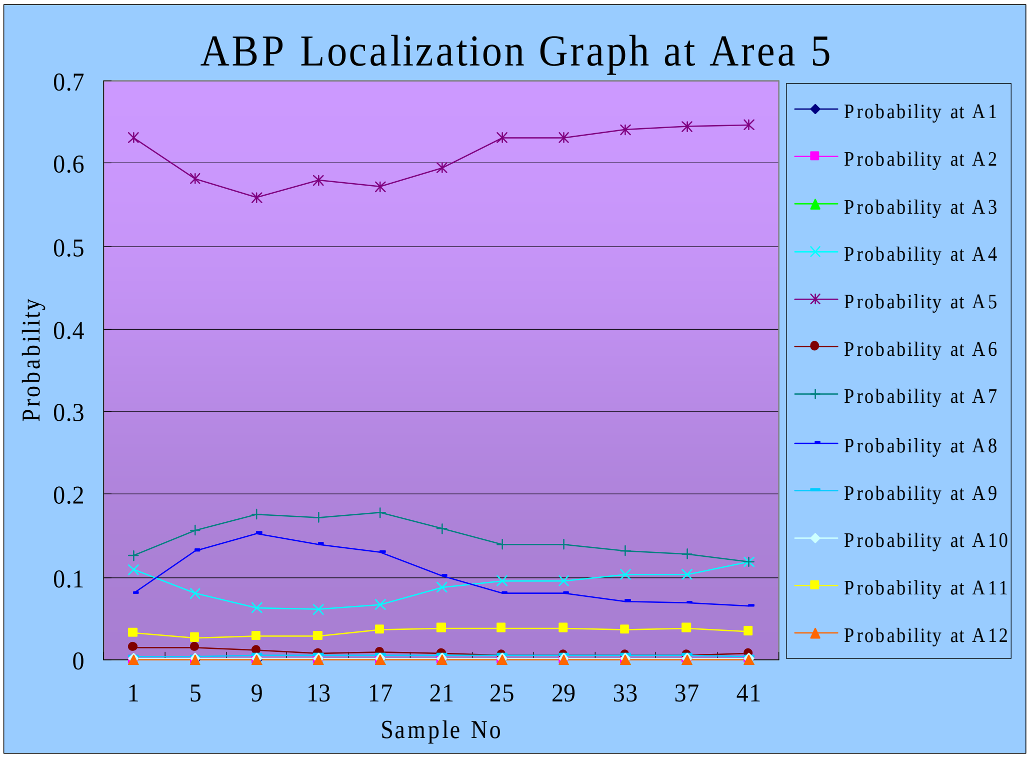


Figure 4.16

Area 5 is near the North-West stairway on the first floor. From the graph, we can see that the deep purple line representing the probability of being at area 5 is always on the top of other lines as the number of samples increases. The probability is quite high and is always around 0.6. For other probabilities, they are quite stable and always stay below a value of 0.2. Therefore, our system always indicates the object is at area 5 which is the correct result.

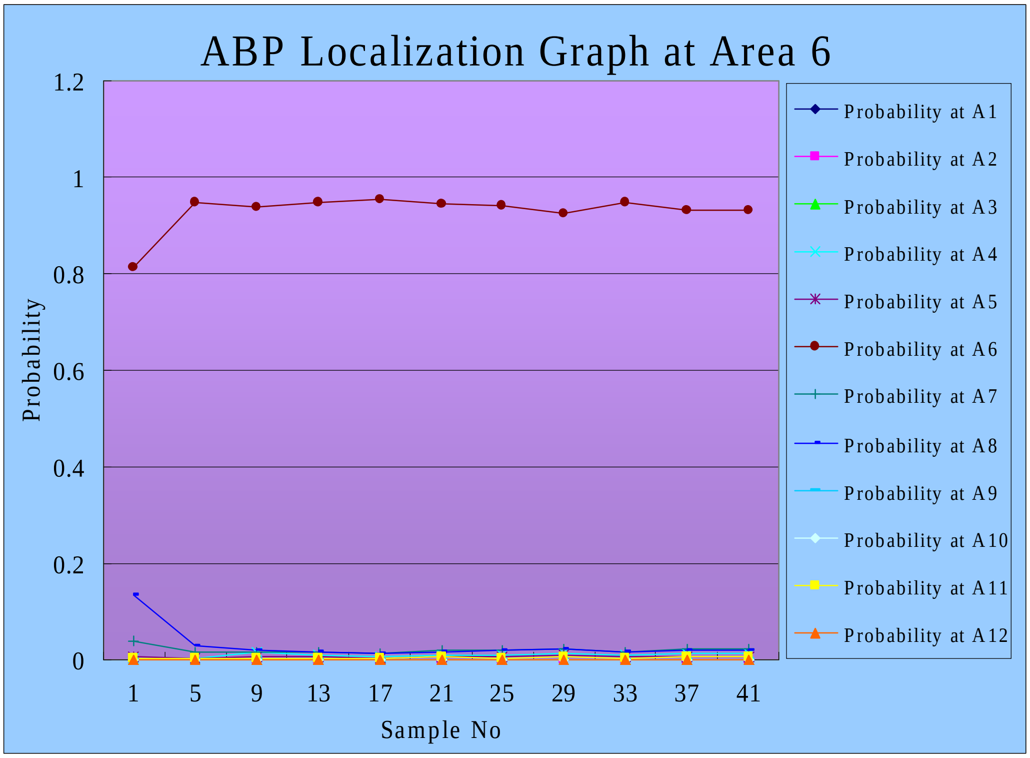


Figure 4.17

Area 6 is the one in front of Rm117. From the graph, we can see that the brown line representing the probability of being at area 6 is always on the top of other lines as the number of samples increases. The probability is very high and is usually around 0.9. For other probabilities, they are very low, stable and always have a value near 0. Therefore, our system always indicates the object is at area 6 which is the correct result.

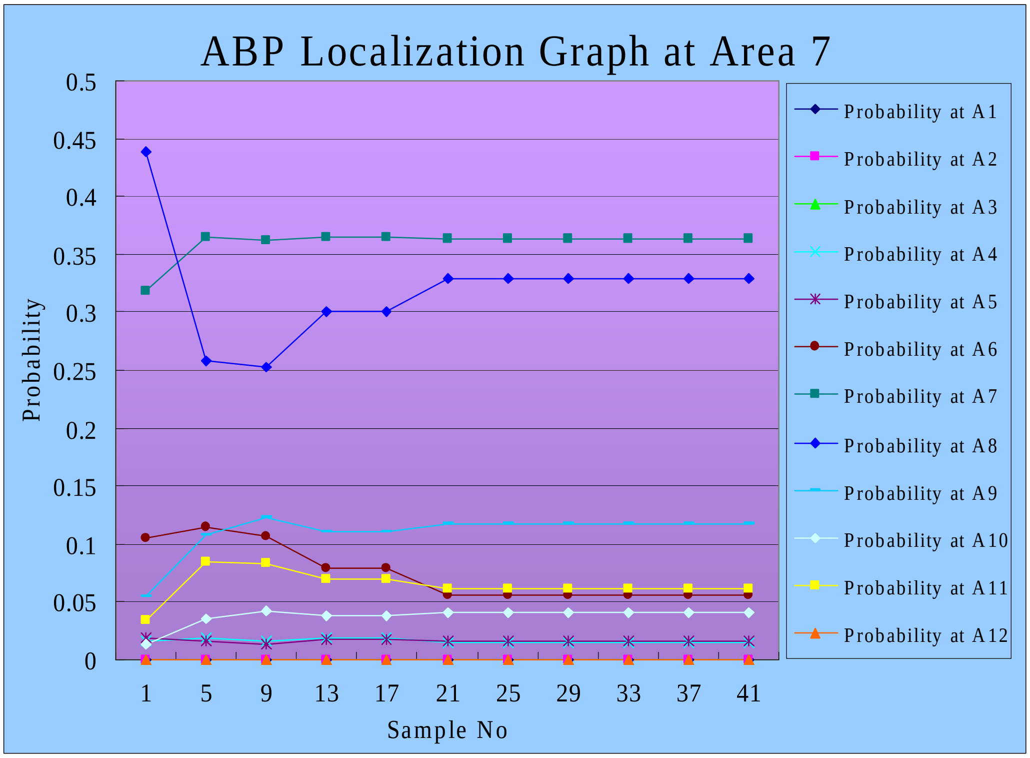


Figure 4.18

Area 7 is the North-East corner of the corridor. From the graph, we can see that the deep green line representing the probability of being at area 7 is below the blue one when the number of samples equals to 1. As the number of samples increases, the probability of deep green line increases and remains at a value of 0.36 while the probability of blue line drops and lies below it. Therefore, our system first indicates the object is at area 8 which is an incorrect result and then indicates the object is at area 7 when the number of samples reaches 5.

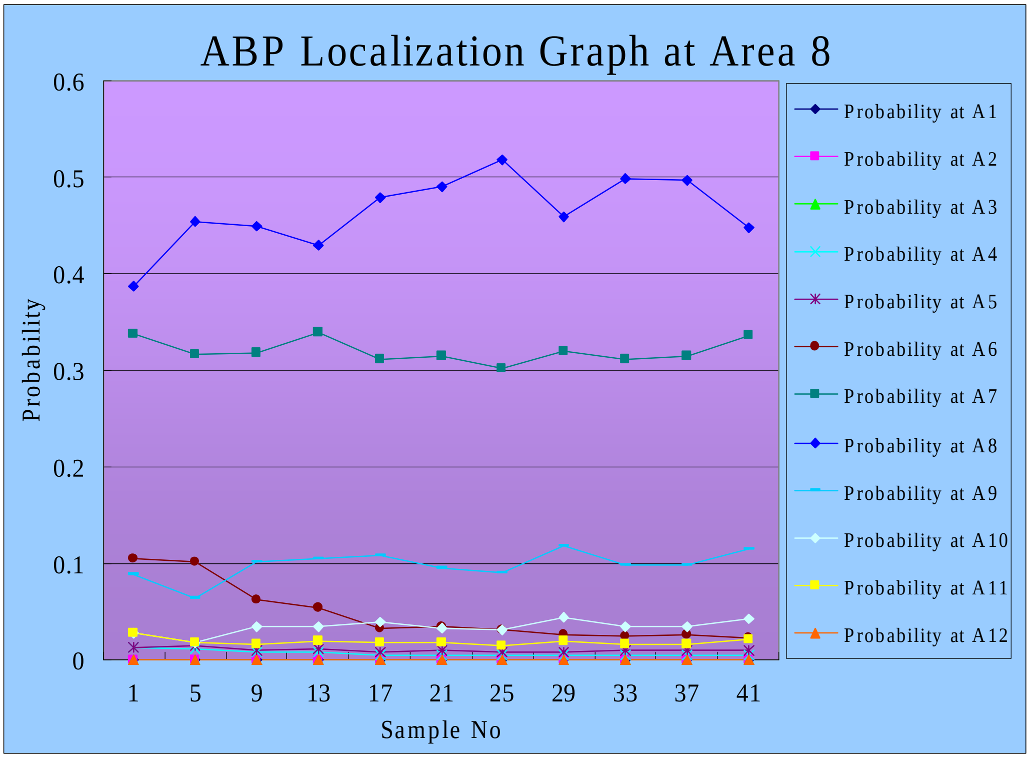


Figure 4.19

Area 8 is near the North-East stairway on the first floor. From the graph, we can see that the blue line representing the probability of being at area 8 is always on the top of other lines as the number of samples increases. The probability is quite high and is always around 0.45. The probability of being at area 7 remains at a comparatively high value over 0.3. For other probabilities, they are quite low, stable and always stay below a value of 0.12. Therefore, our system always indicates the object is at area 8 which is the correct result.

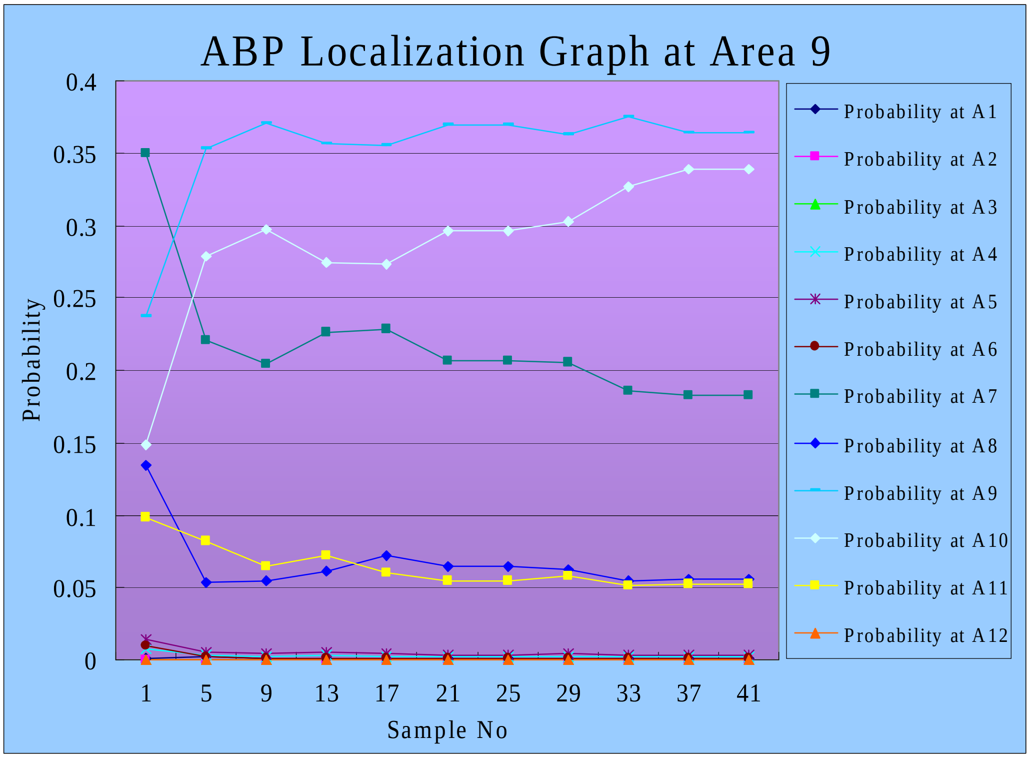


Figure 4.20

Area 9 is the area in front of Rm121. From the graph, we can see that the light blue line representing the probability of being at area 9 is below the deep green one when the number of samples equals to 1. As the number of samples increases, the probability of the light blue line increase and remains at a value over 0.35 while the probability of deep green line drops and lies below it. Therefore, our system first indicates the object is at area 7 which is an incorrect result and then indicates the object is at area 9 when the number of samples reaches about 5.

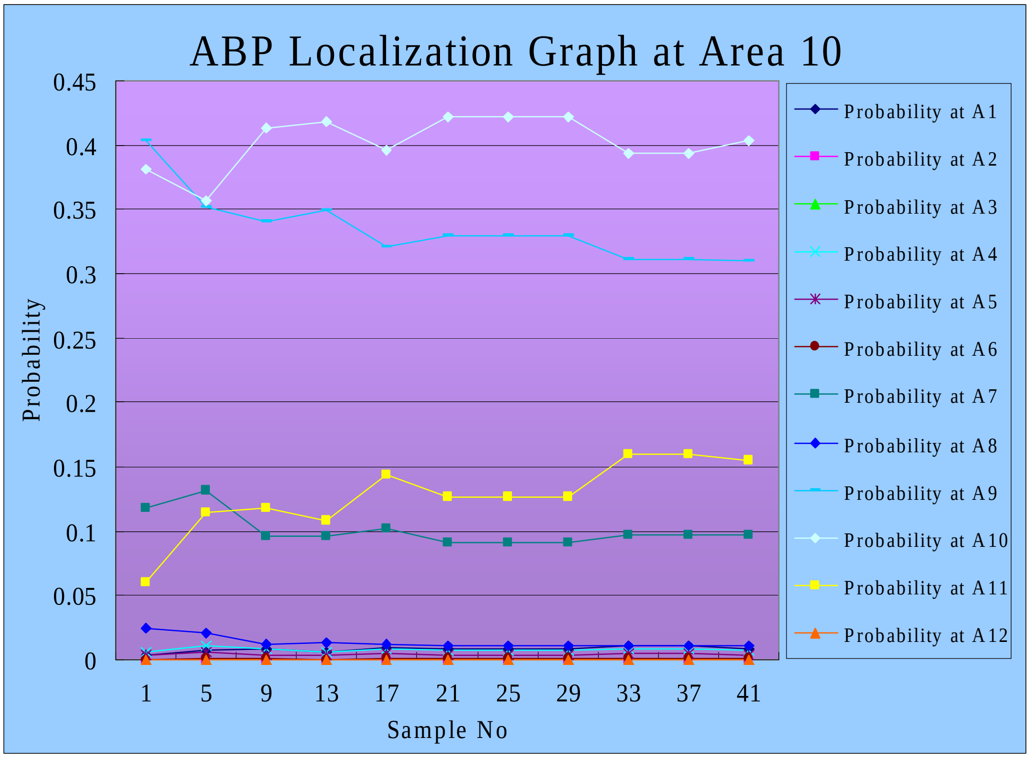


Figure 4.21

Area 10 is South-East corner of the corridor. From the graph, we can see that the pale blue line representing the probability of being at area 10 is below the light blue one when the number of samples equals to 1. As the number of samples increases, the probability of the pale blue line increase and remains at a value around 0.4 while the probability of light blue line drops and lies below it. Therefore, our system first indicates the object is at area 4 which is an incorrect result and then indicates the object is at area 10 when the number of samples reaches about 5.

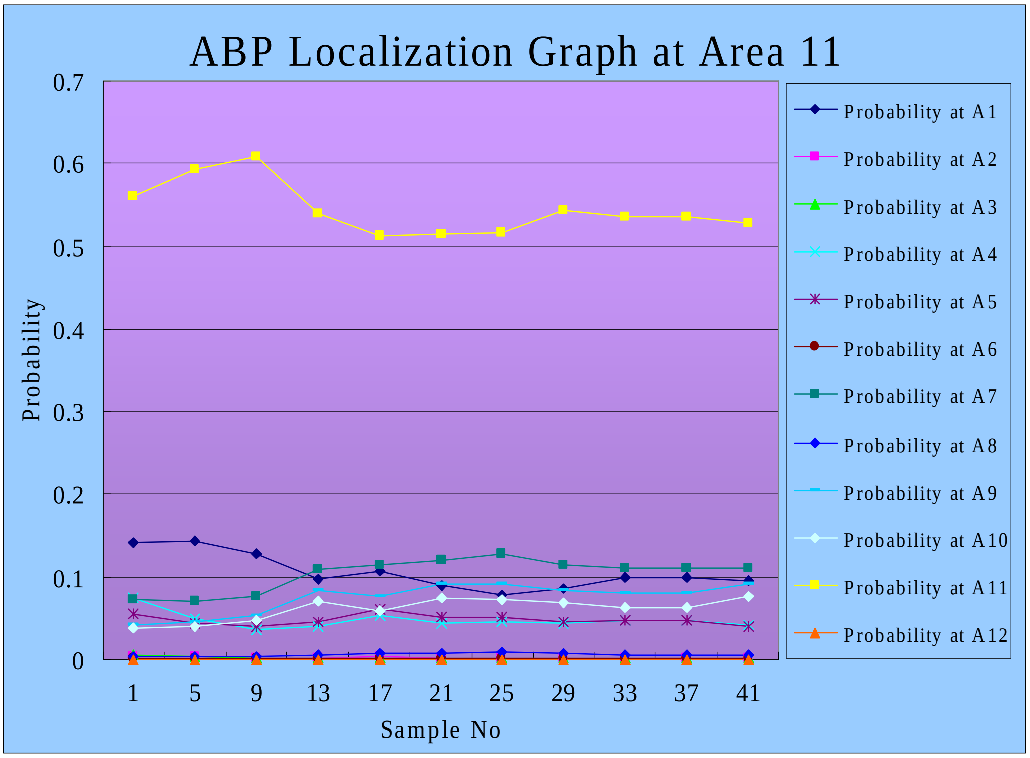


Figure 4.22

Area 11 is the one near the door of Rm123. From the graph, we can see that the yellow line representing the probability of being at area 11 is always on the top of other lines as the number of samples increases. The probability is quite high and is always over 0.5. For other probabilities, they are quite low, stable and always stay below a value of 0.15. Therefore, our system always indicates the object is at area 11 which is the correct result.

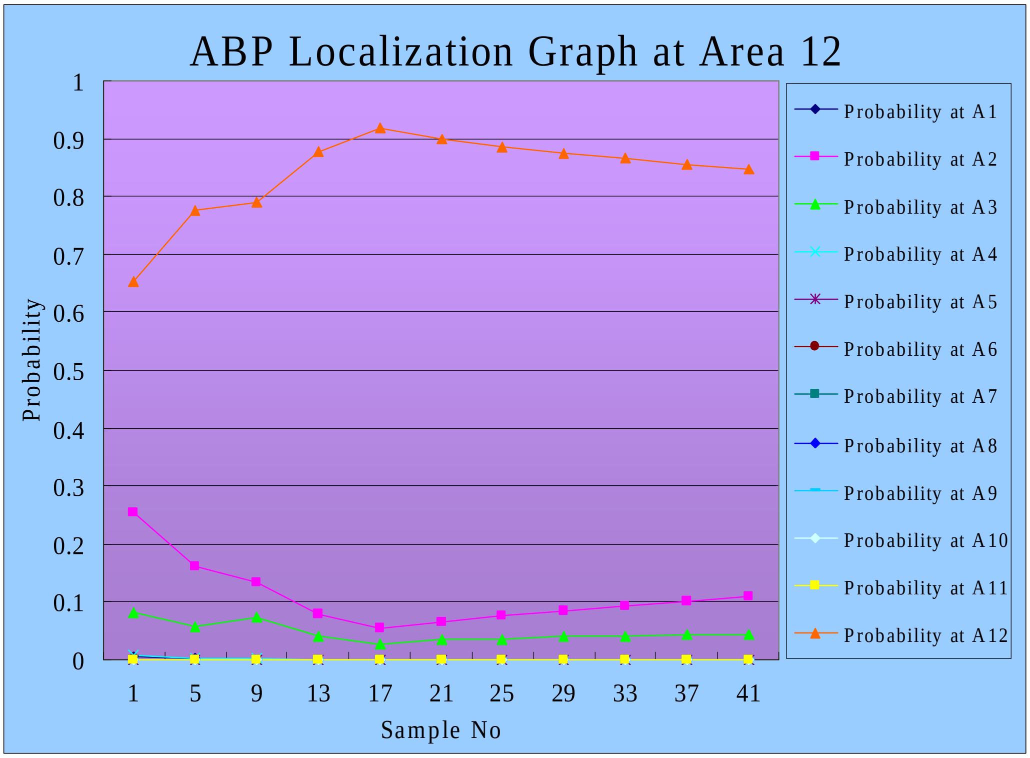


Figure 4.23

Area 12 is at Rm101. From the graph, we can see that the orange line representing the probability of being at area 12 is always on the top of other lines as the number of samples increases. The probability is quite high and remains over 0.8 as the number of samples reaches 13. For other probabilities, they are quite low, stable and always stay below a value of 0.15 as the number of samples increases. Therefore, our system always indicates the object is at area 12 which is the correct result.

5. Analysis of Experiment Results

5.1 Performance of Area-Based Probability Algorithm

The performance of ABP algorithm is good. It usually returns the correct location with a high probability over 0.3 and for most of other locations, it usually return a probability as low as 0.001. Therefore, it works well in distinguishing different locations.

However, when two areas are close together, the signal strengths collected from one area is similar to that of the other. In this case, the algorithm usually returns comparable probabilities for these two areas. Therefore, the algorithm may fail to give the correct location.

There is a tradeoff between time and accuracy in localization. A testing set is usually more accurate if its sample size is larger, but it requires a longer time to collect the signals. ABP algorithm has an acceptable accuracy even though the sample sizes of the testing sets are small. In other words, ABP algorithm can be applied to some location-based applications which require real-time localization.

5.2 Accuracy of our localization system

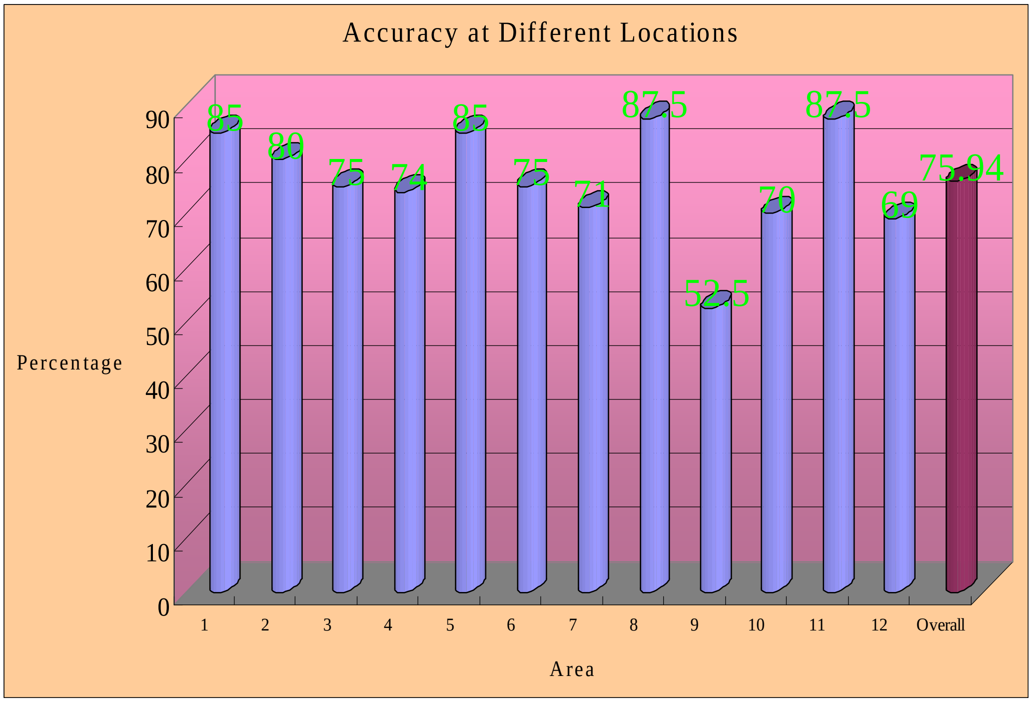


Figure 5.1

Figure 5.1 shows the accuracy of our system at different locations. The accuracy is in terms of the percentage of our system returning a correct location during localization. We have taken 80 testing sets for each of the 12 locations and test on the result of our system. The overall accuracy of our system is about 75.94%. That means on average, our system can indicate 76 correct locations when taking 100 trials.

We have found that the accuracy varies with different locations. It is because the localization algorithm will depends on the testing set generated at different locations. The accuracy at area 9 is particularly low because the distance between area 9 and area 10 is too close. Therefore, the signal strengths received from area 9 is very similar to those from area 10. Hence, our system may fail to distinguish between the two locations.

5.3 The effect of number of samples on accuracy

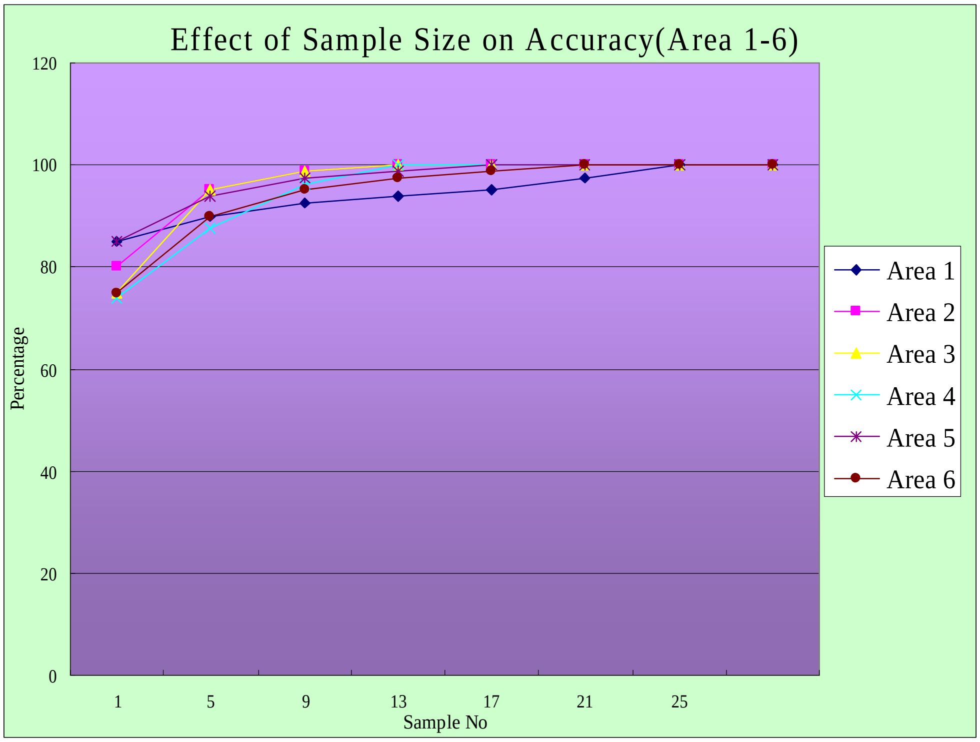


Figure 5.2

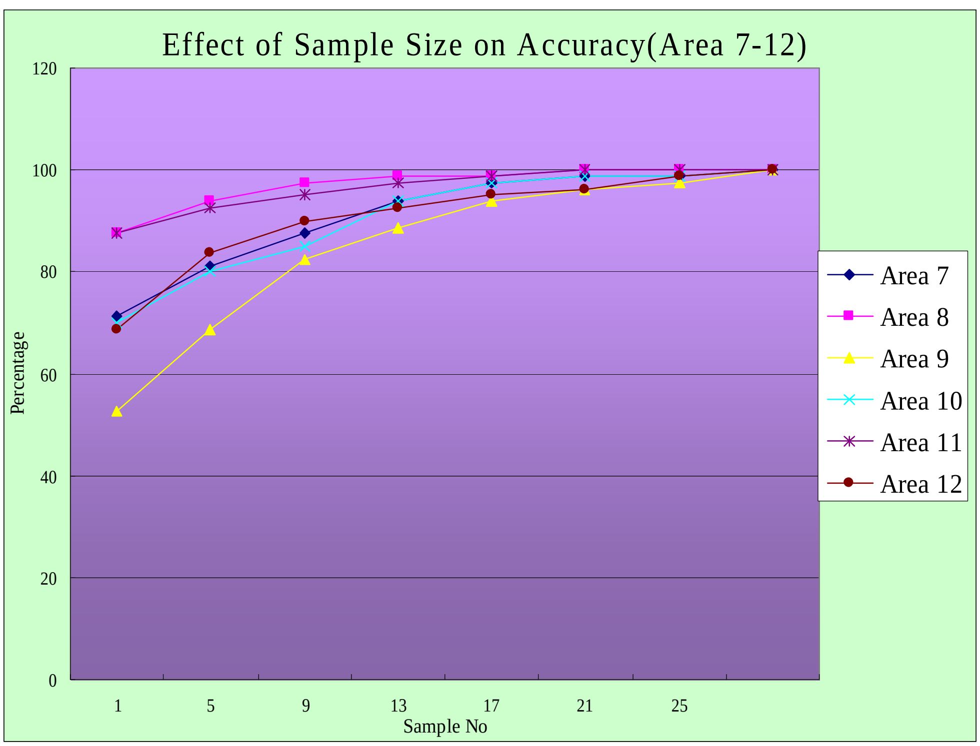


Figure 5.3

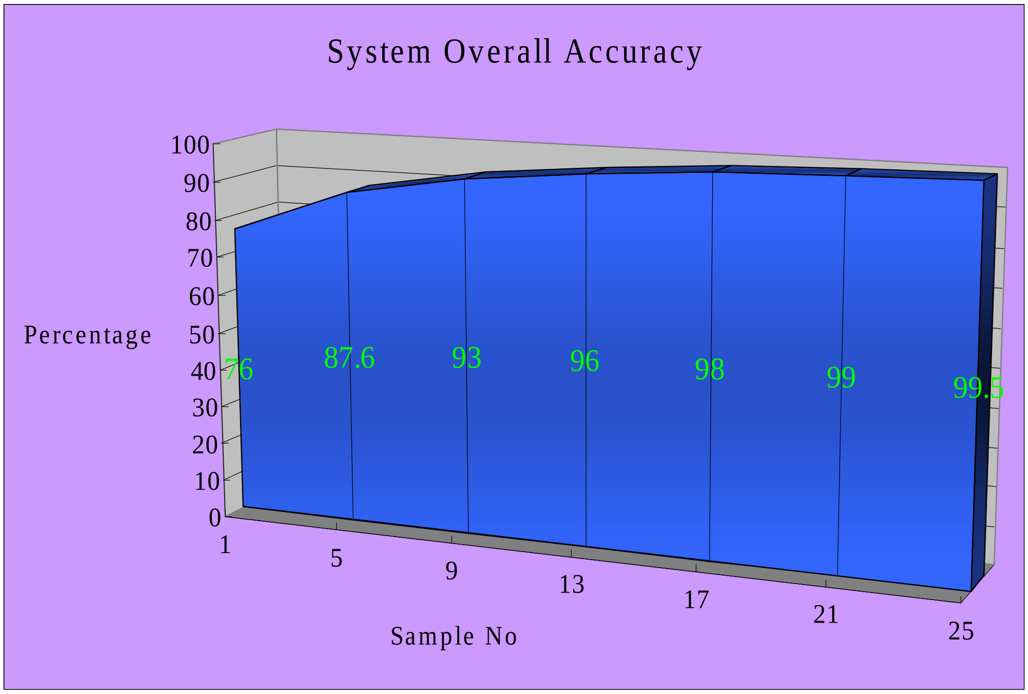


Figure 5.4

Figure 5.2, Figure 5.3 and Figure 5.4 show how the accuracy changes with the number of samples included in a testing set. Figure 5.2 and Figure 5.3 show the accuracy changes at different locations while Figure 5.4 represents the overall accuracy change of the system. We can conclude that the accuracy increases with the sample size. It first increases significantly and finally levels off when the sample number equals 21.

The accuracy will never reach 100% because there are some other factors affecting the accuracy. It will be discussed in the next section.

5.4 Other factors affecting the accuracy

5.4.1 Hardware failure

When any of the access points fails to give out signals or give out signals at unusual strength, the signals received from the mobile device will be different. Therefore, the accuracy of the system will decrease.

5.4.2 Change in environment

When there is one or more access points in the floor, its signals will cause the disturbance for those of original access points.   Also, when someone opens the door or change the location of some large enough objects in the floor, it will affect the signals received. Because the signals may be originally weaken by the doors or these objects.

5.4.3 Orientation in collecting signal

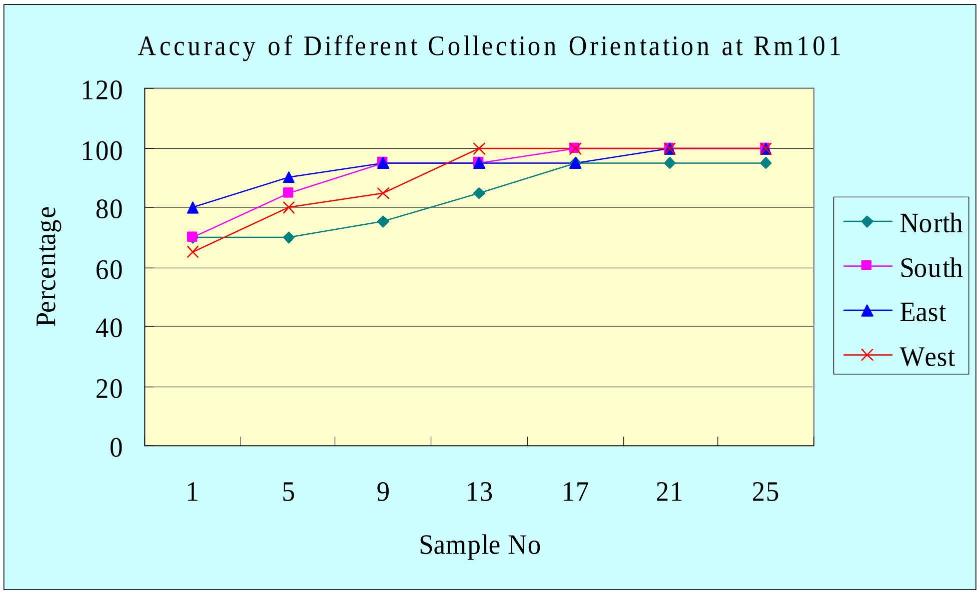


Figure 5.5

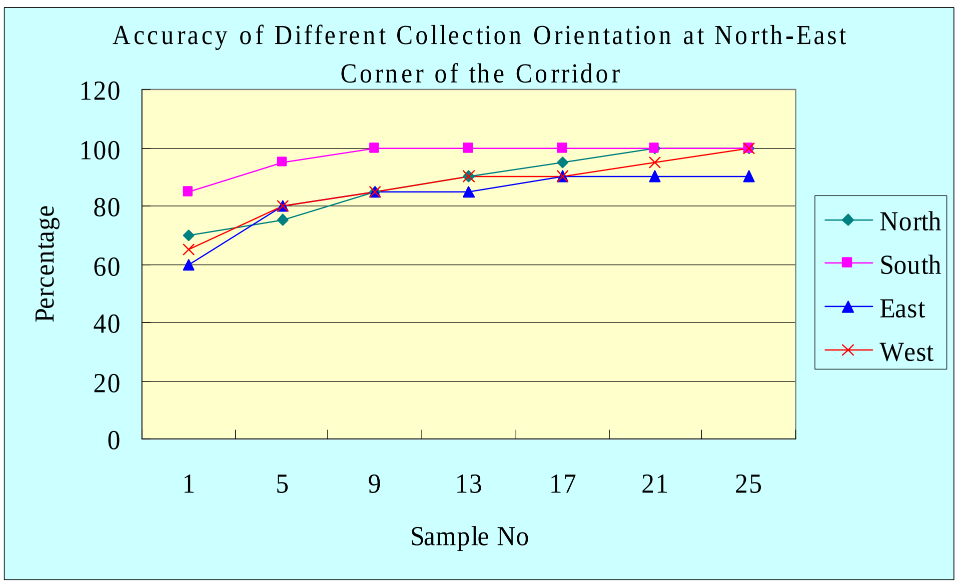


Figure 5.6

Figure 5.5 and Figure 5.6 show the accuracy of localization when using different orientation in collecting the signals. The accuracy is measured at Rm101 and the North-East Corner of the 1st floor corridor respectively. We can see that there are slightly differences in the accuracies for different orientations. The orientation in collecting signals has some but not significant effect on localization.

6. Wi-Fi Location System (WLS)

6.1 Introduction

Wi-Fi Location System is a tool that helps non-professional developers to deploy location-based system. With this tool, a developer can deploy the Location-Based System within a short period of time. It simplifies the steps and increases the efficiency and productivity in implementing the System.

The System consists of three components: Wi-Fi Signal Scanner (WSS), Wi-Fi Data Processor (WDP) and Wi-Fi Location Detector (WLD).

Wi-Fi Signal Scanner
This is a tool installed in a mobile device to collect the data from access points at target place. It records the mean of signal strength received by the mobile device. After collecting the data from all the target position, the data is then being analyzed in the WDP.

Wi-Fi Data Processor
The main function of Wi-Fi Data Processor (WDP) is to generate Location-Based Data (LBSData) which is being used in the Wi-Fi Location Detector. Before generating LBSData, some useless data collected from the WSS need to be filtered out. This is to simplify the computation in WLD. Specific information of each position need to be set in the tool, such as name, position in the Map Picture. This tool provides a user-friendly interface to simplify the procedure in constructing the LBSData.

Wi-Fi Location Detector
It is the core part of the program in the WLS. The developer can deploy the WLS with the LBSData and WLD. The WLD then provide the Location-Based Service within the WLAN network at the target place. WLD shows the Location Information to client. A user can be located by our System.

With these 3 tools, a developer can implement the Wi-Fi Location System in an efficient way. And it reduces the complexity in deploying the System.

6.2 Deploying Procedure

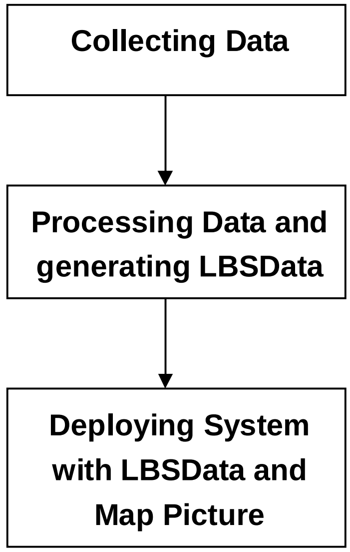


Figure 6.1

Collecting Data (Step 1):
First, the developer needs to collect the source data from the Wi-Fi network at target place. At each target position, it must collect mean strength signal data from Access Points. The data then are processed in the next step and used in localization algorithm.

Processing Data and generating LBSData (Step 2)
Second, the developer needs to process the source data and filter out some useless data (noise data). Normally, WSS collects some weak signal from outside target place or failed Access Point. This noise data affects the accuracy of localization algorithm. The data must be filtered out manually. The name of the position and the actual location in the Map Picture will be shown in the WLD. They should be set in this step. After processing the data, LBSData is generated by WDP.

Deploying System with LBSData and Map Picture (Step 3)
After generating the LBSData, it is used as data source for the WLD to perform the normal function. The Map Picture is used as the media to show the actual position in the target place on the screen. The WLD will estimate the actual

position according to the LBSData. The client can use WLD to locate his position at the target place.

6.3 Comparison between Tradition Development Procedure and the Wi-Fi Location System Development Procedure

6.3.1 Tradition Development Procedure (TDP)

TDP is mainly divided into three steps:

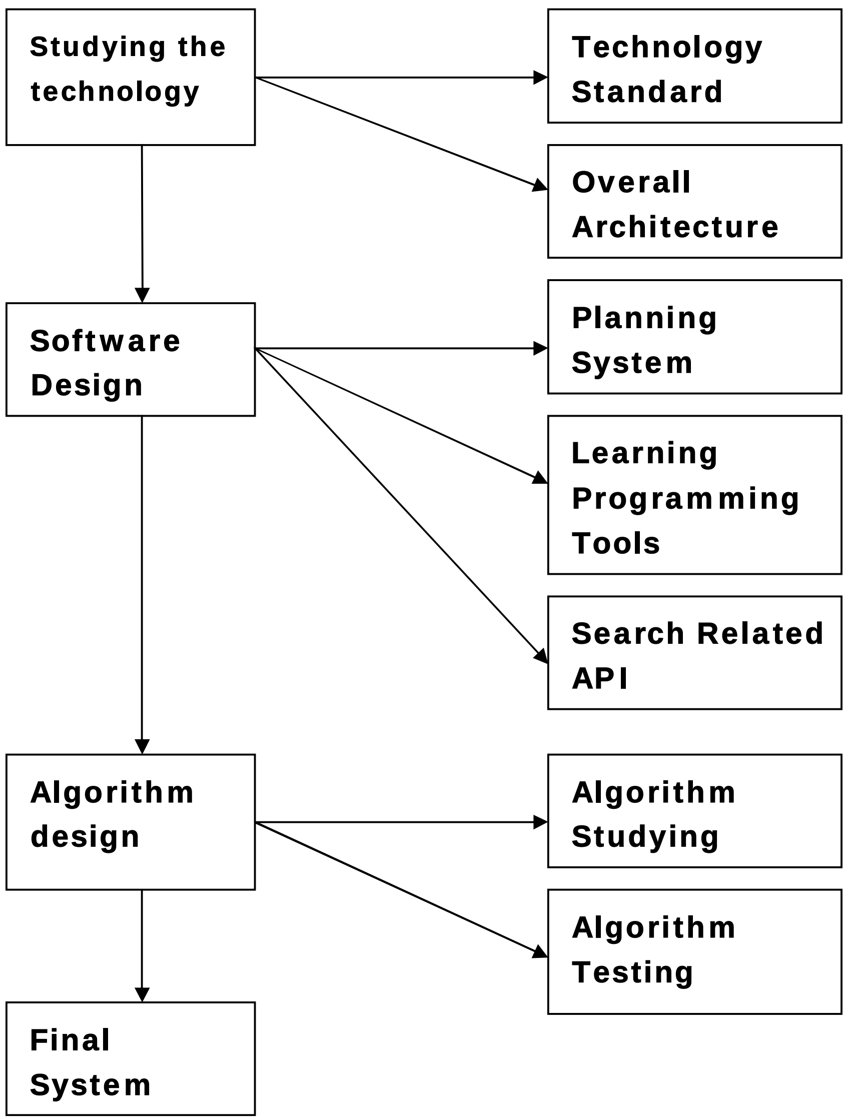


Figure 6.2

In the TDP, the developer needs to study the Wi-Fi network and lots of lower level technology, such as network standard and network architecture. They need to fully understand the technology before planning the system. They may need to have base knowledge in network. The developer must be a professional in network field.

After that, developer need to study related Application Program Interface (API) before implementing the Location-Based System. However, only several programming languages can support the Location-Based Development. They need to waste much time in finding target tools for development. Sometimes, they target tools may difficult to find. After finding the target tools, they need to learn to use the tools before implementing the Location-Based System. They must use much time in finding the method to implement the System. The developer must have experience in software design such that they can do this work well.

After finding the related tools and technology, they need to study the algorithm for the system. There are lots of algorithms for locating position at target place. The algorithms vary from the technology and situation. Sometimes, the algorithm is not efficient enough, so they need to modify it. They also need to spend time in understanding the algorithm and need to test the algorithms when implementing system.

In conclusion, the developer wastes lots of time in planning, designing and implementing the Location-Based System. It may require the developer to be professional in many fields, such as networking, software engineering and algorithm studying. Here is the summary of TDP:

a. Study the technology of the Location-based System
b. Design the software architecture and find the program tools
c. Study algorithm in implementing System

The developer need to waste lots of time in deploying Location-Based System. We have estimated the developer may need to waste 2-3 weeks to finish implementation of the System and spend 1 more week for testing.

6.3.2 Wi-Fi Location System Development Procedure (WLP)

Wi-Fi Location System Development Procedure is divided into 3 steps. Less work would be done in each step. And the work is more simple.

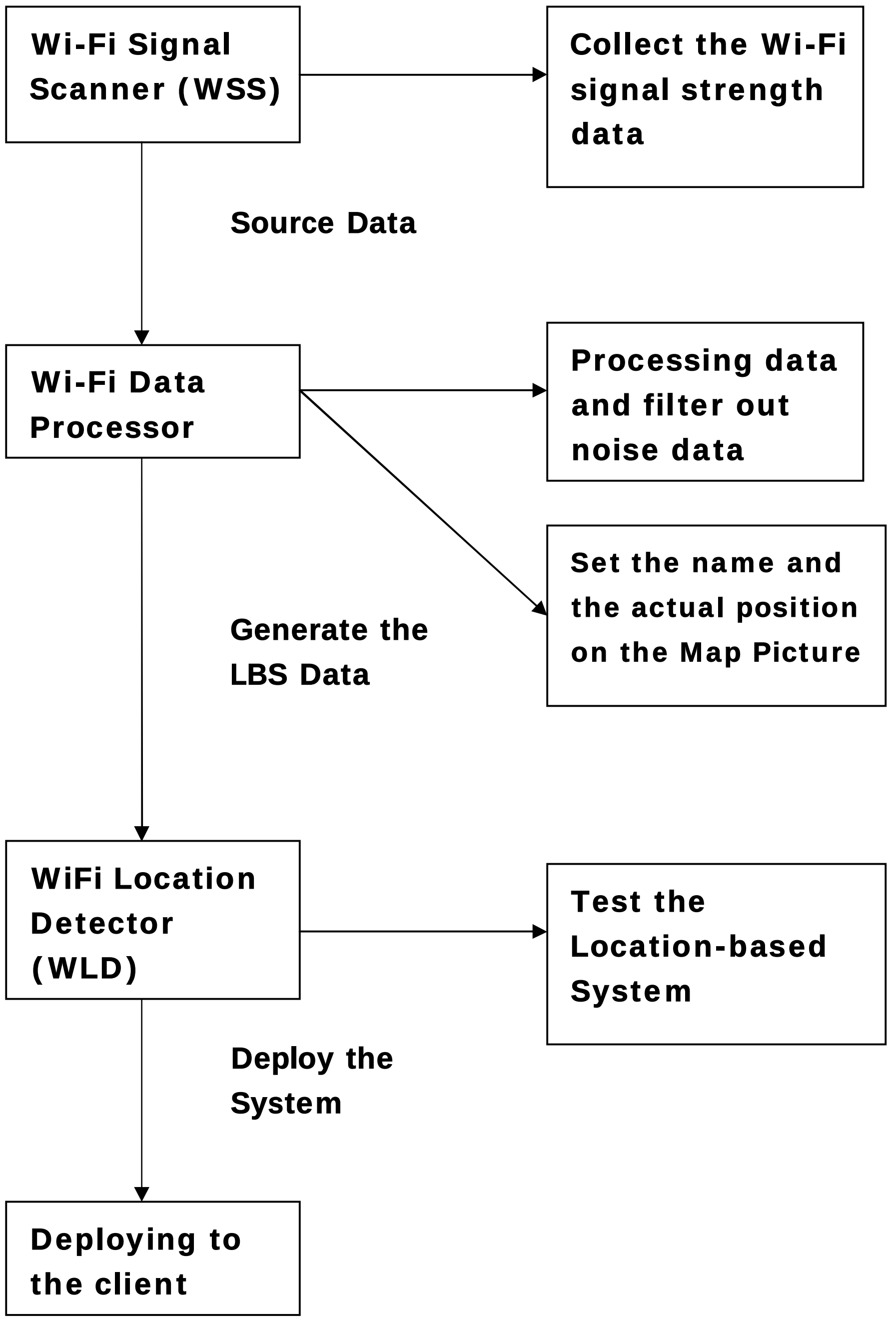


Figure 6.3

In WLP, the developer only need to follow the procedures .Then he/she can deploy the system in a short period of time. The WLP is a very efficient way to develop Location-based System.

First, the developer needs to collect the source data from the target place. The source data is then processed in the next step and is used in the location algorithm. The developer needs to collect the mean of signal strength in each target position. And the WSS atomically store the source data in the mobile device. Then the data is passed to WDP for further processing. This work is pretty simple. It only trains the developer to use this tool. No specific knowledge requirement is needed.

Second, the developer then processes the source data. There are some data collected in the WSS containing some noise. The noise data may come from the weak signal from other buildings or places. It needs to be filtered out to increase the accuracy of localization algorithm. The noise data may also affect the system performance if the noise data is too much. The data needs to be careful filtered out in order to make the system work well. The developer only needs to know the filer technique. Therefore, this step is also simple.

After that, the developer sets the name of target position and actual position in the Map Picture. These things will be shown in the WLD. WLD tells the location information through this graphics interface. These must be set in the step. Since the WDP contains a fancy user interface, the developer can done this without knowledge about the programming. They can finish this job by click the buttons only.

The Location-Based Data (LBSData) is generated in the previous step. There are two main data sources in the WLD. They are LBSData and the Map Picture. The localization algorithm uses the mean signal strength to estimate the location of the client at the target place. After estimation, WLD will show the estimated position in the mobile device through the fancy interface. There would be a point shown in the Map Picture. The developer only needs to perform testing on the system. If system does not work fine, it may need to collect the data once more and perform the step 1 and step 2 again.

In conclusion, the WLP is mainly divided into four procedures:
a. Collect the Source Data using the WSS
b. Process the data using the WDP
c. Test the System with WLD
d. Deploy the System

Using the WLP, the developer can develop the program within 1-3 day. It reduces the time of implementing the system. The development can also be done by non-professional people. It much increases the productivity and efficiency of the Location-based System development.

6.4 Features of Wi-Fi Location System (WLS):

a. Scalability:
The WLS can apply in any flexible environment. In case the change of environment, it can generate the System with perform the procedure once. The new System can be run in the new environment. WLS can be scaled at any time and anywhere.

b. Abstraction
The WLS would do all the prepared work for the developer. And most of prepared work is only the low level technology studying and research work. The developer only focuses on how to increase the accuracy of System rather than studying the prepared work. The prepared work is transparent for the developer.

c. Transparent
No matter what changes in the WLS, they can still deploying the system without any problem. Even though the core part of the WLS system is changed, it only affects core part, such as algorithm. The data source remains unchanged in the WLS.

d. User-friendly
The three components are made by using fancy user interface. They are very easy to use. The developer only needs to know how to use them correctly. It can be used by anyone with basic computer knowledge. They can handle the WLS well.

The WLS is a powerful and user friendly system. It can deploy the Location-based System within a short period of time.

6.5 Wi-Fi Signal Scanner (WSS)

Wi-Fi Signal Scanner is a tool to collect the signal strength received at the target position. The data then are used in the WDP. We will detail explain the WSS's architecture, method of usage in the following section.

6.5.1 Software Architecture

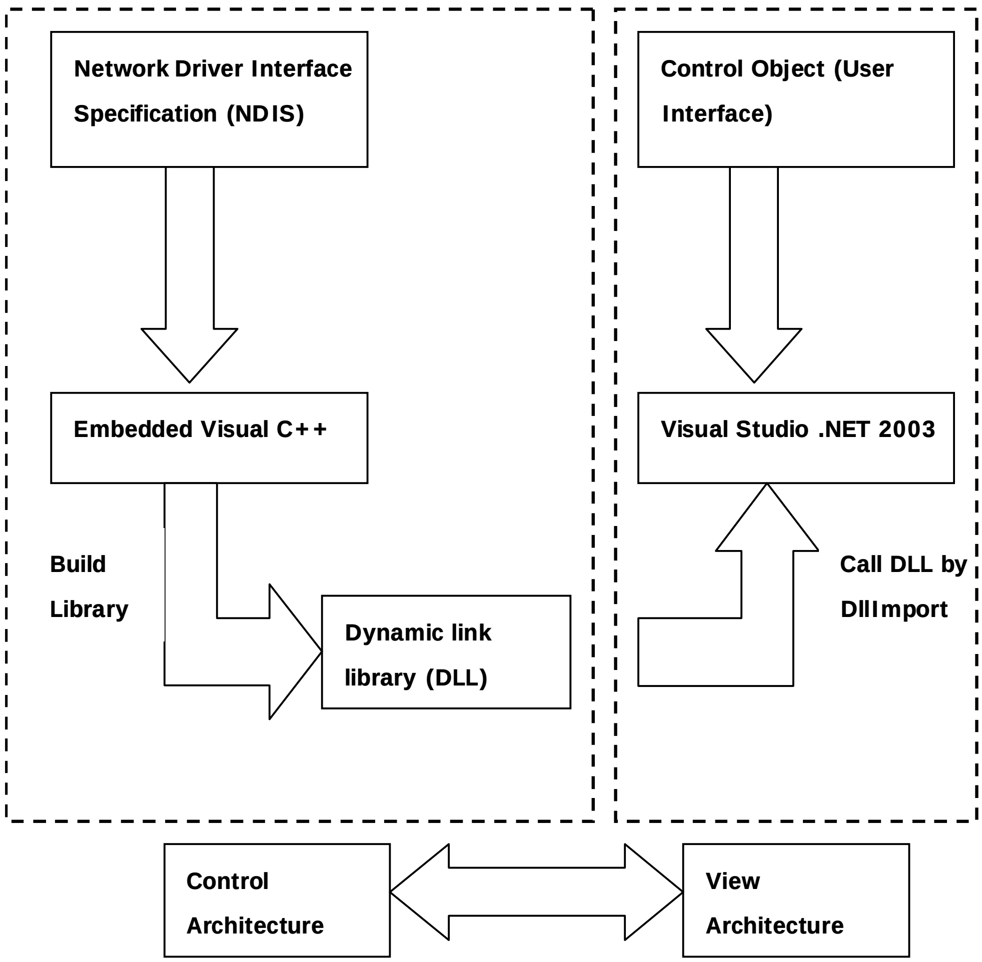


Figure 6.5

The Software Architecture is divided into two main parts: Control Architecture and View Architecture.

Control Architecture (CA) controls the program flow in the WSS. Its main function is to get the Wi-Fi information for the program. CA is made by tools Embedded Visual C++. Embedded Visual C++ supports lots of libraries for programming Pocket PC. It is a desirable tool for CA. The CA gets the Wi-Fi information though Network Driver Interface Specification (NDIS) Application Programming Interface (API). NDIS is explained in the following section.

View Architecture (VA) is used to display the result for WSS. Since the Visual Studio .NET 2003 supports lots of user interface control, it is desirable for building a fancy interface for WSS. Many fancy user interface controls are supported, such as Data Grid, Label, Tree View and Scroll Bar.

The main bridge for VA and CA is the Dynamic link library (DLL). The CA information is passed by the DLL to VA such that it can display in the VA interface. The CA project is build by Dynamic link library (DLL) project in Embedded Visual C++. The main DLL role is to act as a bridge for CA and VA to communicate. This is an important part of Software Architecture.

6.5.2 Network Driver Interface Specification (NDIS)

The Network Driver Interface Specification (NDIS) is the interface by one or more network adapter drivers communicate with network adapters, protocol drivers, miniport drivers, and operating system. NDIS is abstracted all the lower level network adapter development for developer.

NDIS provides a pair of abstraction layers that connect network adapter drivers to an overlying protocol stack. Transmission Control Protocol/Internet Protocol (TCP/IP) and Infrared Data Association (IrDA) are the example for overlying protocol.

NDIS is supported in the Windows CE. NDIS provided a abstracted programming interface to write an target-specific programming/driver for the network adapter. All Kernel-mode functions required for development can be exported by NDIS Dynamic Link Library (Ndis.dll). NDIS also store the state information for the network adapter. That means all the information related to network adapter can be access by the user program through NDIS.

NDIS support these following functions:

a. A network adapter driver provided an abstracted layer for the upper-layer (Application Layer) to sending the data into the network.

b. A network adapter driver provided an abstracted layer for the upper-layer (Application Layer) to receive the data from the network.

c. It provides an interface for upper-layer (Application Layer) to configure a network adapter or a network adapter driver

d. It provides an interface for upper-layer (Application Layer) to queries a network adapter driver for specific configuration data from an underlying network adapter or from the network adapter driver

The following diagram shows the overall NDIS architecture for Windows based Architecture:

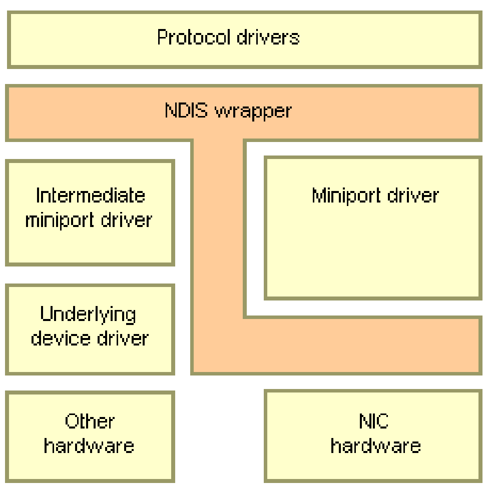


Figure 6.6

This diagram shows the NDIS acts as the bridge for the Protocol and the network adapter (Lower Layer). NDIS acts as midpoint role for the application to communicate with network driver. NDIS is interface for the application to get the lower layer information in the network adapter.

6.5.3 Feature of Wi-Fi Signal Scanner

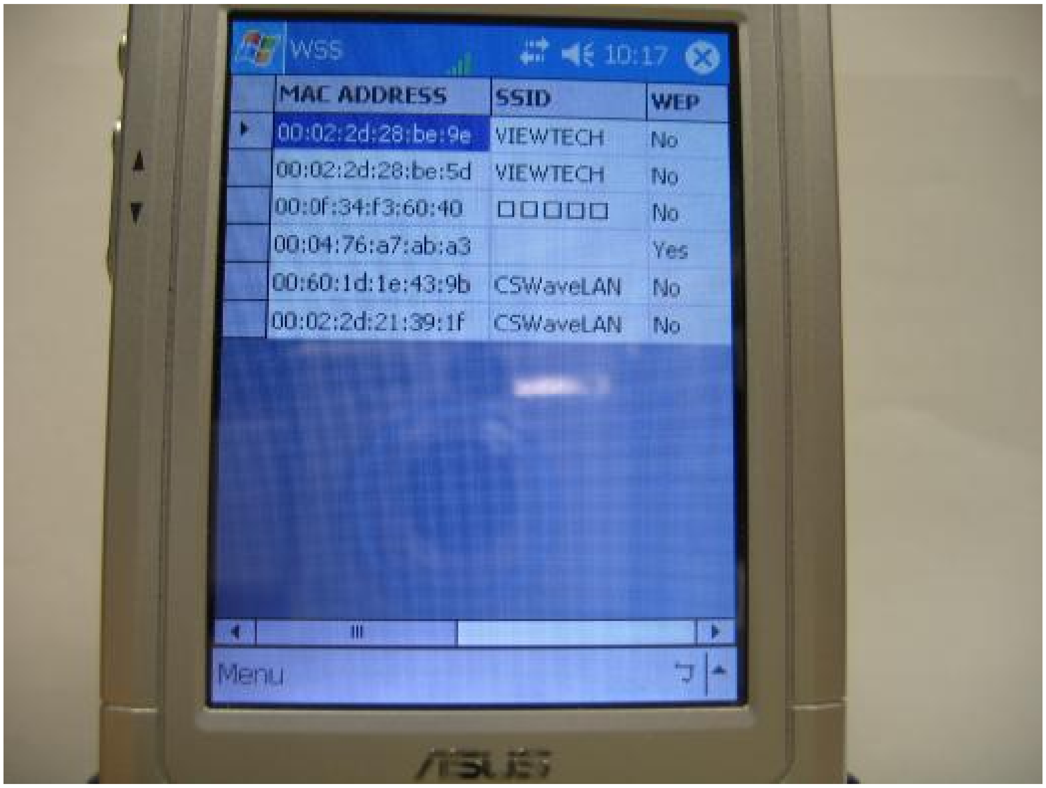


Figure 6.7

The above is the View of the WSS. They are made by the fancy user interface by Visual Studio .NET 2003. They are mainly made by Data Grid, Menu Item, etc. WSS is a highly user friendly software. The developer can learn the software in a short period of time.

There is much information shown in the WSS:

MAC Address:

It is the MAC Address of Access Point.

SSID:

It is the SSID of Access Point. Most of SSID is broadcast by the Access Point. SSID is not secured in WLAN.

WEP:

It is shown whether the WEP is enabled in the Access Point.

RSSI:

The signal is received by the mobile device for that particular Access Point

Num:

The number of times that signal received by the mobile device

Total:
The total number of signal strength received by mobile device.
Mean:
The mean of signal strength is received by the mobile device for that particular Access Point. The mean would used in localization algorithm.

There are several functions in the WSS:

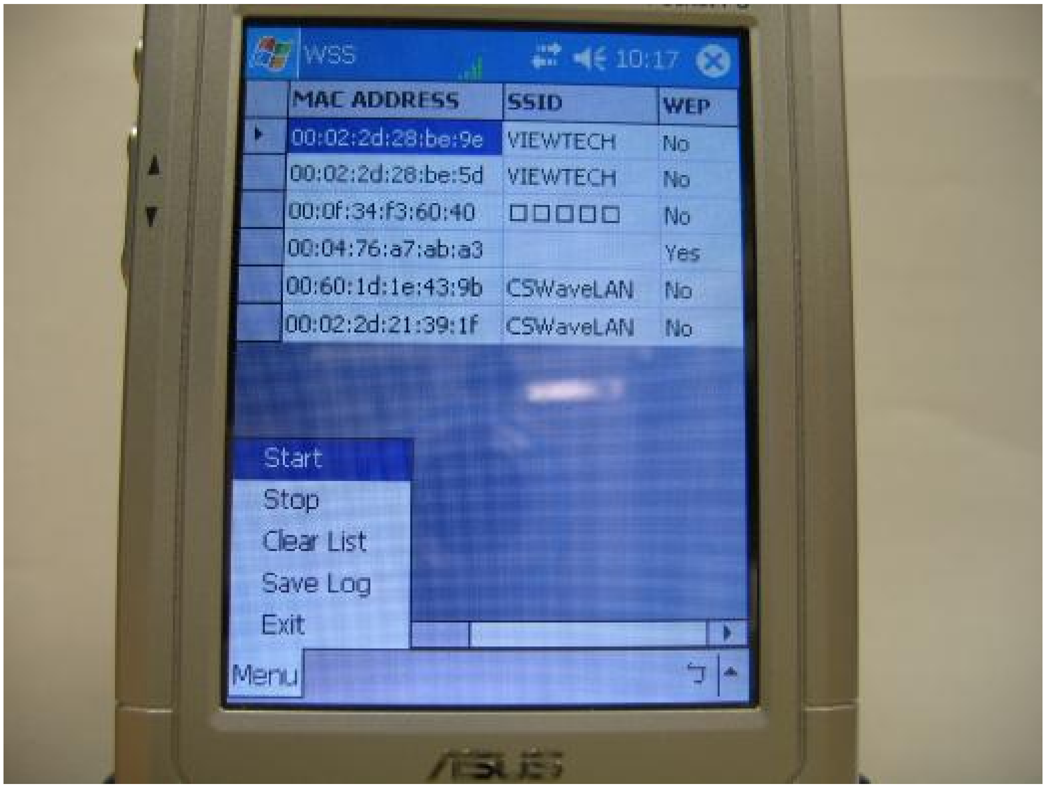
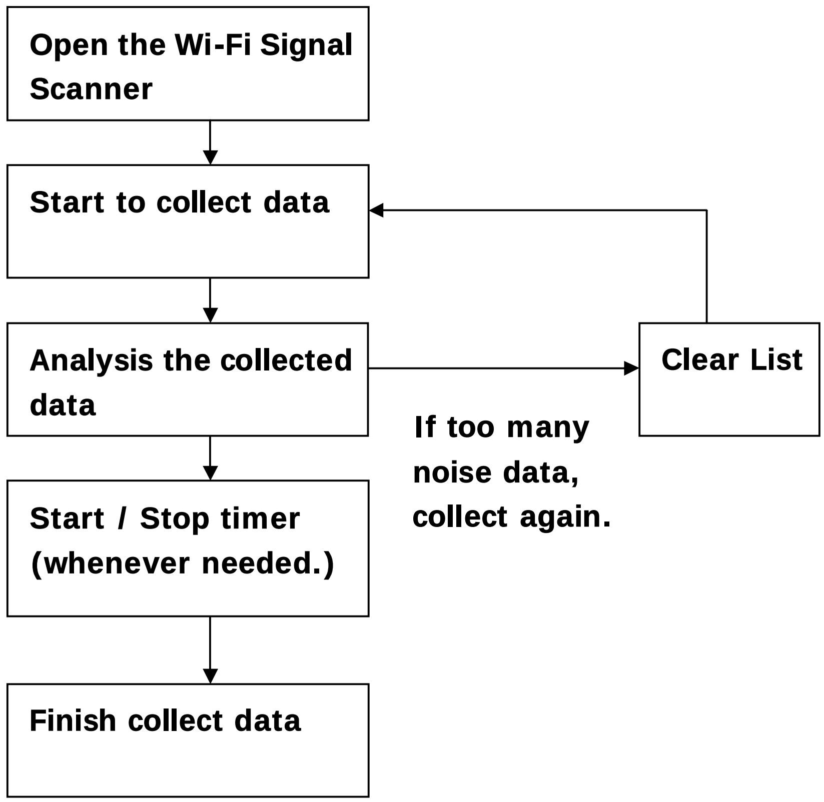


Figure 6.7

Start:
It is to start the timer of WSS. The time is enabled initially.
Stop:
It is to stop the timer of WSS.
Clear List:
It is to clear the current in the Data Grid.
Save Log:
It is to store the Wi-Fi information in the Data Grid. The Log file is stored the same path as the WSS program.
Exit:
It is to exit the WSS.

6.5.4 Collecting Data Procedure using Wi-Fi Signal Scanner



The above flow diagram shows the common procedures of collecting the source data in one target position. They are divided into five main states. Two main states are important in collecting data. They include "Analysis the collected data" and "Start / Stop timer".

Analysis the collected data:
In collecting the data in target place, some weak signals are received by device. They mainly come from other building, broken-down Access Point, etc. If there are too many weak signals in the collected data, it should be re-collected. The details would be explained in the following section.

Start / Stop timer:
WSS provides the flexibility for the developer to analysis data by stopping timer. It is convenient for them to analyze data in the middle of collecting process. If mean signal of target position tends to be static in short period of time, the developer should stop the timer and collect data in other target position.

6.5.5 Analysis the Data in Wi-Fi Signal Scanner

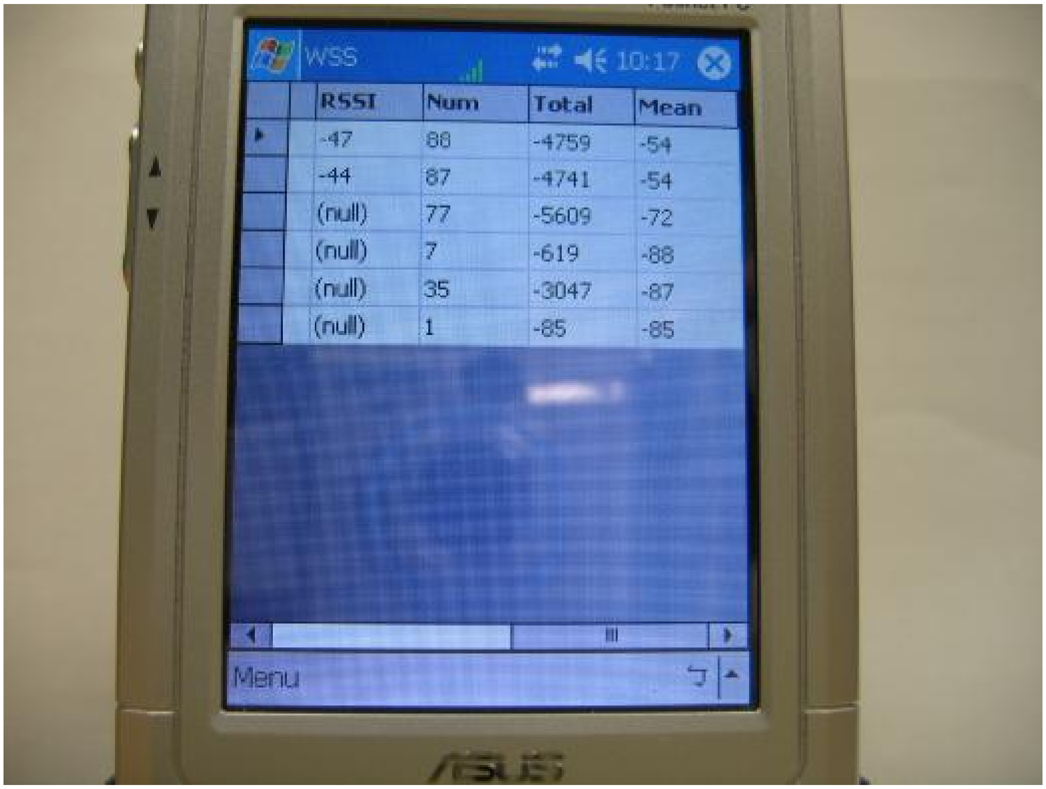


Figure 6.8

The above is the data got by the WSS at one target position. There is different meaning in each type of data. There are explanations in the following:

RSSI:

The signal is received by the mobile device. If RSSI shows a “null”, it means that no signal is received by the mobile device. Normally, RSSI shows the negative number. The value is the signal strength of that Access Point received by the device. The larger value (smaller in magnitude) means that signal strength is stronger for that Access Point. The value of RSSI is in terms of “dBm”.

Num:

It shows the number of signal received by device. If value of “Num” is too small, it means the connection between Access Point and device is very weak. We may ignore this data in localization algorithm. The data is labeled as “noise data”. This data is useless in locate the position. The data need to be filtered out in the WDP.

Mean:
This is an important data in the WLS. It is used in the localization algorithm. This mean data normally tends to be static after one to two minutes. If the mean data fluctuates for a long time, it means that signal received by device is not stable. This data should also be ignored because it would affect the accuracy of WLS. In localization algorithm, the stability of mean signal is very important. It should be very careful to analyze the data in WSS.

Total:
It is the same meaning as the "Mean". It can be used to analysis the quality of signal strength with num. If num is very small and total is large, it means the quality is very poor. If num is very small and total is small, it means the quality is very good. The Total normally is used to calculate the "Mean".

To conclude, it is important for us to analyze source data. It would affect the accuracy of localization algorithm. It should be careful when collecting the data in target position. There are different kinds of meaning for each field in the WSS. This would help the developer to filter out the noise data in the WDP.

6.6 Wi-Fi Data Processor (WDP)

WDP is a tool to process the source data that are collected by WSS. After analyzing the data, we need to filter out some noise data if needed. The parameters of target position which are shown in the WLD are also set in the WDP. The developer need to process all the data in WDP.

6.6.1 Overview of WDP

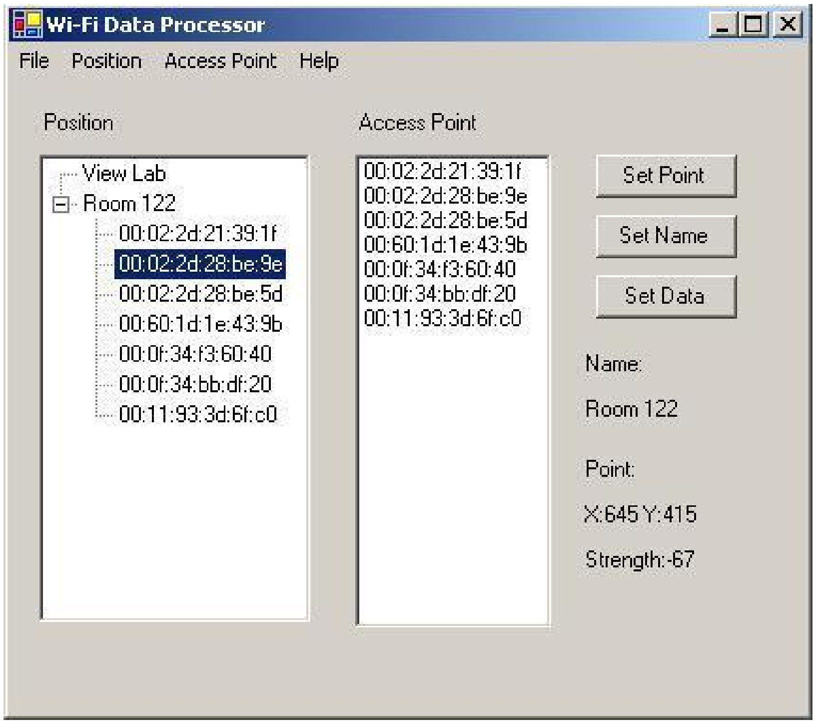


Figure 6.9

WDP is mainly divided into 3 main regions: Position, Access Point, Setting and Information.

Position:

It shows the number of target position added in WDP. The default name of newly added position is positionX (where X is any integer). After setting the data for that particular position, there would be a list of Access Point under that particular position.

Access Point

It shows the total list of Access Point being used in the WDP. The list is the MAC Address of Access Point. And it can be delete the items in list whenever is needed. The change of Access Point is also updated in the Position region.

Setting and Information:

The name, data and point of position are set in this region. All the parameters are set normally. Otherwise, they are set to default value initially. At the bottom of region, it shows the information of particular region. By clicking the list in the position region, the most updated information of that selected position will be displayed here.

6.6.2 Data Processing Procedure in WDP

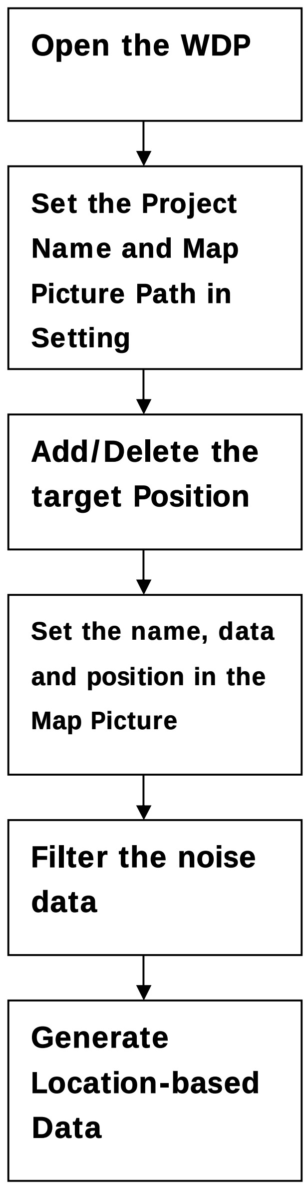


Figure 6.10

The Data Processing for source data contains six steps. It is explained in the followings.

Open the WDP (Step1)
First, it needs to open the WDP to perform processing source data.

Set the Project Name and Map Picture Path in Setting (Step 2)

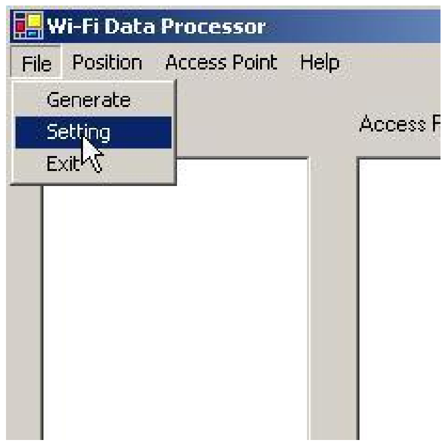


Figure 6.11

First, select the setting in the File Menu. And a setting dialog would be shown immediately.

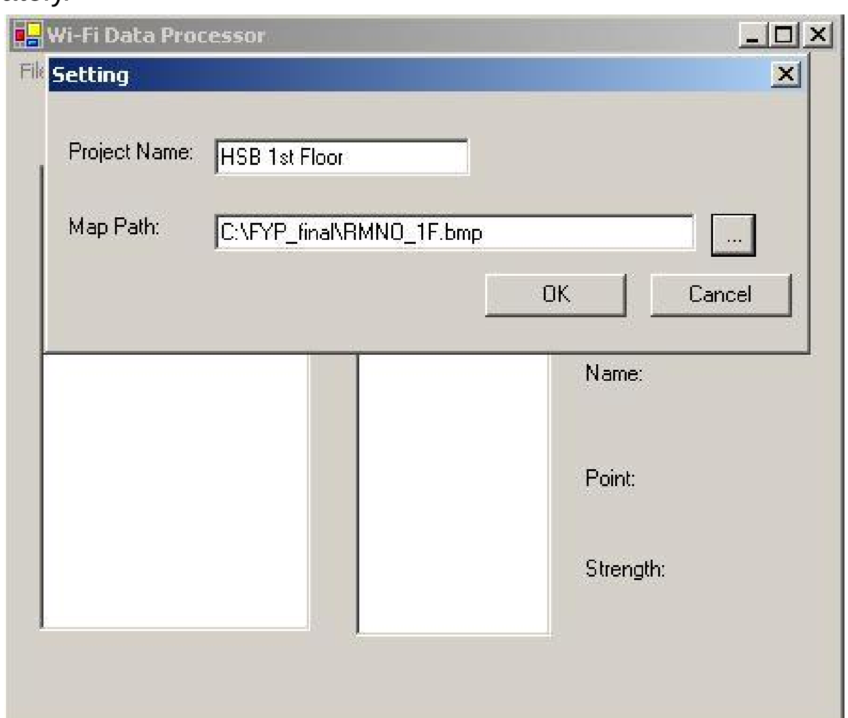


Figure 6.12

Then developer inputs the “Project Name” in the corresponding field and sets the Map Picture Path by browse the File Dialog. These two fields must be set. If not, the data processing procedure cannot go on.

Add/Delete the target Position (Step 3)

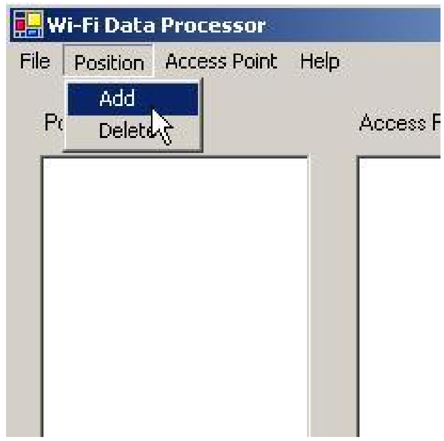


Figure 6.13

When adding or deleting the position, it should click the Position Menu in the Menu Bar. After adding the position, a new position is added in the list. The default name is “PositionX”. (where X is an integer)

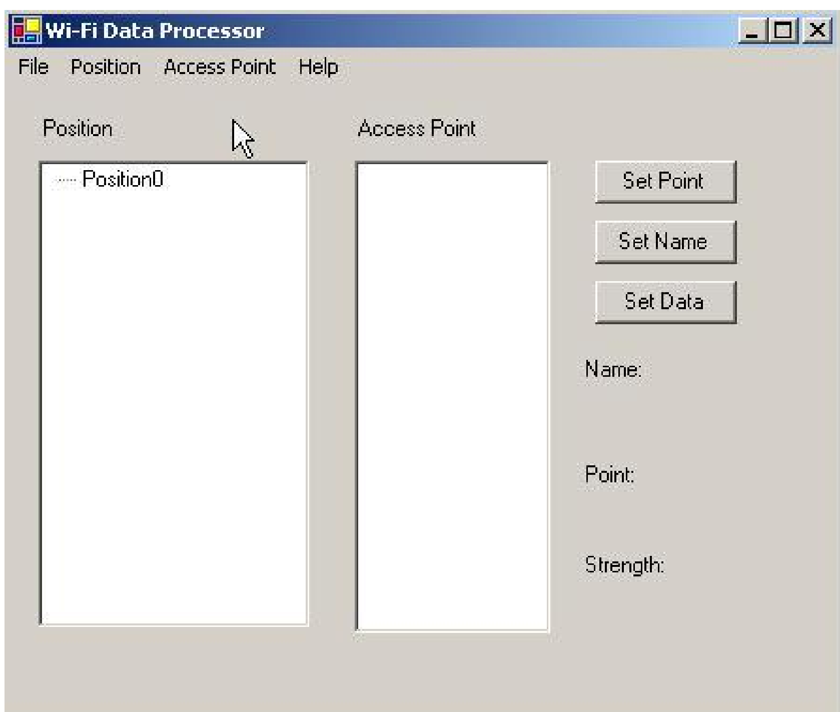


Figure 6.14

Set the name, data and position in the Map Picture (Step 4)

We should set the name, data and position in the Map Picture for target .position.

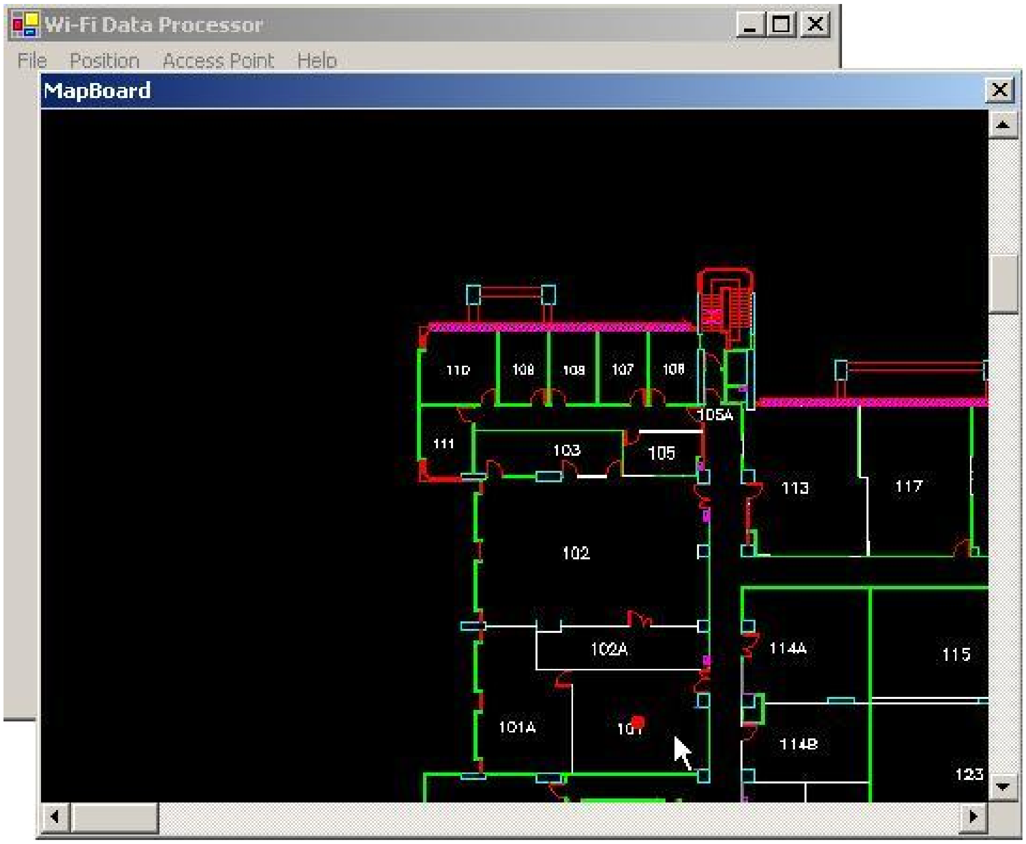


Figure 6.15

After clicking "Set Point", MapBoard is shown immediately. We can set the point on the Map Picture by clicking on the Map.

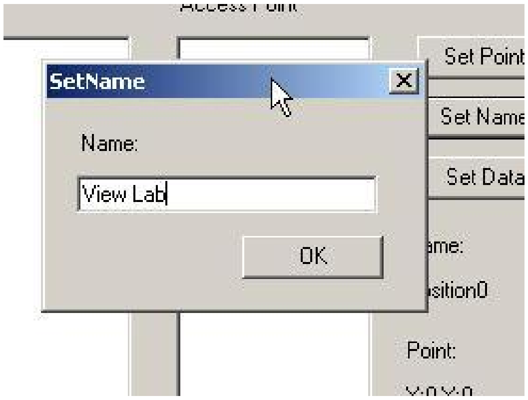


Figure 6.16

We set the name of a position by click the "Set Name" button. We should input the name in the Set Name dialog.

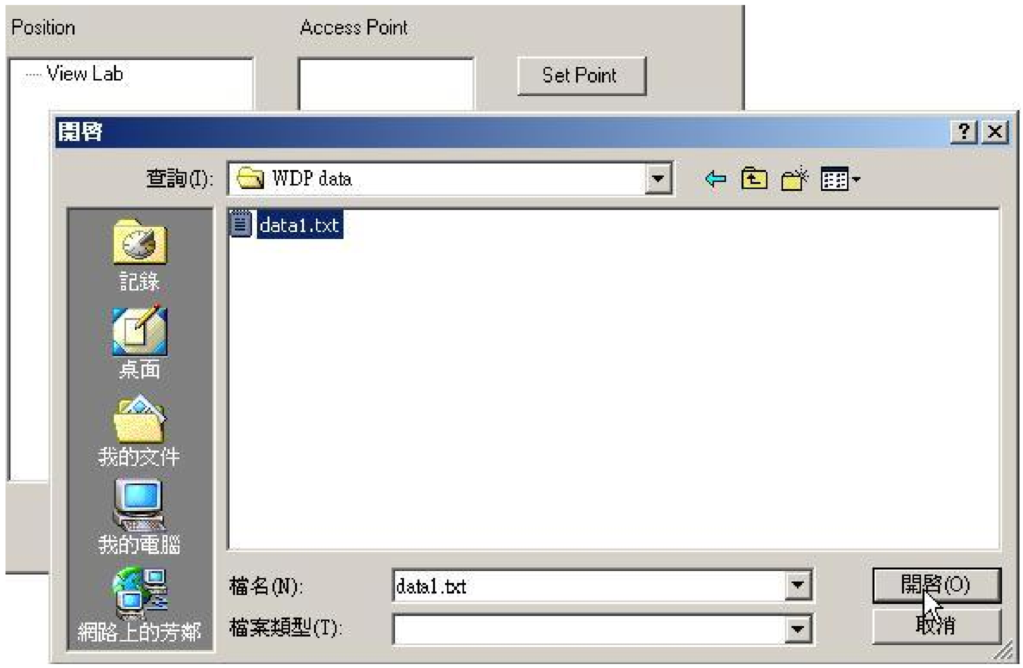


Figure 6.17

For each position, it should set the mean strength signal data to it. We should import the “saved log” of WSS to WDP. After setting the Data, there would be a list of Access Point under the name of position in the “position region”.

Filter the noise data (Step 5)

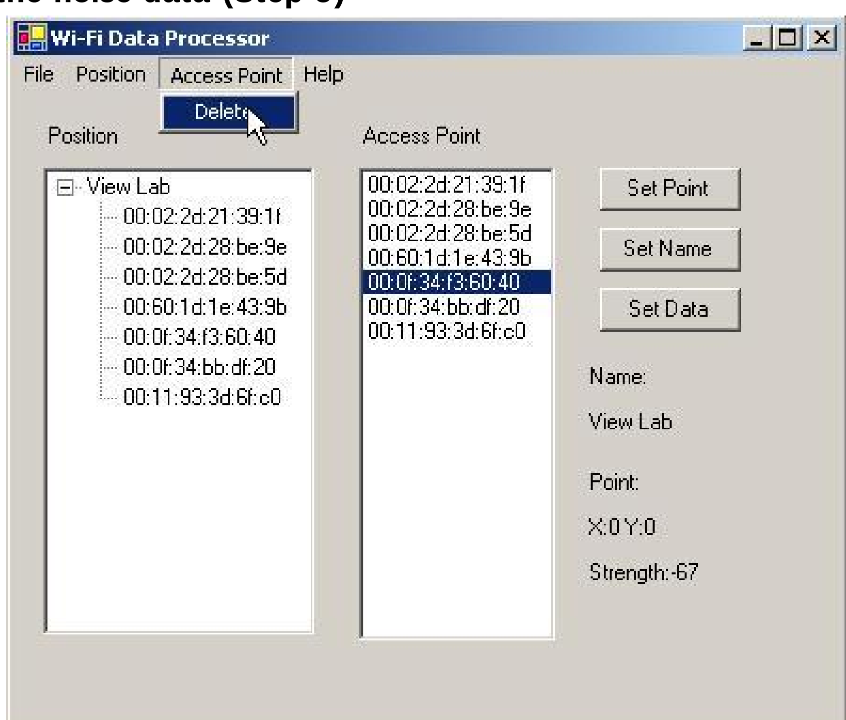


Figure 6.18

After analyzing the source data, we may need to filter out some noise data. It is done by deleting item in the Access Point List. The access point listed in the position would also be updated.

Generate Location-based Data (Step 6)

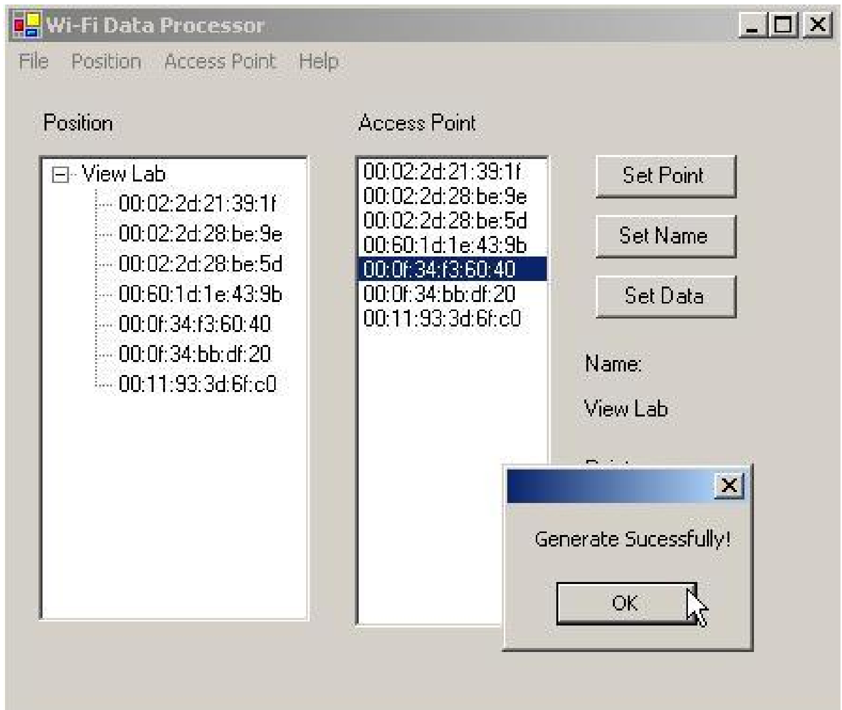


Figure 6.19

After finishing processing the source data, we can generate the Location-Based Data (LBSData). This is done by click “Generate” in File Menu. To generate successfully, there should be at least one position and one Access Point. And also the “Project Name” should be set. The LBSData and the Map Picture would be the data source for Wi-Fi Location Detector (WLD).

6.6.3 Feature of Wi-Fi Data Processor

Fancy User Interface:

The WDP is made Visual Studio .NET 2003. Graphics User Interface (GUI) simplifies the complexity of using WDP. The user can learn WDP in a short period of time. The user can set the Position parameters by just click the buttons or Map Picture in the interface. It is convenient for user to process the data in WDP.

Abstraction:

The WDP provides abstraction for the users to hidden the complex format of LBSData. They only to set the Data by GUI interface and WDP then generate LBSData for user. Whenever the format of LBSData changed, the WDP don't change except the Generating part. WDP abstracts all the core part in WLS.

Flexible

WDP is a flexible system. The user can set the position parameter according to the user need. There is no restriction on setting the parameters. The user can set the parameter freely. The "Add/Delete Position" Option can let the user to add or delete position in the WDP. This option provides another flexible for user.

Robustness:

There is an error handling in WDP. Whenever error occurs, error message would be shown on the screen immediately. This is to prevent the System crash in WDP when generating LBSData. It is an important feature in WDP. This is to ensure the LBSData generating successfully.

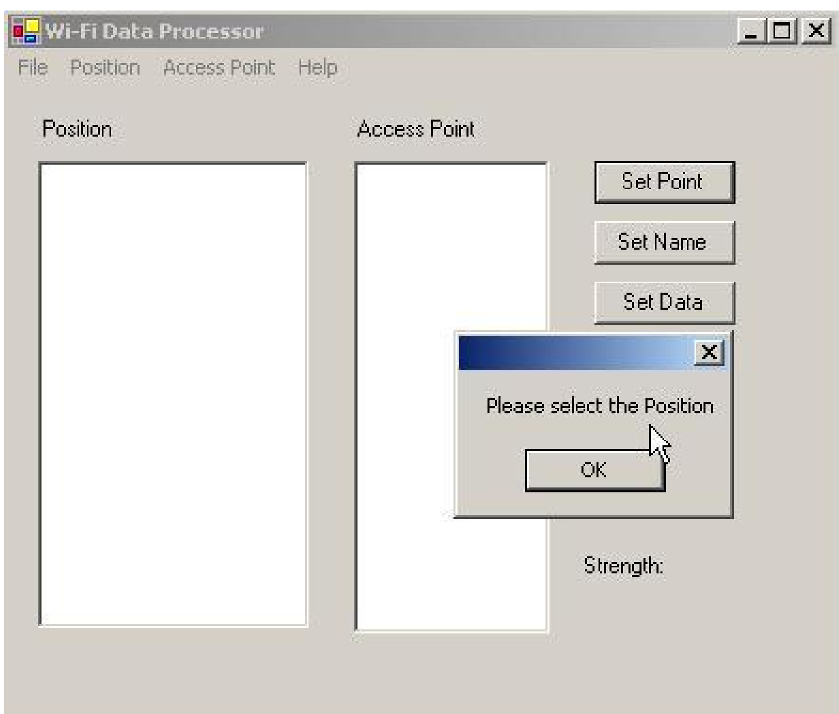


Figure 6.20

6.7 Wi-Fi Location Detector (WLD)

Wi-Fi Detector is a tool to detect the location in the target place. WLD would show the target position name and the position in the Map Picture. WLD estimates the position based on localization algorithm. WLD is a two dimensions detector which it can detect the position in two dimensions space.

6.7.1 System Architecture

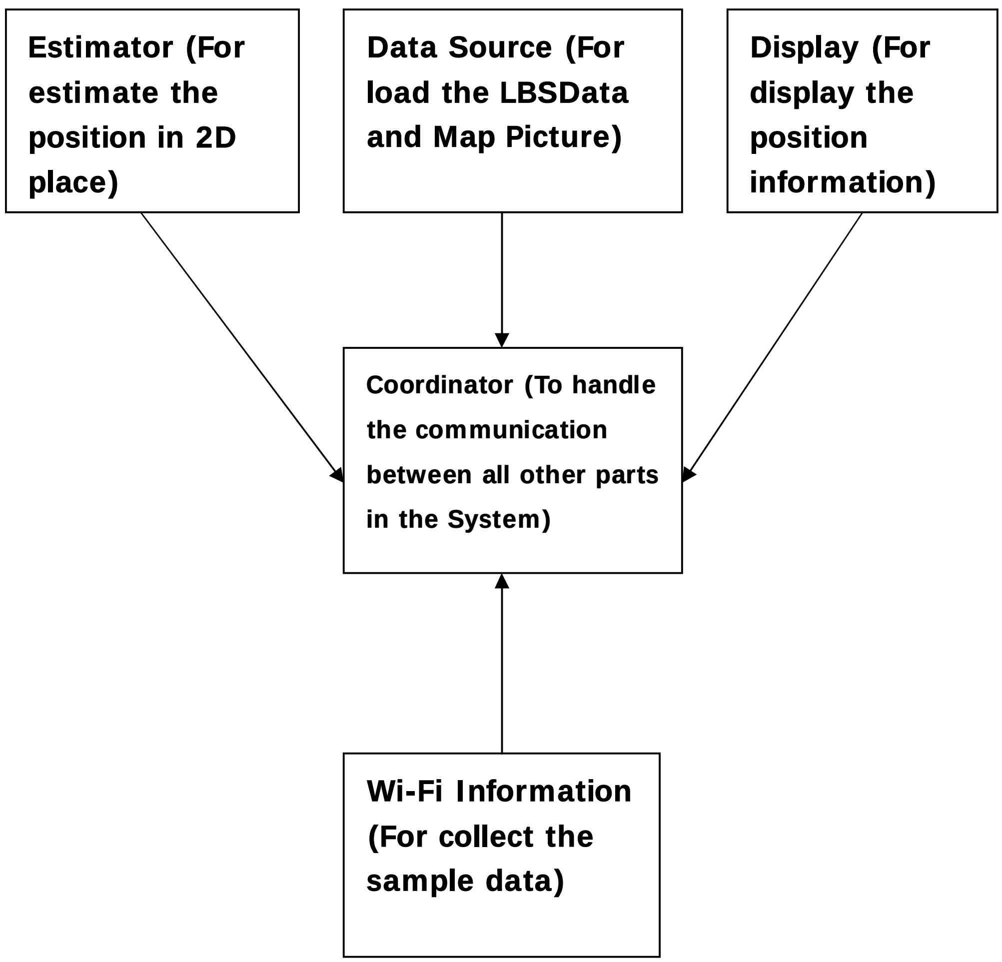


Figure 6.21

The system is divided into 5 parts. They are Estimator, Data source, Display, Wi-Fi Information and Coordinator. Each of them has different kinds of function. The detail would be briefly explained in the followings.

Data Source:
It is to control the data source for WLD. It is to read the data form the LBSData and assign it to the Object in the WLD. It is bridge for WLD and LBSData File to communicate. The Map Picture is also read by Data Source. Data Source is the interface for WLD to read data into the System.

Estimator:
It is core part of localization algorithm. The Estimator detects the position is based on the algorithm and sample data collected by Wi-Fi Information. After the calculation, it returns the result to the Core Part for display. This is main function of the Estimator.

Wi-Fi Information:
It is to collect sample data for the Access Point. The sample data is used in the Estimator for calculation. The quality of data by Wi-Fi Information is very important. It controls the accuracy of WLD in estimating position.

Display:
It controls the output of WLD. Display mainly shows results of Wi-Fi Information and Estimator. It processes this result and show in a fancy user interface.

Coordinator:
It is the Core Part of WLD. It controls the all parts in the System. It coordinates all the sub-parts in WLD. It acts the communication bridge for sub-parts to interact. It is the heart of the WLD.

With this architecture, WLD becomes a highly abstracted system. The work of WLD is divided into several parts. Each of them hides all detail parts. Other part only interact it with its interface. Whenever there are changes in one sub-part, it does not affect the system. The system can work normally. It is convenient for developers to modify the system with little effort. They can only update the parts that they want. In other words, it is a highly maintainable system.

6.7.2 Usage of Wi-Fi Location Detector (WLD)

There are three modes in WLD. They are Data Mode, Map Mode and Probability Mode. In each Mode, WLD shows different kinds information. It is convenient for the developer to testing the WLD.

Data Mode:

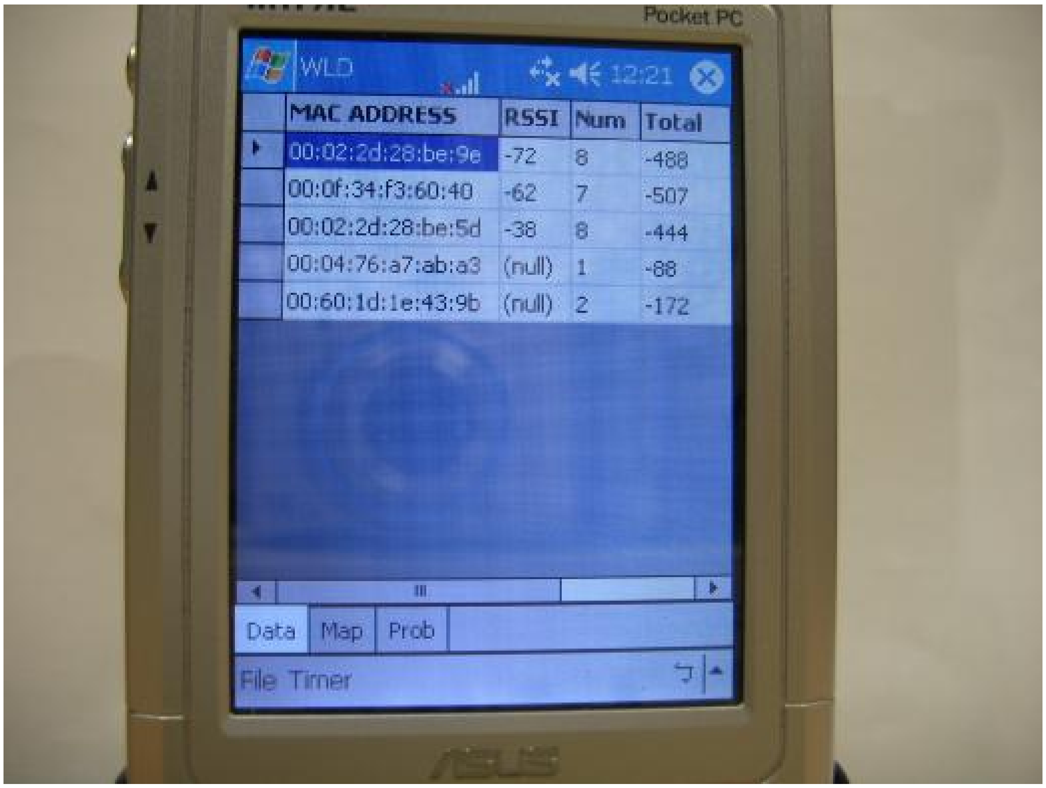


Figure 6.22

In the Data Mode, it shows the sample data collected by mobile device. The developer can analyze the quality through different field of data. We take the RSSI and Mean as an example. If there is too much null RSSI value in sample data, it means no signal or weak signal received by device. The WLD does not have enough data to estimate position. In the “Mean” field, if it is fluctuated in a large range, the data received by device is not stable. The stability of data affects the accuracy of localization algorithm.

Map Mode

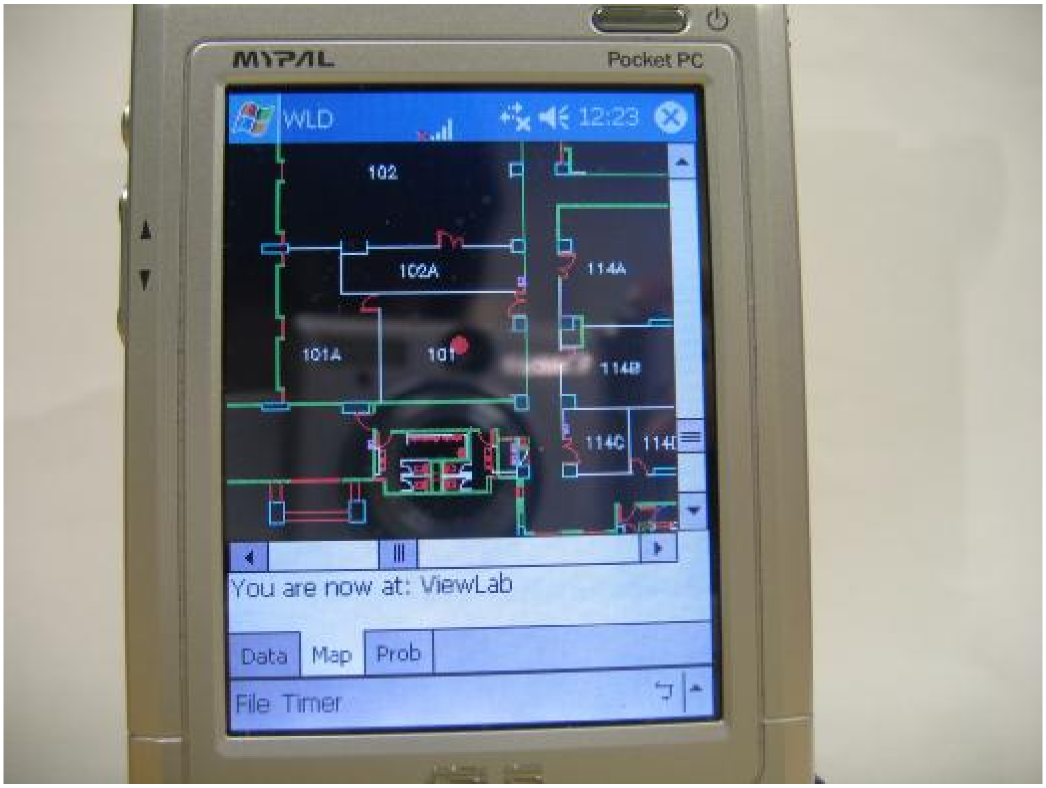


Figure 6.23

The Map Mode is to show the position in the Map Picture. It shows the position by draw a dot on the position. There is a description about the position at the bottom part.

The Map Picture can be move by several modes:
Scroll Bar:
It can be move by directly click the scroll bar on the Map Mode. This is tradition method of moving the picture.

Key Pad:
It can move by press the Key Pad on the mobile device. This is most convenient way to move the Picture.

Drag:
It can move picture by directly drag on the screen. It can drag any position that the user wants. They can move picture freely

Probability Mode

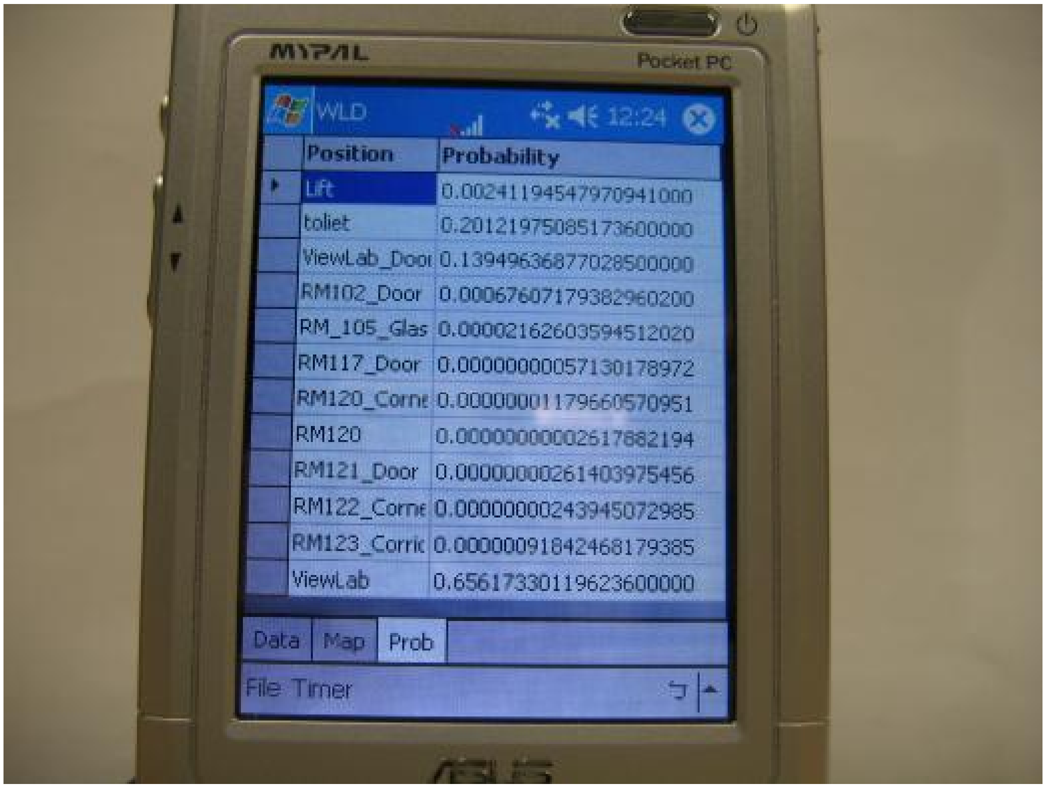


Figure 6.24

The Probability Mode shows the occurring probability in each position. The occurring probability is calculated by the localization algorithm. The highest probability means that you are now at that position. The developer can use it to test any error in the localization algorithm.

There several functions in the Menu Bar.

Clear List: To clear the sample data collected by device.
Start: To start the timer
Stop: To stop the timer
Exit: To exit the program

6.8 Summary of Wi-Fi Location System (WLS)

The advantages of WLS are in the followings.

User-friendly System
WLS is made by fancy user interface. User can learn WLS within short period of time. They can control the WLS by clicking the buttons or control in the system. WLS is a user-friendly System.

Hidden all the low-layer technique/technology
WLS have done all low-layer architecture for user. They can only to know how to process the data from the low-layer architecture. The user does not need to know the low-layer architecture before processing the data. The system hides all low-layer things for the user and done it for users. The user can focus on the work rather than the technology using in the task.

Reusability
WLS is a highly reusable System. It can be used to implement any Location-Based System (LBS). The user only needs to follow the procedures in implementing the System. WLS can generate any LBS in any area. WLS can be reused in deploying LBS in any area that user wants.

Efficiency System
WLS saves lots of time in developing LBS since all the low-layer things are done by WLS. The users only need to collect the data at the target place. And then they process the data and deploying the system to the client. Using WLS, it is much faster than tradition methods to develop LBS. The following tables show period of time to implement the system.

Tradition Method

Task	Time
Studying the technology	1-2 week
Software Design	2-3 week
Algorithm design	1-2 week

WLS Method

Task	Time
Wi-Fi Signal Scanner	3-4 days
Wi-Fi Data Processor	1 week
Wi-Fi Location Detector	1 week

The disadvantages of WLS is in the followings.

Fixed Technology

The WLS is only to support Wi-Fi (Wireless LAN) network. It cannot be using the GSM network or Bluetooth Network to deploy LBS. The technology is limited in WLS.

Fixed User Interface

Since all the user interface is done by WLS, the developers cannot change the interface that they want. The Interface is also fixed in WLS.

In conclusion, WLS is a desirable tool for developing LBS. With WLS, it develops LBS in a short time. It saves much effort on studying technology. It increases the efficiency of implementing system. However, WLS is only fixed on deploying Wi-Fi Network. And also the client interface cannot be changed by developer. However, WLS is done the fancy user interface. It can fulfill the common requirement on most LBS. WLS is focus on 2D detection environment. The most suitable technology for 2D detection environment is Wireless LAN network. At this moment, it has little chance to develop 2D detection environment with other technology. WLS is a suitable tool for developing the LBS.

7. Summary

Our final year project is entitled Location-Based Multimedia Mobile Service. In the first semester, we mainly focus on the problem of localization. Therefore, base on our knowledge and developed tools in localization, we are able to further develop a location-based service.

In this semester, we have chosen the 1st floor of the Ho Sin-Hang Engineering Building in the Chinese University of Hong Kong to study the problem of localization. We succeed in achieving our goal which is to locate a person when he/she is walking around on the floor.

There are a lot of problems in achieving the goal, for instance, how to collect the signals from the Wireless LAN access points, how to the collected signals, how to choose and apply algorithms to achieve localization and so on. And we have solved every problem mentioned and achieve the localization purpose.

In the first section, we have given an introduction about the current issue of localization in recent years so you realize that localization is a heat topic recently. Then in Section 2, it talks about Wireless LAN fundamentals including Wireless LAN access points, Wireless LAN terminology, and Wireless LAN standard which help you to get basic idea on Wireless LAN and our project.

In Section 3, we have talked about the point-based approach and the area-based approach, so you know the advantages and disadvantages of different approach for localization. We have also introduced three localization algorithms, namely, the Distance Mapping algorithm, the Simple Point Matching algorithm and the Area-Based Probability, so you should have a brief idea on how to achieve localization.

In Section 4, we have discussed the advantages of using the Area-Based Probability algorithm for localization over other algorithms. Then, we have given a detail description on how we apply the Area-Based Probability algorithm and solve the problems we face. It includes the way we generate the training set and the testing set. We have also described how we apply some Mathematics assumptions, distribution models and some theories so that we are able to find out the probability used in localization. At the end of the section, we have also listed out the result of the experiments. In Section 5, we have performed some studies on the accuracy of our system and we have also given some analysis of them.

We have implemented our first application Wi-Fi Location System (WLS). It is a develop tool for Location-based System. WLS consists of 3 components. They are Wi-Fi Signal Scanner (WSS), Wi-Fi Data Processor (WDP) and Wi-Fi Location Detector (WLD). WSS is used to collect signal strength data. The data is then processed by WDP. WLD estimates the position by output data of WDP. And we have discussed the features and functions of them one by one in Section 6.

These three components are closely related to each other. After the detection of access point signals and retrieval of signal strengths in WSS, the data will then pass to WDP for data processing. After that, the output data of WDP will become the training set for WLD, so that it can be able to carry out the location detection.

In the coming semester, we will complete the second half of the project. We will improve the accuracy, efficiency, scale and features of our Wi-Fi Location System. Then base on the system, we will be able to implement the tour guide service or some other location-based applications in the future. The detail of our future work will be discussed in Section 11.

8. Problem and solution

Problem 1

We can find any API to achieve the MAC Address of Access Point. We can get MAC address of Access Point.

Solution

We find API (NDIS) in Microsoft MSDN homepage that indirectly support to achieve MAC address of Access Point. NDIS is originally support to write the network adapter driver.

Problem 2

We don not how to using the API in NDIS since there is no sample code in MSDN. In the Internet, many people ask this question but no one knows how to solve it.

Solution

We ask this question in the MSDN homepage. However, the problem still does not be solved. We send the email to the Microsoft Researcher about the NDIS. The Researcher gives the sample code about the NDIS. We follow the code and carry on out project

Problem 3

The sample code seem do not work. We can only obtain one MAC address in Access Point List. We cannot get others MAC Address

Solution

There are is a bug in our program. We find the answer in the Microsoft MSDN homepage. We need follow structure defined in NDIS. We cannot move the pointer by the structure size. We should move the pointer by the length field in the structure.

Problem 4

We have differently in writing a GUI in Embedded Visual C++

Solution

Since there is lots if control don’t support in Embedded Visual C++, we move the Development Platform to Visual Studio .NET 2003. It supports more control and must easier in developing a beautiful GUI.

Problem 5

There is no NDIS support in .NET framework. We can get the MAC address in Visual Studio .NET 2003.

Solution

We make our Embedded Visual C++ to a DLL project. We import the DLL by Method “DllImport” in Visual Studio .NET 2003. We find this solution Microsoft MSDN homepage.

Problem 6

We have problem in write a Visual Studio .NET 2003 program. And we have not written any .NET program before.

Solution

We study the C# language on the NET. And we can find the samples codes in Microsoft MSDN homepage. Whenever we have problem, we solved by search in www.google.com.

Problem 7

We don’t know a good way to achieve the localization by using the signal strengths collected.

Solution

We read a lot of papers about the Localization Algorithm, so that we have learnt different approaches and different localization algorithms. We also find out the advantages and disadvantages of different algorithms, so that we are able to choose a suitable one to implement.

Problem 8

We do not know very clearly how to calculate the probability using the Gaussian distribution assumption in the Area-Based Probability algorithm.

Solution

We have found information on the Internet about the probability in mathematics. We have learnt the equation of the Gaussian distribution and

how to find the probability by integrating the Gaussian distribution function.
Problem 9
We do not know how to write a program to perform integration.

Solution
We have found information on the Internet about the error function in mathematics. The Gaussian distribution function can be rewritten in terms of the error function. Then we can estimate the error function by a series and write a function in a program to calculate the value of the series.

Problem 10
We do not know how many iterations we need to calculate the series in problem 9. We need to minimize the number of iterations to reduce the overhead in calculating the probability.

Solution
We test on the accuracy of the series by increasing the number of iterations. We choose the number of iterations which is large enough so that the probability calculated will not be negative due to estimation error.

Problem 10
We do not know whether the orientation of collecting signal will affect the accuracy of our system.

Solution
We have performed a study on the orientation. We collect the signals in different orientation and compare their accuracy.

Problem 11
We do not know how large we should choose in the sample size of a testing set.

Solution
We have performed the studies on how the sample size of the testing set affects the accuracy of our localization system. With the study result, we can find a suitable sample size.

Problem 12

We do have a floor plan of the first floor of the Ho Sin-Hang Engineering Building.

Solution

We have asked for the help from the administrators of the engineering building. We have signed a special request form as a formal procedure. Finally, we get the floor plan from them.

9. Contribution of work

9.1 Introduction

In this semester, we have implemented the Wi-Fi Location System (WLS) for the localization purpose. Hence, we would be able to design a location-based application based on the knowledge and technology we have discovered in WLS.

The WLS can be divided into three modules. The first module is the detection of WLAN access point signals and retrieval of signals strength. The second is the data processing of collected signals and the third module is the location detection. These three modules are included in the three programs in WLS. They are the Wi-Fi Signal Scanner (WSS), the Wi-Fi Data Processor (WDP) and the Wi-Fi Location Detector (WLD) respectively.

These three modules are closely related to each other. After the detection of WLAN access point signals and retrieval of signal strengths in the first module, the data will then pass to the second module for data processing. After that, the processed data will become the training set for the third module, so that it can be able to carry out the location detection.

Most of my work focuses on the location detection module. The detail will be described in the next sections.

9.2 Preparation Work

Starting from the last summer, we have been preparing for the final year project. Before the project, I had little chance to use the PDA. Therefore, in the summer, I have got more familiar with it. I learnt the operating system Window CE in the PDA, so that I am able to use the PDA for simple purpose such as running programs, file management and memory control. Then, I spent a couple of weeks in learning the embedded Visual C++ and the use of Pocket PC 2003 SDK, so that I am able to write some programs which can be run inside the PDA. Moreover, we study a final year project recently done for localization purpose to get a deeper understanding on localization. We also read a lot of materials on papers, books or Internet about wireless network so that we can know more about the terms, standards and the properties in Wireless LAN.

9.3 Detection of Access Point Signals and Retrieval of Signal Strengths

Although our PDA has the ability to scan for access points automatically or manually, it does not show the strength of the signal received. Also, this function is a build-in one in the PDA and there is no way to get a source code from it. Therefore, we have spent more than three weeks in order to find a way to detect and retrieve the signal strength. We searched the entire library and functions in Window CE and at the end we found a secret library which contains functions to scan the information from the access points. We understood the data structures, input and output parameters and the ways to use of these functions, however, we still spent one more week on extracting the information from the data structures because they are used in an unusually way.

When we are able to get the information from our program, Wi-Fi Signal Scanner (WSS), I had designed the ways to collect the signals so that the data can be later used by generating a training set for localization. Later, we have spend one more week in learning the programming by using the C# .Net, so that we can change WSS originally written in C++ to C#. It is because there are more libraries in C# to write a fancier user interface.

9.4 Data Processing of Collected Signals

The main goal of the module is to process the data collected, mainly the signal strengths, so that it can become the training set later used by the localization algorithm.

In order to generate the first training set for the algorithm, I need to use the Microsoft Office Excel to process thousands of data manually. I had to design how to handle some missing signals and set up some criteria in selecting the access points for localization in order to let the calculation in the algorithm more efficient.

Later, Wi-Fi Data Processor (WDP) is implemented mainly by my partner, and is also based on my experience and suggestions. It becomes more convenient in the data processing.

9.5 Location Detection

The goal of this module to detect the location of an object based on the signals it received in its current position. The signal data collected in this way is known as the testing set.

My main duty in this module is to design a suitable algorithm to use both the training set and testing set so as to achieve the goal of localization. I have spent several weeks in reading papers on the related topics. As a result, I have learnt the principles of different approaches and algorithms in localization. After my investigation on the advantages and disadvantages of different algorithms, I compare them and finally choose the Area-Based Probability algorithm.

I am also responsible for the implementation of the algorithm. I find many difficulties in applying the algorithm, such as understanding the mathematics equations, finding the integral of the Gaussian distribution function and increasing the accuracy of the algorithm. Therefore, I had chances in having a deeper knowledge on the algorithm, mathematics probability and applying numerical computation in computer. I had modified and improved the algorithm so that it is more accurate and specific.

In order to know the accuracy of our system, I have made several studies on the accuracy and factors affecting it. I have tested the localization of our system in different locations and constructed several sets of statistics. I have found that the accuracy of our system is ranged from 76% to 99%, mainly depends on the way we collect and generate a testing set.

9.6 Conclusion

The workload of the final year project is evenly distributed among two of us. Even though sometimes we are not doing on the same section, my partner Clarence and I always work together at the same time. We distribute the work among ourselves evenly based on our interest and strength. I really enjoy the partnership in doing this final year project.

In this project, I have gained much knowledge, for example, I have learnt two more programming languages, C++ and C#. I have also learnt more about Wireless LAN, embedded operation system, mathematics probability, numerical computation and data processing. All this knowledge and problem solving techniques is important for my future career. I am very glad that I have this change to conduct this project.

10. Schedule of Project

The following table is the Project schedule in this semester:

Date	Progress
June 2004	-Plan the project aim and start to design project -Study the wireless technology, such as Bluetooth, GSM network, Wireless LAN network
July 2004	-Study the OS platform that can implement project. E.g. Symbian OS, Windows CE. -Get familiar with Embedded Visual C++ and Symbian programming with Visual C++.
August 2004	-Confirm the project using the Windows CE platform with Wireless LAN network -Start to study the Application Program Interface (API) related to achieve MAC Address from Access Point
September 2004	-Implement the draft version program to achieve the MAC address with Embedded Visual C++ -Study localization algorithm
October 2004	-Modify the draft program to fully planned version(Wi-Fi signal Scanner) with Visual Studio .NET 2003 -Collect data using Wi-Fi signal Scanner -Plan to implement first application using Location-based Technology -Design appropriate algorithm in first application

November 2004	-Test our first application(Wi-Fi Location Detector) -Study the accuracy of Area-Based probability algorithm -Plan and Develop the Wi-Fi Location System(WLS) -Finish Develop the Wi-Fi Data Processor

11. Future Work

1. Ho Sin-Hang Engineering Building Tour Guide Service
We will introduce this location-based application based on our localization system. This service allows a user to locate its position in the Ho Sin-Hang Engineering Building. Besides, the service will give some introduction about the rooms and laboratories when the user arrives at these locations. This service can be applied during the open-day when the visitors visit the engineering building.

2. Location Monitoring System
This is the second location-based application we will introduce based on our localization system. We will build a centralized server which can monitor the locations of different users. Each user will have a mobile device with our system which can locate him/her and also indicate the locations of the other users in the system.

3. Multimedia Application with video streaming
We will study on the field of multimedia and then apply video streaming on the tour guide service. The video can be some descriptions about the Ho Sin-Hang Engineering Building and the facilities inside.

4. Improvement in Localization Algorithm
We will study on localization algorithm more deeply, so that we can modified it to meet the requirement of the location-based applications. We will research on a faster and more accurate algorithm.

5 Increase the Accuracy in Localization
We will study on the more factors that affect the accuracy in localization, for example, the way to collect training set and the way to divide the floor into different areas. With this prior knowledge, we are able to increase the accuracy of our system.

6 Increase the Scale of the System
We will increase the scale of the localization system, so that more rooms or floors can be located in our system.

7 Improvement of Wi-Fi Location System

We would like to add more features in Wi-Fi Data Processor, such as saving the work in WLD. We will also improve the interface of Wi-Fi Location System so that it is fancier. And we will improve the generality of the system, so that it can be applied not on in the engineer building, but also the other places.

8 Research on 3D localization algorithm in an building

We would like to make a study on 3D localization algorithm. We would make some experiment on 3D algorithm whether it is workable and implement it in an application.

12. Acknowledgement

We would like to thank Prof. Michael Lyu, our project supervisor, who has given us many useful suggestions and directions throughout this project.

We would also like to thank Edward Yau, Technical Manager of VIEW project. He has given us lots of background knowledge about Wireless Technology. He has also given us various innovative ideas on our project.

We would like to special thank John Krumm, Microsoft Researcher, who has given a technique support on using NDIS API.

13. References

1. Moustafa Youssef, and Ashok K. Agrawala, "Continuous Space Estimation for WLAN Location Determination Systems," IEEE Thirteenth International Conference on Computer Communications and Networks, Chicago, IL USA, October 11-13, 2004.
2. Moustafa Youssef, and Ashok K. Agrawala, "Handling Samples Correlation in the Horus System," IEEE Infocom, Hong Kong, March 2004.
3. P. Krishnan, A.S. Krishnakumar, Wen-Hua Ju, Colin Mallows, and Sachin Ganu, "A System for LEASE: Location Estimation Assisted by Stationary Emitters for Indoor RF Wireless Networks," IEEE Infocom, Hong Kong, March 2004.
4. Kamol Kaemarungsi, and Prashant Krishnamurthy, "Modeling of Indoor Positioning Systems Based on Location Fingerprinting ," IEEE Infocom, Hong Kong, March 2004.
5. Youngjune Gwon, Ravi Jain, and Toshiro Kawahara, "Robust Indoor Location Estimation of Stationary and Mobile Users," IEEE Infocom, Hong Kong, March 2004.
6. Moustafa Youssef and Ashok Agrawala, "On the Optimality of WLAN Location Determination Systems," Communication Networks and Distributed Systems Modeling and Simulation Conference, January 18-24 2004, San Diego, California.
7. S. Ganu, A.S.Krishnakumar, P.Krishnan, "Infrastructure-based Location Estimation in WLAN Networks," IEEE Wireless Communications and Networking Conference (WCNC 2004).
8. Moustafa Youssef, Ashok Agrawala, A. Udaya Shankar, "WLAN Location Determination via Clustering and Probability Distributions," IEEE International Conference on Pervasive Computing and Communications (PerCom) 2003, Fort Worth, Texas, March 23-26, 2003.
9. Moustafa Youssef, Ashok Agrawala, "Small-Scale Compensation for WLAN Location Determination Systems," IEEE Wireless Communications and Networking Conference (WCNC) 2003 New Orleans, Louisiana, March 16-20, 2003.
10. Moustafa Youssef, Ashok Agrawala, "On the Optimality of WLAN Location Determination Systems," CS-TR 4459, Department of Computer Science, University of Maryland, College Park, March 2003.

11. Siddhartha Saha, Kamalika Chaudhuri, Dheeraj Sanghi, Pravin Bhagwat, "Location Determination of a Mobile Device Using IEEE 802.11b Access Point Signals," IEEE Wireless Communications and Networking Conference (WCNC) 2003 New Orleans, Louisiana, March 16-20, 2003.
12. Saikat Ray, Rachanee Ungrangsi, Francesco de Pellegrini, Ari Trachtenberg and David Starobinski, "Robust Location Detection in Emergency Sensor Networks," IEEE Infocom 2003, San Francisco, pp. 1044-1053, March 2003.
13. Saikat Ray, David Starobinski, Ari Trachtenberg and Rachanee Ungrangsi, "Robust Location Detection with Sensor Networks," IEEE JSAC (Special Issue on Fundamental Performance Limits of Wireless Sensor Networks), in press.
14. Andrew Howard, Sajid Siddiqi, and Gaurav S Sukhatme, "An Experimental Study of Localization Using Wireless Ethernet", International Conference on Field and Service Robotics (FSR'03), Lake Yamanaka, Japan.
15. Sajid Siddiqi, Gaurav S. Sukhatme, and Andrew Howard, "Experiments in Monte-Carlo Localization using WiFi Signal Strength," International Conference on Advanced Robotics (ICAR), Coimbra, Portugal, 2003.
16. Ping Tao, Algis Rudys, Andrew M. Ladd, Dan S. Wallach, "Wireless LAN location-sensing for security applications", Wireless Security Workshop, San Diego, CA, 2003.
17. Ping Tao, Algis Rudys, Andrew M. Ladd, Dan S. Wallach, "Wireless LAN locationsensing for security applications", Wireless Security Workshop, San Diego, CA, 2003.
18. T. He, C. Huang, B. M. Blum, J. A. Stankovic, and T. Abdelzaher. “Range-free localization schemes for large scale sensor networks. In Proceedings of the Ninth Annual Internationa Conference on Mobile Computing and Networking (MOBICOM)”, San Diego, CA, Sept. 2003
19. Andrew M. Ladd, Kostas E. Bekris, Algis Rudys, Guillaume Marceau, Lydia E. Kavraki, and Dan S., "Robotics-Based Location Sensing using Wireless Ethernet," The Eighth ACM International Conference on Mobile Computing and Networking (MOBICOM) (Atlanta, GA), September 2002.
20. P. Prasithsangaree, P. Krishnamurthy, and P. K. Chrysanthis, "On Indoor Position Location With Wireless LANs ," The 13th IEEE

International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC 2002), Lisbon, Portugal, September 2002.

21. Arun Kishan, Michael Michael, Sahil Rihan, and Rahul Biswas, “Halibut: An Infrastructure for Wireless LAN Location-Based Services.”
22. Michael Wallbaum, "Wheremops: An Indoor Geolocation System", The 13th IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications, Pavilhão Atlântico, Lisbon, Portugal, September 15-18, 2002.
23. T. Roos, P.Myllymäki, and H.Tirri, "A Statistical Modeling Approach to Location Estimation," IEEE Transactions on Mobile Computing, Vol. 1, No. 1, January-March 2002, pp. 59-69.
24. K. Pahlavan, X. Li, and J. Makela, "Indoor Geolocation Science and Technology", IEEE Communications Magazine, Feb. 2002.
25. Mauro Brunato, Csaba Kiss Kalló, "Transparent Location Fingerprinting for Wireless Services," in Proceedings of Med-Hoc-Net 2002.
26. Roberto Battiti, Thang Lee Nhat, Alessandro Villani, " Location-aware computing: a neural network model for determining location in wireless LANs," Technical Report DIT-02-0083, 2002.
27. Roberto Battiti, Mauro Brunato, Alessandro Villani, " Statistical Learning Theory for Location Fingerprinting in Wireless LANs," Technical Report DIT-02-0086, 2002.
28. Paul Castro, Patrick Chiu, Ted Kremenek, and Richard Muntz, "A Probabilistic Location Service for Wireless Network Environments," In Proceedings of Ubicomp 2001, pp. 18--24. Springer Verlag, September 2001.
29. Asim Smailagic, Daniel P. Siewiorek, Joshua Anhalt, David Kogan, and Yang Wang, “Location Sensing and Privacy in a Context Aware Computing Environment,“ Pervasive Computing, 2001.
30. P. Myllymaki, T. Roos, H. Tirri, P. Misikangas, J. Sievanen, "A Probabilistic Approach to WLAN User Location Estimation," The Third IEEE Workshop on Wireless LANs, 2001.
31. G. Shreve, D. Kell, "A Precision Location Network Using Ultra Wideband WLAN Radios," The Third IEEE Workshop on Wireless LANs, 2001.
32. Personnel, T. Kochanski, "A Portable Wideband System for Monitoring and Locating Firefighters and Other Emergency," The Third IEEE Workshop on Wireless LANs, 2001.

33. J. Hightower and G. Borriello. “Location systems for ubiquitous computing.” IEEE Computer, 34(8):57–66, Aug. 2001
34. Paramvir Bahl and Venkata N. Padmanabhan, “RADAR: An In-Building RF-based User Location and Tracking System,” IEEE Infocom 2000, volume 2, pages 775-784, March 2000.
35. P. Bahl, V. N. Padmanabhan, and A. Balachandran, “Enhancements to the RADAR User Location and Tracking System,” Microsoft Research Technical Report: MSR-TR-00-12 (February 2000)
36. Asim Smailagic, Jason Small, and Daniel P. Siewiorek, “Determining User Location For Context Aware Computing Through the Use of a Wireless LAN infrastructure,” December 2000.
37. Paul Castro and Richard Muntz, "Managing Context for Smart Spaces," IEEE Personal Communications, October 2000.
38. Jay Werb, Colin Lanzl, "Designing a Positioning System for Finding People and Things Indoors," IEEE Spectrum, Vol. 35, No. 9., pp. 71-78, September 1998.
39. S. Thrun, W. Burgard and D. Fox, “A probabilistic approach to concurrent mapping and localization for mobile robots”. *Machine Learning*, 31(1-3):29–53, 1998.
40. Thomas Christ, Philip Godwin, "A Prison Guard Duress Alarm Location System," Proc. IEEE International Carnahan Conference on Security Technology, October 1993.

14. Appendix

14.1. Signal Strengths Collected by Wi-Fi Signal Scanner

AP MAC address	Position											
	1	2	3	4	5	6	7	8	9	10	11	12
00:02:2d:28:be:9e	-70	-62	-58	-67	-73	-78	-83	-86	-84	-81	-78	-55
00:02:2d:28:be:5d	-67	-59	-60	-71	-76	-79	-81	-86	-81	-83	-79	-52
00:60:1d:1e:43:9b	-79	-87	-85	-84	-89	-80	-76	-77	-66	-63	-77	-90
00:0f:34:f3:60:40	-63	-69	-65	-74	-76	-72	-77	-84	-76	-74	-66	-79
00:02:2d:21:39:1f			-82	-78	-82	-59	-78	-73	-83	-85	-82	
00:11:93:3d:6f:c0				-90	-85	-86	-89	-88				
00:11:20:93:65:c0					-89	-89						-90
00:0f:34:bb:df:20				-89	-90	-82	-88	-88				
00:0c:ce:21:1b:9d						-87						
00:0c:85:35:33:d2					-88			-88				
00:11:20:93:63:90						-89						-88
00:0c:85:35:33:d4									-87			
00:04:76:a7:ab:a3												-90

14.2 Probability Changes at Different Locations

14.2.1 Probability Changes at Area 1

Probability at Locations	Sample Number										
	1	5	9	13	17	21	25	29	33	37	41
Probability at A_1	0.386747	0.393857	0.238336	0.265182	0.272068	0.290128	0.308682	0.292967	0.307465	0.316625	0.329145
Probability at A_2	0.327626	0.343003	0.231836	0.237417	0.218078	0.214042	0.177562	0.168522	0.141765	0.130703	0.13968
Probability at A_3	0.104723	0.114281	0.230239	0.232544	0.216576	0.224654	0.173917	0.162797	0.142748	0.124529	0.148646
Probability at A_4	0.084524	0.069957	0.188417	0.163456	0.177236	0.162337	0.19027	0.204509	0.211683	0.209133	0.19464
Probability at A_5	0.047294	0.035533	0.057381	0.049096	0.057045	0.050825	0.066537	0.082119	0.085	0.091239	0.076025
Probability at A_6	1.42E-05	1.08E-05	0.000214	0.000199	0.000241	0.000227	0.000355	0.000404	0.000493	0.000567	0.000466
Probability at A_7	0.0019	0.001314	0.003083	0.002787	0.003471	0.00305	0.005961	0.006587	0.008745	0.010778	0.007714
Probability at A_8	2.84E-05	1.66E-05	6.34E-05	5.06E-05	6.75E-05	5.38E-05	0.000128	0.000153	0.000215	0.000276	0.000172
Probability at A_9	0.000665	0.000466	0.000629	0.000584	0.000728	0.000648	0.001455	0.0014	0.002135	0.002668	0.001909
Probability at A_{10}	0.00047	0.00033	0.000347	0.000309	0.000396	0.000363	0.000814	0.000773	0.001262	0.001513	0.001161
Probability at A_{11}	0.028162	0.023308	0.03611	0.036472	0.044171	0.04519	0.067931	0.074031	0.094286	0.10845	0.096834

Probability at A_{12}	0.017845	0.017924	0.013346	0.011902	0.009924	0.008483	0.006388	0.005737	0.004203	0.003517	0.003606

14.2.2 Probability Changes at Area 2

Probability at Locations	Sample Number										
	1	5	9	13	17	21	25	29	33	37	41
Probability at A_1	0.068129	0.107231	0.123569	0.147312	0.191963	0.158606	0.176342	0.147312	0.147312	0.133969	0.133969
Probability at A_2	0.512777	0.490644	0.520394	0.525546	0.505246	0.520797	0.518411	0.525546	0.525546	0.51928	0.51928
Probability at A_3	0.21611	0.201143	0.183242	0.192893	0.188023	0.202018	0.190274	0.192893	0.192893	0.193246	0.193246
Probability at A_4	0.018576	0.055224	0.040887	0.032643	0.037045	0.030188	0.0322	0.032643	0.032643	0.038074	0.038074
Probability at A_5	0.003345	0.015913	0.011946	0.00915	0.01176	0.008231	0.009539	0.00915	0.00915	0.010821	0.010821
Probability at A_6	2.86E-07	1.56E-06	1.35E-06	1.19E-06	1.8E-06	1.13E-06	1.4E-06	1.19E-06	1.19E-06	1.29E-06	1.29E-06
Probability at A_7	1.58E-05	0.000117	0.000106	8.49E-05	0.000132	7.53E-05	0.0001	8.49E-05	8.49E-05	9.5E-05	9.5E-05
Probability at A_8	1.06E-07	1.44E-06	1.1E-06	7.01E-07	1.11E-06	5.64E-07	7.83E-07	7.01E-07	7.01E-07	8.88E-07	8.88E-07
Probability at A_9	3.64E-06	2.62E-05	2.53E-05	2.1E-05	3.37E-05	1.89E-05	2.55E-05	2.1E-05	2.1E-05	2.29E-05	2.29E-05
Probability at A_{10}	2.69E-06	2.16E-05	1.74E-05	1.43E-05	2.32E-05	1.32E-05	1.71E-05	1.43E-05	1.43E-05	1.62E-05	1.62E-05
Probability at A_{11}	0.00048	0.002219	0.002078	0.002158	0.003557	0.002229	0.002768	0.002158	0.002158	0.002162	0.002162
Probability at A_{12}	0.18056	0.127459	0.117732	0.090176	0.062214	0.077823	0.070321	0.090176	0.090176	0.102311	0.102311

14.2.3 Probability Changes at Area 3

Probability at Locations	Sample Number										
	1	5	9	13	17	21	25	29	33	37	41
Probability at A_1	0.005386	0.007436	0.007618	0.007618	0.008573	0.009604	0.012213	0.011177	0.012972	0.015014	0.016391
Probability at A_2	0.170708	0.188935	0.18828	0.18828	0.189694	0.19026	0.199357	0.198225	0.205982	0.213446	0.214464
Probability at A_3	0.498389	0.493851	0.472144	0.472144	0.482312	0.490487	0.513938	0.483526	0.475417	0.466141	0.494995
Probability at A_4	0.033402	0.04245	0.05351	0.05351	0.06364	0.075349	0.078951	0.084121	0.093669	0.10401	0.097527
Probability at A_5	0.00137	0.002113	0.002935	0.002935	0.00374	0.004745	0.005255	0.005756	0.006964	0.008401	0.007663
Probability at A_6	1.64E-06	2E-06	2.74E-06	2.74E-06	3.65E-06	4.82E-06	6.05E-06	6.27E-06	8.13E-06	1.05E-05	1.01E-05
Probability at A_7	3.28E-06	5.5E-06	8.3E-06	8.3E-06	1.13E-05	1.54E-05	1.93E-05	2.15E-05	2.98E-05	4.13E-05	3.71E-05
Probability at A_8	4.33E-08	7.78E-08	1.39E-07	1.39E-07	2.03E-07	2.95E-07	3.51E-07	4.29E-07	6.21E-07	8.97E-07	7.32E-07
Probability at A_9	3.31E-07	5.94E-07	8.83E-07	8.83E-07	1.21E-06	1.64E-06	2.11E-06	2.32E-06	3.26E-06	4.58E-06	4.18E-06
Probability at A_{10}	2.65E-07	5.17E-07	7.69E-07	7.69E-07	1.08E-06	1.51E-06	1.92E-06	2.05E-06	2.76E-06	3.72E-06	3.49E-06
Probability at A_{11}	6.24E-05	0.000104	0.000132	0.000132	0.000175	0.000231	0.000315	0.000301	0.00039	0.000504	0.000528
Probability at A_{12}	0.290677	0.265102	0.275369	0.275369	0.251847	0.2293	0.189941	0.216862	0.204563	0.192423	0.168377

14.2.4 Probability Changes at Area 4

Probability at Locations	Sample Number										
	1	5	9	13	17	21	25	29	33	37	41
Probability at A_1	0.074468	0.081733	0.053506	0.05051	0.05019	0.037372	0.037372	0.032369	0.030088	0.035229	0.043938
Probability at A_2	0.076556	0.06735	0.058134	0.043988	0.048821	0.039497	0.039497	0.03821	0.02847	0.033334	0.04274
Probability at A_3	0.079248	0.067817	0.0636	0.046169	0.054155	0.047601	0.047601	0.044795	0.03202	0.03265	0.043636
Probability at A_4	0.413499	0.40632	0.456074	0.43653	0.452128	0.475657	0.475657	0.479652	0.452058	0.436158	0.442102
Probability at A_5	0.273119	0.272113	0.281121	0.313268	0.298634	0.305607	0.305607	0.321224	0.352469	0.359407	0.330702
Probability at A_6	0.00187	0.002261	0.002682	0.003339	0.00297	0.003953	0.003953	0.004577	0.005609	0.004162	0.003883
Probability at A_7	0.018053	0.02472	0.024164	0.033133	0.027507	0.030164	0.030164	0.027996	0.037798	0.036468	0.030884
Probability at A_8	0.001288	0.001812	0.00221	0.003385	0.002696	0.00349	0.00349	0.003284	0.004953	0.00446	0.003199
Probability at A_9	0.00187	0.002981	0.002609	0.003627	0.00297	0.002998	0.002998	0.002169	0.00297	0.00307	0.002636
Probability at A_{10}	0.000816	0.001301	0.001092	0.001497	0.001278	0.00122	0.00122	0.000847	0.001144	0.001216	0.001044
Probability at A_{11}	0.053073	0.066886	0.048908	0.060874	0.054155	0.048264	0.048264	0.040663	0.049836	0.050817	0.05151
Probability at A_{12}	0.006141	0.004705	0.005899	0.003678	0.004497	0.004177	0.004177	0.004212	0.002586	0.003028	0.003725

14.2.5 Probability Changes at Area 5

Probability at Locations	Sample Number										
	1	5	9	13	17	21	25	29	33	37	41
Probability at A_1	0.001478	0.000877	0.00082	0.000974	0.001267	0.001718	0.002183	0.002183	0.002509	0.002595	0.002358
Probability at A_2	0.000517	0.00029	0.000217	0.000258	0.000309	0.000523	0.000743	0.000743	0.000953	0.000986	0.001001
Probability at A_3	0.000363	0.000164	0.000102	0.000106	0.000134	0.000236	0.000309	0.000309	0.000391	0.000404	0.000498
Probability at A_4	0.108889	0.080614	0.062955	0.06164	0.066979	0.087156	0.095102	0.095102	0.10341	0.102631	0.117956
Probability at A_5	0.630232	0.582096	0.559332	0.578789	0.570915	0.595471	0.632046	0.632046	0.641353	0.645387	0.64598
Probability at A_6	0.015185	0.014418	0.01126	0.008478	0.008961	0.008484	0.006115	0.006115	0.005953	0.005363	0.006885
Probability at A_7	0.1259	0.15762	0.176332	0.172649	0.177512	0.159027	0.139092	0.139092	0.131714	0.128928	0.118774
Probability at A_8	0.07978	0.131691	0.153564	0.140315	0.129162	0.102174	0.080009	0.080009	0.070705	0.069209	0.065546
Probability at A_9	0.003707	0.004641	0.00564	0.005918	0.006431	0.005922	0.005475	0.005475	0.005184	0.005363	0.004547
Probability at A_{10}	0.000997	0.001087	0.001339	0.001445	0.001614	0.001507	0.001493	0.001493	0.001375	0.001422	0.001223
Probability at A_{11}	0.032933	0.02649	0.028431	0.02942	0.036708	0.037761	0.037404	0.037404	0.036413	0.037669	0.035185
Probability at A_{12}	1.94E-05	1.18E-05	7.3E-06	8.67E-06	9.04E-06	1.86E-05	2.9E-05	2.9E-05	4.1E-05	4.25E-05	4.75E-05

14.2.6 Probability Changes at Area 6

Probability at	Sample Number

Locations	1	5	9	13	17	21	25	29	33	37	41
Probability at A_1	9.59E-07	1.62E-06	1.13E-05	1.2E-05	1.07E-05	1.67E-05	1.35E-05	2.1E-05	1.18E-05	1.84E-05	1.84E-05
Probability at A_2	2.29E-08	2.23E-08	1.38E-06	8.24E-07	6.04E-07	9.41E-07	8.29E-07	1.29E-06	7.26E-07	1.13E-06	1.13E-06
Probability at A_3	3.02E-07	1.34E-06	4.22E-05	3.15E-05	2.31E-05	3.13E-05	2.61E-05	3.53E-05	2.29E-05	3.09E-05	3.09E-05
Probability at A_4	0.00182	0.002032	0.014919	0.010539	0.008051	0.010336	0.010033	0.01283	0.009155	0.011724	0.011724
Probability at A_5	0.006142	0.001697	0.007578	0.00528	0.004204	0.005704	0.006099	0.008243	0.005489	0.007429	0.007429
Probability at A_6	0.814285	0.947613	0.937498	0.948696	0.955517	0.943374	0.941373	0.925737	0.946306	0.931926	0.931926
Probability at A_7	0.03836	0.015611	0.01678	0.015414	0.014289	0.018345	0.01882	0.024067	0.017412	0.022299	0.022299
Probability at A_8	0.134973	0.028288	0.019536	0.015202	0.013335	0.015977	0.018306	0.021847	0.016937	0.020242	0.020242
Probability at A_9	0.000325	0.000486	0.000345	0.000379	0.000342	0.00047	0.000457	0.000626	0.0004	0.000549	0.000549
Probability at A_{10}	6.37E-05	0.000243	9.53E-05	0.000125	0.000112	0.000158	0.000149	0.00021	0.00013	0.000184	0.000184
Probability at A_{11}	0.004029	0.004027	0.003194	0.004321	0.004117	0.005587	0.004723	0.006383	0.004135	0.005596	0.005596
Probability at A_{12}	5.2E-11	4.65E-11	2.28E-08	7.14E-09	4.14E-09	6.44E-09	6.52E-09	1.01E-08	5.71E-09	8.87E-09	8.87E-09

14.2.7 Probability Changes at Area 7

Probability at Locations	Sample Number										
	1	5	9	13	17	21	25	29	33	37	41
Probability at A_1	0.000117	0.000456	0.000434	0.000371	0.000371	0.000304	0.000304	0.000304	0.000304	0.000304	0.000304
Probability at A_2	4.85E-06	1E-05	8.56E-06	9.64E-06	9.64E-06	7.69E-06	7.69E-06	7.69E-06	7.69E-06	7.69E-06	7.69E-06
Probability at A_3	1.45E-05	3.71E-05	3.25E-05	2.97E-05	2.97E-05	1.98E-05	1.98E-05	1.98E-05	1.98E-05	1.98E-05	1.98E-05
Probability at A_4	0.015988	0.018525	0.016462	0.018544	0.018544	0.015427	0.015427	0.015427	0.015427	0.015427	0.015427
Probability at A_5	0.019672	0.015912	0.013195	0.017546	0.017546	0.016994	0.016994	0.016994	0.016994	0.016994	0.016994
Probability at A_6	0.105522	0.114104	0.10716	0.079731	0.079731	0.055417	0.055417	0.055417	0.055417	0.055417	0.055417
Probability at A_7	0.318911	0.364453	0.361736	0.364817	0.364817	0.363245	0.363245	0.363245	0.363245	0.363245	0.363245
Probability at A_8	0.438292	0.257953	0.252514	0.30062	0.30062	0.329739	0.329739	0.329739	0.329739	0.329739	0.329739
Probability at A_9	0.054344	0.107965	0.123048	0.111103	0.111103	0.116914	0.116914	0.116914	0.116914	0.116914	0.116914
Probability at A_{10}	0.013638	0.035233	0.041856	0.037793	0.037793	0.040884	0.040884	0.040884	0.040884	0.040884	0.040884
Probability at A_{11}	0.033497	0.085352	0.083553	0.069436	0.069436	0.061048	0.061048	0.061048	0.061048	0.061048	0.061048
Probability at A_{12}	6.53E-08	6.59E-08	5.39E-08	8.35E-08	8.35E-08	6.94E-08	6.94E-08	6.94E-08	6.94E-08	6.94E-08	6.94E-08

14.2.8 Probability Changes at Area 8

Probability at Locations	Sample Number										
	1	5	9	13	17	21	25	29	33	37	41
Probability at A_1	0.000128	7.75E-05	6.88E-05	9.4E-05	5.04E-05	5.3E-05	4.19E-05	6.28E-05	5.25E-05	5.22E-05	8.2E-05

Probability at A_{2}	3.83E-06	3.05E-06	2.11E-06	2.66E-06	9.14E-07	1.07E-06	9.23E-07	1.14E-06	1.16E-06	1.15E-06	1.66E-06
Probability at A_{3}	1.2E-05	8.31E-06	5.15E-06	5.49E-06	1.74E-06	1.99E-06	1.62E-06	1.89E-06	1.76E-06	1.75E-06	2.33E-06
Probability at A_{4}	0.01265	0.011877	0.008218	0.008176	0.00438	0.00494	0.004548	0.004499	0.004694	0.004734	0.005262
Probability at A_{5}	0.013744	0.015445	0.009436	0.011237	0.007828	0.00946	0.008955	0.008497	0.009766	0.009067	0.01036
Probability at A_{6}	0.10416	0.101943	0.062281	0.053964	0.03274	0.034937	0.031292	0.025858	0.024831	0.026108	0.02294
Probability at A_{7}	0.337322	0.316731	0.318299	0.339351	0.311703	0.314723	0.302065	0.320134	0.311703	0.314418	0.335262
Probability at A_{8}	0.387334	0.453729	0.449715	0.429257	0.478484	0.489844	0.517916	0.458603	0.498745	0.496183	0.448198
Probability at A_{9}	0.088237	0.063712	0.102448	0.104786	0.107505	0.094531	0.089484	0.118315	0.098947	0.098439	0.114043
Probability at A_{10}	0.0284	0.018615	0.033898	0.033726	0.039732	0.033517	0.030862	0.044952	0.035083	0.034903	0.042734
Probability at A_{11}	0.02801	0.017858	0.01563	0.0194	0.017575	0.017992	0.014832	0.019077	0.016176	0.016093	0.021114
Probability at A_{12}	5.1E-08	5.57E-08	3.91E-08	4.05E-08	9.48E-09	1.16E-08	1.15E-08	1.18E-08	1.43E-08	1.43E-08	1.8E-08

14.2.9 Probability Changes at Area 9

Probability at Locations	Sample Number										
	1	5	9	13	17	21	25	29	33	37	41
Probability at A_{1}	0.001285	0.001834	0.001344	0.001547	0.001062	0.000923	0.000923	0.001099	0.000949	0.00092	0.00092
Probability at A_{2}	1.2E-05	1.01E-05	7.23E-06	9.29E-06	6.2E-06	4.83E-06	4.83E-06	6.42E-06	4.96E-06	4.81E-06	4.81E-06
Probability at A_{3}	1.28E-05	6.84E-06	4.74E-06	5.93E-06	3.8E-06	3.04E-06	3.04E-06	4.27E-06	3.39E-06	3.52E-06	3.52E-06
Probability at A_{4}	0.007438	0.002894	0.002305	0.002842	0.002401	0.001948	0.001948	0.002416	0.001947	0.002079	0.002079
Probability at A_{5}	0.013856	0.005032	0.003845	0.005079	0.004727	0.003578	0.003578	0.004085	0.003073	0.003236	0.003236
Probability at A_{6}	0.009344	0.0017	0.001392	0.001623	0.00149	0.001278	0.001278	0.001361	0.00116	0.001188	0.001188
Probability at A_{7}	0.349612	0.220717	0.204653	0.225878	0.228394	0.206964	0.206964	0.205778	0.185259	0.182063	0.182063
Probability at A_{8}	0.134681	0.053879	0.054278	0.060742	0.072501	0.064797	0.064797	0.061808	0.054881	0.055446	0.055446
Probability at A_{9}	0.237396	0.353162	0.370845	0.356458	0.355479	0.369884	0.369884	0.362714	0.374961	0.363435	0.363435
Probability at A_{10}	0.148366	0.279193	0.297254	0.274114	0.273361	0.296483	0.296483	0.303048	0.326547	0.339161	0.339161
Probability at A_{11}	0.097997	0.081572	0.064073	0.071703	0.060575	0.054138	0.054138	0.05768	0.051215	0.052463	0.052463
Probability at A_{12}	4.35E-08	2.71E-08	2.13E-08	2.85E-08	1.99E-08	1.48E-08	1.48E-08	2.17E-08	1.61E-08	1.56E-08	1.56E-08

14.2.10 Probability Changes at Area 10

Probability at Locations	Sample Number										
	1	5	9	13	17	21	25	29	33	37	41
Probability at A_{1}	0.003394	0.007267	0.008551	0.005951	0.009531	0.008155	0.008155	0.008155	0.011054	0.011054	0.009073
Probability at A_{2}	5.08E-05	0.000106	9.43E-05	5.26E-05	9.68E-05	7.41E-05	7.41E-05	7.41E-05	0.000103	0.000103	7.59E-05
Probability at A_{3}	3.98E-05	9.93E-05	0.000111	5.92E-05	0.000118	9.31E-05	9.31E-05	9.31E-05	0.000133	0.000133	9.94E-05

Probability at A_4	0.00615	0.010556	0.008318	0.005869	0.009018	0.007201	0.007201	0.007201	0.008983	0.008983	0.007793
Probability at A_5	0.003637	0.006506	0.003888	0.003195	0.004709	0.003509	0.003509	0.003509	0.004563	0.004563	0.004241
Probability at A_6	0.000514	0.000824	0.000658	0.000604	0.000775	0.000654	0.000654	0.000654	0.000794	0.000794	0.000769
Probability at A_7	0.117685	0.131592	0.095438	0.096477	0.10205	0.091017	0.091017	0.091017	0.097532	0.097532	0.097148
Probability at A_8	0.024002	0.021217	0.011833	0.013361	0.012479	0.010977	0.010977	0.010977	0.010827	0.010827	0.011556
Probability at A_9	0.402795	0.351171	0.340484	0.348981	0.321478	0.329233	0.329233	0.329233	0.311522	0.311522	0.310295
Probability at A_{10}	0.381125	0.356059	0.413193	0.41769	0.39556	0.422256	0.422256	0.422256	0.394056	0.394056	0.403507
Probability at A_{11}	0.060607	0.114601	0.117431	0.10776	0.144184	0.12683	0.12683	0.12683	0.160433	0.160433	0.155443
Probability at A_{12}	4.98E-07	7.44E-07	5.11E-07	2.35E-07	4.32E-07	3.17E-07	3.17E-07	3.17E-07	4.01E-07	4.01E-07	2.68E-07

14.2.11 Probability Changes at Area 11

Probability at Locations	Sample Number										
	1	5	9	13	17	21	25	29	33	37	41
Probability at A_1	0.141613	0.143711	0.128264	0.0978	0.108039	0.089499	0.078352	0.085776	0.099869	0.099869	0.095873
Probability at A_2	0.004468	0.003345	0.002393	0.002095	0.003137	0.002083	0.001824	0.001838	0.00239	0.00239	0.002054
Probability at A_3	0.006356	0.003921	0.002344	0.001787	0.002429	0.001681	0.00169	0.0018	0.002308	0.002308	0.00204
Probability at A_4	0.073945	0.048885	0.036453	0.040934	0.053379	0.043612	0.046334	0.043568	0.047997	0.047997	0.042999
Probability at A_5	0.055313	0.044376	0.040716	0.046357	0.062147	0.050775	0.051042	0.046686	0.047997	0.047997	0.040127
Probability at A_6	0.002449	0.001735	0.001293	0.001413	0.001518	0.001464	0.002023	0.001955	0.001929	0.001929	0.001826
Probability at A_7	0.072428	0.071497	0.076375	0.108485	0.114974	0.120478	0.127996	0.115466	0.11078	0.11078	0.110851
Probability at A_8	0.003459	0.003143	0.0035	0.006555	0.007241	0.008019	0.009129	0.007476	0.006693	0.006693	0.006606
Probability at A_9	0.042242	0.046576	0.054057	0.084585	0.076998	0.092647	0.091854	0.084015	0.080606	0.080606	0.092616
Probability at A_{10}	0.0373	0.040005	0.047077	0.070671	0.058398	0.075295	0.072615	0.068281	0.063723	0.063723	0.076318
Probability at A_{11}	0.560412	0.592799	0.607523	0.539312	0.511727	0.514439	0.517137	0.543135	0.535702	0.535702	0.528685
Probability at A_{12}	1.45E-05	7.79E-06	4.59E-06	5.84E-06	1.15E-05	6.67E-06	5.84E-06	5.12E-06	7.34E-06	7.34E-06	6.05E-06

14.2.12 Probability Changes at Area 12

Probability at Locations	Sample Number										
	1	5	9	13	17	21	25	29	33	37	41
Probability at A_1	0.00459	0.001554	0.000872	0.000275	0.000124	0.000184	0.000274	0.000333	0.000406	0.000494	0.0006
Probability at A_2	0.252903	0.161699	0.133611	0.079721	0.0543	0.064525	0.077184	0.083991	0.091568	0.099744	0.10855
Probability at A_3	0.081964	0.056938	0.074245	0.041916	0.027772	0.036355	0.036333	0.040088	0.041354	0.042622	0.04389
Probability at A_4	0.007243	0.002894	0.001624	0.000499	0.000244	0.000445	0.000555	0.000713	0.000832	0.000972	0.001133
Probability at A_5	0.001304	0.000364	0.000126	2.77E-05	1.12E-05	2.31E-05	3.45E-05	4.74E-05	6.02E-05	7.63E-05	9.67E-05

Probability at A_6	2.44E-08	4.82E-09	1.28E-09	2.11E-10	6.91E-11	1.51E-10	2.66E-10	3.82E-10	5.19E-10	7.05E-10	9.58E-10
Probability at A_7	1.14E-06	1.85E-07	2.91E-08	4.3E-09	1.33E-09	2.95E-09	6.22E-09	9.17E-09	1.34E-08	1.95E-08	2.83E-08
Probability at A_8	1.29E-08	1.78E-09	1.93E-10	2.28E-11	6.5E-12	1.69E-11	3.79E-11	5.98E-11	9.09E-11	1.38E-10	2.09E-10
Probability at A_9	1.09E-07	1.67E-08	2.49E-09	3.57E-10	1.08E-10	2.35E-10	5.16E-10	7.62E-10	1.13E-09	1.66E-09	2.45E-09
Probability at A_{10}	6.51E-08	1.04E-08	2.17E-09	2.82E-10	8.16E-11	2.02E-10	3.81E-10	5.78E-10	8.19E-10	1.16E-09	1.64E-09
Probability at A_{11}	1.71E-05	3.11E-06	9.76E-07	1.48E-07	4.46E-08	9.74E-08	1.72E-07	2.46E-07	3.35E-07	4.55E-07	6.18E-07
Probability at A_{12}	0.651978	0.776549	0.789522	0.877561	0.91755	0.898469	0.885619	0.874828	0.865779	0.856091	0.84573

14.3 The Number of Samples Required to Give Correct Localization Result at Different Locations

Area\ Sample Number	1	5	9	13	17	21	25	>=25	Total
1	68	4	2	1	1	2	2	0	80
2	64	12	3	1	0	0	0	0	80
3	60	16	3	1	0	0	0	0	80
4	59	11	7	3	0	0	0	0	80
5	68	7	3	1	1	0	0	0	80
6	60	12	4	2	1	1	0	0	80
7	57	8	5	5	3	1	0	1	80
8	70	5	3	1	0	1	0	0	80
9	42	13	11	5	4	2	1	2	80
10	56	8	4	7	3	1	0	1	80
11	70	4	2	2	1	1	0	0	80
12	55	12	5	2	2	1	2	1	80