國立臺灣大學電機資訊學院電信工程學研究所

碩士論文

Graduate Institute of Communication Engineering
College of Electrical Engineering and Computer Science
National Taiwan University

Master thesis

無線網路下藉由混和網路存取點的重要性之精準的室內定位技術

Accurate Indoor Location Estimation by Incorporating the Importance of Access Points in Wireless Local Area Networks

黃浩儒

Hao-Ru Huang

指導教授：林宗男 博士

Advisor: Tsung-Nan Lin, Ph.D.

中華民國 98 年 7 月

July, 2009

國立臺灣大學（碩）博士學位論文
口試委員會審定書

論文中文題目：無線網路下藉由混和網路存取點的重要性之精準的室內定位技術

論文英文題目：Accurate Indoor Location Estimation by Incorporating the Importance of Access Points in Wireless Local Area Networks

本論文係黃浩儒君（r96942101）在國立臺灣大學電信工程所、所完成之碩（博）士學位論文，於民國 98 年 7 月 15 日承下列考試委員審查通過及口試及格，特此證明

口試委員：

林宗男 （簽名）
（指導教授）

蔡子傑　　陳俊良

系主任、所長　王暉　（簽名）

誌謝

這篇論文首先要感謝我的指導教授林宗男博士，在老師的指導下，不論是課業研究或是生活上都有很大的啟發，尤其在研究上的嚴謹的態度以及邏輯思考的指正使我能夠順利完成我的研究。其次也要感謝口試委員蔡子傑以及陳俊良教授給我良好的研究意見，使論文更加完善。

接下來要感謝實驗室的學長方士豪，從碩一剛進來一直到畢業都給予我很多的指導以及意見。感謝伯江學長在網路領域的指導。感謝同屆的大頭以及震謙使我在碩士班的生活不無聊。感謝實驗室的學弟書銘，育宏，武億與琮訓，在課餘之間帶動研究室的氣氛。感謝承先在最後的量測以及整理資料幫了我很大的忙。感謝研究所籃球隊的所有隊友，在研究之於一起揮灑汗水。最後感謝我的家人，感謝老爸老媽支持，感謝老姊老哥在各方面給我意見，感謝純綺，感謝每位曾幫助過我的人，因為有你們我才能順利的完成學業。

摘要

本篇論文的焦點是無線網路的是內定位。我們觀察到在做定位估算時，每個無線網路存取點(AP)的貢獻度是不同的。這篇論文的主要貢獻可以分為兩個部分。第一，我們提供一個新的方法去量測 AP 階級的重要性。在不同的地方利用訊號的鑑別度量化每一個 AP 的重要性。我們利用如此的數值關連來選擇重要的 AP 來避免不必要的計算。選擇重要的 AP 意味著佈建 AP 在較有鑑別度的位置。我們的方法可以有效的應用在 AP 的佈建。第二，這些重要性進一步地嵌入我們的定位系統。我們提供一個加權重的零核函數這些權重影響的 AP 是有區別的。那就是愈重要的 AP 分配愈大的權重。此外，我們發展一個概似熵的函數來避免突然劇烈改變的權重。我們的定位系統是發展在一個真實的無線網路環境，我們蒐集真實量測到的訊號強度(RSS)。實驗結果顯示考慮這些不同的重要性在定位的精準度有重大的提升。

關鍵字:室內定位，無線網路，選擇無線網路存取點，特徵指紋定位，無線網路存取點的布建

Abstract

This study focuses on indoor localization in Wireless Local Area Networks (WLANs). We investigate the unequal contribution of each access point (AP) on location estimation. The main contribution is two parts. First, a novel mechanism is proposed to measure the degrees of the AP importance. The importance of each AP is quantified by the signal discrimination between distinct locations. We utilize such numerical relevancies to select important APs to avoid unnecessary calculations. To select the important AP means place the AP in the discriminative positions. Our method can efficiently apply to AP placement. Second, the importance is further embedded into our positioning system. We provide a weighted kernel function where the effect of APs is differentiated. That is, the larger weights are assigned to the more important APs. Moreover, we develop a quasi entropy function to avoid an abrupt change on the weights. Our positioning system is developed in a real-world WLAN environment, where the realistic measurement of receive signal strength (RSS) is collected. Experimental results show that the positioning accuracy is significantly improved by taking the different importance into consideration.

Keyword: indoor localization, wireless local area networks, AP selection, location fingerprinting, AP placement

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Nowadays wireless local area network (WLAN) is one of the most popular wireless infrastructures in an indoor environment. Positioning in such an environment is highly desirable for many location aware applications such as museum tour guide and fraud detection [1–3]. Thus, the location fingerprinting technique is developed to provide a high accuracy in a challenging indoor environments [4–6]. This method collects the received signal strength (RSS) at the sample locations to build a radio map for the target environment. When a user wants to estimate his/her location, the positioning system measures RSS in real-time and estimate the location by matching the measurements with the previously stored radio map.

In WLAN location fingerprinting, the multi-dimensional measurements are described by RSSs from the detectable APs [7]. Although signals sent from every AP deployed in the area are mutually independent, RSS from different AP has different importance to the estimation of user location. For radio signals, the long distance they travel, the more time they are influenced by the environmental noises, and the more uncertainties factors are added in the signal strength received by the user. Thus, it is believed that some RSSs are strongly relevant, some are weakly relevant and others are irrelevant to the location information [8–10]. In other words, the RSS from different AP has different contribution to the location estimation. These importance should be embedded in the positioning algorithm. However, traditional approaches treat each RSS in an equal way. That is, the effect of APs on the location computing has not been differentiated. In this article, we argue that it would be more profitable to take such properties into consideration while designing a location system. Two questions arise from such considerations. The first question is how to effectively quantify the importance of each AP. The second is how the different importance is embedded in the location estimation.

The proposed positioning algorithm contains the answers to these two questions. First, a novel mechanism is proposed to measure the degrees of the AP relevancies. The importance is quantified by the signal discrimination between distinct locations through a quasi entropy function. In our approach, we utilize such numerical importance to select reliable APs so as to avoid unnecessary calculations.

Next, the importance is further embedded into our positioning system. Our fingerprinting system adopts the kernel-based method which computes the similarity between the online measurements and the training data. The kernel function in our algorithm is modified by incorporating the previously quantified importance. We assign larger weights to the relevant APs and smaller weights to the irrelevant APs. This way, RSS from different APs are fused with different importance and the location computation can be dominated by the more relevant ones. In the experiments, our localization system is developed by collecting realistic RSS data in an indoor WLAN environment. Experimental results indicate superior performance of our algorithm, as compared to the existing methods.

In the following chapter, we illustrate the location fingerprinting systems and introduce the kernel approach model. Chapter 3 illustrates the important of AP selection and reviews the existing AP selection methods. Then we propose our AP selection method by calculating the discrimination of RSS.

Chapter 4 proposes our positioning system which contains four stages. The first stage is offline modeling to build the radio map. In the second stage, we calculate the importance of each AP and select the most important APs for positioning. The third stage evaluates the weights of each AP to the kernel distance. Finally, the weights are incorporating in the positioning system.

In chapter 5, we implemented our algorithms in an indoor environment. All experiments are conducted in realistic environment of fifth floor of BL building in NTU. The results show that our AP selection method can correctly select the more important APs. The significant reduction of mean error, as compared to the existing algorithms.

In chapter 6, we introduce a significant application of AP selection. That is, AP selection is an efficient application in AP placement.

Chapter 2

Background Description

2.1 Characteristics of Signal Propagation

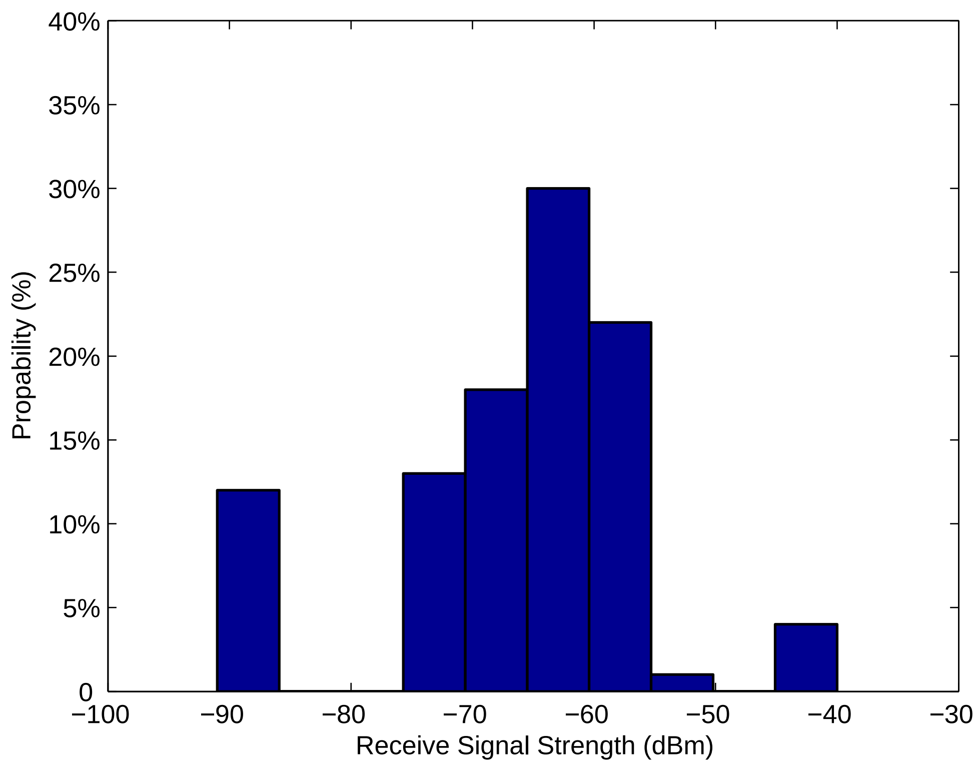


Figure 2.1: An example of receive signal strength distribution.

The IEEE 802.11b standard works over the radio frequencies in the 2.4 GHz band. It is widespread since the band is license-free in most places around the world. It is attractive because the RF-based techniques are popular and inexpensive, providing much ubiquitous coverage and requiring little overhead. However, accurate location estimation using measurements of signal strength is a longstanding difficult task due to the noisy characteristics of signal propagation. Subject to reflection, refraction, diffraction, and absorption by structures and even human bodies, signal propagation suffers from severe multipath fading effects in an indoor environment. As a result, a transmitted signal can reach the receiver through different paths, each having its own amplitude and phase. These different components combine and reproduce a distorted version of the original signal. Moreover, even changes in the environmental conditions, such as temperature or humidity, also affect the signals to a large extent. As a consequence, the signal strength received from an access point at a fixed location varies with time and its physical surroundings.

Fig. 2.1 gives a typical example of the normalized histogram of the signal strength received from an access point at a fixed location. Several hundred measurements were sampled to construct the histogram. It is obvious that the signal strength received from the same AP varies with time, even at a fixed location. Furthermore, the number of APs covering a location also varies with time.

2.2 Overview of indoor positioning systems

Indoor positioning systems in the wireless networks could provide ubiquitous computing in the indoor environments where the global positioning system (GPS) does not work well [11]. In the past years, many developed indoor positioning systems utilize the location features such as the angle of arrival, time of arrival (TOA) [12] and time difference of arrival [13]. The mentioned two measurements need to be precisely measured and require the line-of-sight (LOS) between the transmitter and the receiver [14]. Meanwhile, such features require specialized hardware integrated into the existing equipments. Due to the high implementation cost, using received signal strength (RSS) gets more interests. Since the WLAN infrastructures are widespread, the RSS-based positioning system is a cost effective solution and is growing rapidly in commercial interest. The most viable solution for RSS-based indoor positioning is location fingerprinting which works like the process of pattern recognition. A user's location is estimated by exploiting the function between currently measured RSS pattern and a pre-stored radio map [6, 15]. In general, two stages of the fingerprinting are the offline modeling and the online positioning [4] and they are presented as follows.

2.3 Two stages of Location Fingerprinting

During the offline stage, the received signal strength (RSS) from different APs is collected different sampling locations to build the databased called "radio map" for the target environment. A radio map thus provides a model of RSS in a development area. A visual picture of the collected fingerprints is reported in Fig. 2.2. This figure shows a typical radio map includes 3 information sources, 5 locations and 50 samples RSS at each location. After constructing the radio map, a wireless client's is estimated by inspecting currently measured RSS. We describe several location estimation methods in the next subsection.

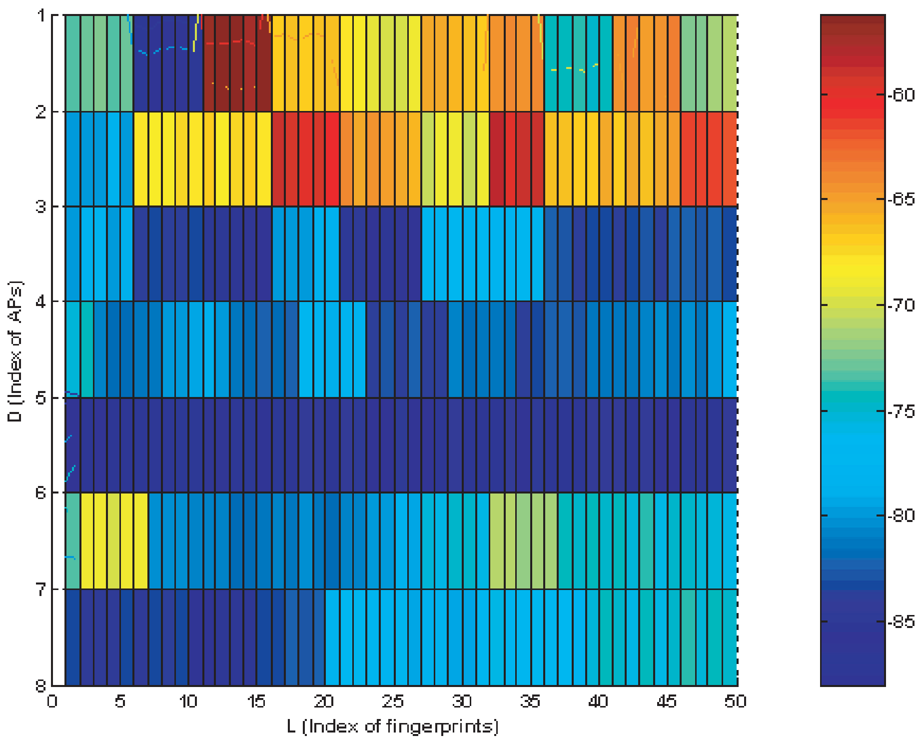


Figure 2.2: A visual picture of the collected WLAN RSS at a fixed indoor location based on temporal and access point diversity.

During the online stage, the positioning systems measure the RSS in real-time and estimates the location by comparing the measured RSS with the pre-recorded radio map. However, the signals of the indoor environment sufer from noise, interference, and multipath. The RSS can be considered as a random variable. Thus, the fundamental objective is seeking a mapping between the radio measurements to a physical location. One of the most popular mapping function is the probabilistic models [16, 17]. The main idea can be regarded as finding $p(\mathbf{l_r}|\mathbf{X})$, where $\mathbf{X}$ is an observed RSS vector, $\mathbf{l_r}$ represents the r-th reference location in the radio map and $p(\mathbf{l_r}|\mathbf{X})$ indicates the posteriori probability of location $\mathbf{l}_r$ given the observation $\mathbf{X}$. By means of Bayes' rule, $p(\mathbf{l_r}|\mathbf{X})$ depends only on the likelihood $p(\mathbf{X}|\mathbf{l_r})$ when the prior probability $\mathbf{p}(\mathbf{l_r})$ follows a uniform distribution. Thus, the location can be regarded as a multivariate multiple regression problem [18] and estimated as

$$\widehat{\mathbf{l}} = \sum_{r=1}^{R} \mathbf{l_r} \cdot p(\mathbf{X}|\mathbf{l_r}) \tag{2.1}$$

where R is the number of reference locations and $\widehat{\mathbf{l}}$ represents the estimated result.

In this artitle, we use kernel approach [18, 19] to compute the likelihood function $p(\mathbf{X}|\mathbf{l_r})$ from data.

2.4 Kernel Approach (Distance Calculation)

This section discusses the distance calculation between an observation and radio maps in the kernel space. It requires that these are decreasing functions of the distance between an observation vector and the training record. That is, survey points whose training records closely match the observation should receive a higher $p(\mathbf{X}|\mathbf{l_r})$. In particular, the likelihood functions $p(\mathbf{X}|\mathbf{l_r})$ should satisfy

$\sum_{r=1}^{R} p(X|l_r) = 1$. Then using the average normalized inner product as the likelihood functions for reasons that will become clear shortly:

$$p(X|l_r) = \frac{1}{n_r} \sum_{t=1}^{n_r} \frac{\langle X, X_r(t) \rangle}{\|X\| \, \|X_r(t)\|} \tag{2.2}$$

where $\langle X, X_r(t) \rangle = X X_r(t)^T$ denotes the canonical inner product in $\Re^D$ and D is the number of APs used for positioning. As seen in Eq. 2.2 the likelihood function is the average of the cosines between the observed RSS vector and the training vectors. The minimum value of Eq. 2.2 occurs when the observed RSS vector is orthogonal to all training vectors. However, this angular measurement can not recognize similarity measurement effictively between two RSS vectors as the maximum angle between them is very small in $\Re^D$. Furthermore, the presence of users and non-line-of-sight (NLOS) propagation results in various RSS distributions for a given survey poins, as shown in Fig. 2.3. In order to improve the complexity of RSS patterns, it can be achieved by using the nonlinear mapping $\phi : X \in \Re^D \mapsto \phi(X) \in \Im$ to map the input data to a higher dimensional space $\Im$.

At first glance, the calculation of $p(\mathbf{X}|\mathbf{l_r})$ in a possibly infinite dimensional space may seem computationally intractable. Fortunately, the kernel trick can be used to calculate the inner product in $\Im$ without the need for explicit evaluation of the mapping ϕ. The kernel trick allows the replacement of inner products in $\Im$ by a kernel evaluation on the input vectors. In the WLAN context, the kernel is a function $k : \Re^D \times \Re^D \mapsto \Re$ such that $k(x, x') = \langle \phi(x), \phi(x') \rangle$. Since the training data only enter $p(\mathbf{X}|\mathbf{l_r})$ throught inner products, the kernel trick can be used to carry out inner products in $\Im$ without the need for explicit evaluation of mapping

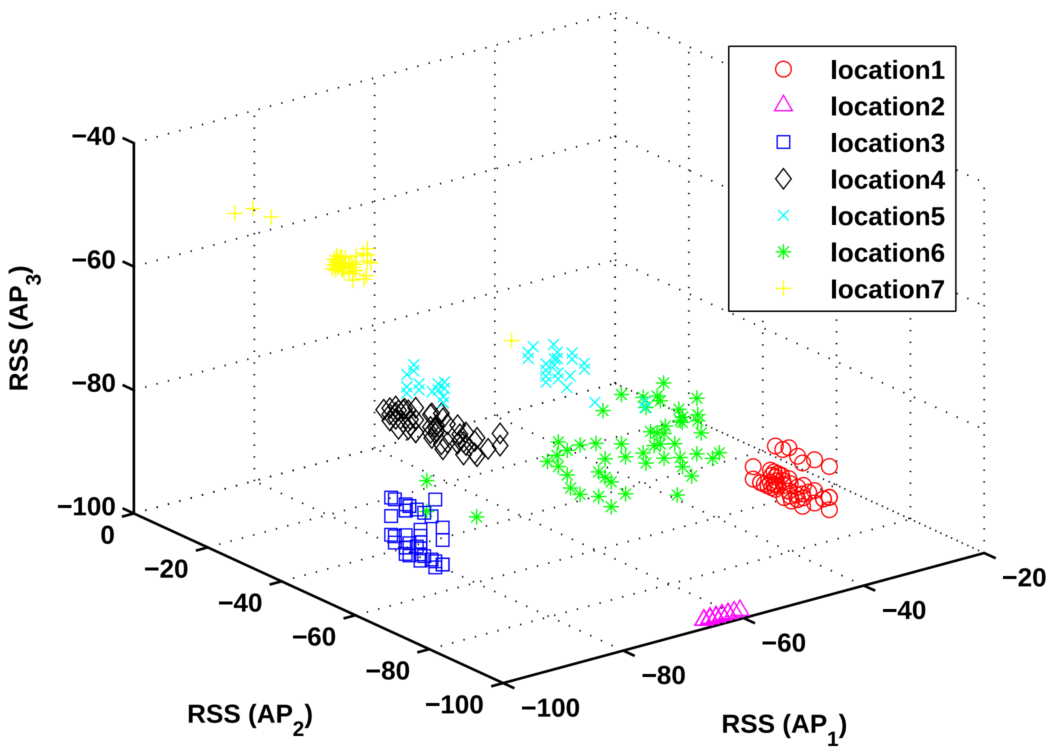


Figure 2.3: An example of the RSS measurement space for various locations in the BL building.

ϕ. The kernelized likelihood function then becomes

$$\begin{aligned} p(\mathbf{X}|l_r) &= \frac{1}{n_r}\sum_{t=1}^{n_r}\frac{\langle\phi(X),\phi(X_r(t))\rangle}{\|\phi(X)\|\,\|\phi(X_r(t))\|} \\ &= \frac{1}{n_r}\sum_{t=1}^{n_r}\frac{k(\mathbf{X},\mathbf{X}_r(t))}{\sqrt{k(\mathbf{X},\mathbf{X})k(\mathbf{X}_r(t),\mathbf{X}_r(t))}} \\ &= \frac{1}{n_r}\sum_{t=1}^{n_r}\hat{K}(\mathbf{X},\mathbf{X}_r(t)) \end{aligned} \tag{2.3}$$

where n_r is the number of collected RSS at the r-th location and $\mathbf{X}_r(t)$ is the t-th collected RSS at the r-th location. The function $k()$ and $\hat{K}()$, respectively, indicate a certain nonlinear kernel and its normalized form. The widely used Gaussian Radial Basis Function (RBF) is defined as

$$k(\mathbf{X},\mathbf{X}_r(t)) = \exp\left(\frac{-1}{2\sigma_r^2}||\mathbf{X}-\mathbf{X}_r(t)||^2\right) \tag{2.4}$$

where σ_r is an adjustable width and the operation $\|(\cdot)\|$ represents the norm function. The most commonly used L^2 norm is adopted which represents the Euclidean distance as $(||\mathbf{X}|| = \sqrt{x_1^2+\cdots+x_D^2})$.

Chapter 3

Access Point (AP) selection

In a tipycal WLAN environment, signals from many APs are detectable here or there within the area of concern. For example, Signals from each AP provide some information for location estimation, and it is a natural way to use as many as possible to improve the accuracy in a location estimation system. However, the increase of accuracy is at the cost of adding mroe computational burden to the system. Using all availiable APs increases the computational complexity of the positioning algorithm. As a consequence, such a location system not only has poor scalability but also is power-insufficient when energy is constrained on the computational unit. Therefore, it is important to only use the number of APs that a target system can afford while maintaining as high a level of accuracy as possible.

Furthermore, the geometric configuration of APs in relation to each other can affect the accuracy of positioning [41]. Since RSS is dependent on the relative distance of user and each AP, as well as the topology of the environment in terms of obstacles causing non-light-of-sight (NLOS) propagation, subsets of available APs may report correlated readings, leading to needless redundancy and possibly biased estimates. This motivates the used of AP selection techniques to select a subset of available APs for positioning. Thus, in order to reduce computational complexity cost and enhance accuracy, an AP selection method is needed.

More importantly, the results in [29, 39] showed that the best positioning accuracy can be produced by using a subset of RSSs in a fingerprinting system. This occurs because, as the number of RSSs increases, more information is added whereas more noise is incurred [29]. Kushki et al. [18] pointed out that the distinct transmitters may produce similar measurements, leading to biased estimates and redundant computation. These works motivate the use of information selection techniques from the view point of performance. The AP selection techniques choose the subset of APs for positioning depend on the value of quantified importance. We introduce the importance quantification method for the following section.

3.1 Importance Quantification for AP selection

The importance quantification methods are originally designed for AP selection. In these methods, some importance evaluation function is used to rank the sensed RSSs according to their estimated importance. Then, the more important APs are selected for positioning. This way, several advantages can be accomplished such as improving the speed of positioning, better power efficient, reducing the storage requirement and avoiding the problem of overfitting. Existing AP selection methods performed the above advantages in positioning system.

For example, Youssef et al. [38] utilized the strongest RSSs to reduce the computational complexity of the positioning algorithm. They mentioned that the strongest APs provide the highest probability of coverage over time. This method named MaxMean, assign the higher importance to the stronger RSS. However, it is also known that the variance measurements from an AP increases with its mean power at a given location. In cases where the measured RSS from an AP exhibits a high degree of variance, the survey values may be very different than the online measurement, degrading the accuracy of estimation [42]. Furthermore, it becomes more difficult to distinguish neighboring points in such cases. Chen et al. [39] provides a selection method based on the discriminant power of each AP quantified through the entropy-based InfoGain criterion. The InfoGain criterion assigns the more importance to the more discriminative APs instead. Thus, InfoGain ranks APs in descending order of their InfoGain values which are calculated as follows:

$$InfoGain(AP_d) = H(G) - H(G|AP_d) \tag{3.1}$$

where $H(G)$ and $H(G|AP_d)$, referred to [39], implies the "entropy of the reference locations when AP_d's value is unknown", and the "conditional entropy of the reference locations given AP_d's value". The recent work of Kushki et al. [18] offers a real-time RSS selection technique which minimizes the correlation between selected RSSs based on different divergence measurements such as Bhattacharyya distance and information potential.

3.2 Proposed AP selection method

In this article, we propose a novel AP selection method in this section. A subset of APs is selected throught an importance evaluation function. In our method, the importance of each AP is quantified by calculating the signal discrimination between different locations. The signal discrimination can be regarded as the signal scatter. The important AP means it has good discrimination. Good discrimination represents the d-th AP has good separation in different location that it can recognize different locaion accurately. Thus, we define the total scatter of d-th AP $S_{T,d}$

$$S_{T,d} = \frac{1}{R \cdot n_r} \sum_{r=1}^{R} \sum_{t=1}^{n_r} (x_{r,d}(t) - m_d)^2 \tag{3.2}$$

where R is the number of reference location, n_r is the number of collected RSS at the r-th location and $x_{r,d}(t)$ is the t-th collected RSS at the r-th location. The m_d denotes the sample mean of the d-th AP.

$$m_d = \frac{1}{R \cdot n_r} \sum_{r=1}^{R} \sum_{t=1}^{n_r} x_{r,d}(t) \tag{3.3}$$

The $S_{T,d}$ is the separation of all measured RSS in all locations. However, Eq. 3.2 can not show the relation between different locations clearly. In order to further analysis the $S_{T,d}$, we define the within-class scatter value $S_{W,d}$ and the between-class scatter value $S_{B,d}$.

$$S_{W,d} = \frac{1}{R \cdot n_r} \sum_{r=1}^{R} \sum_{t=1}^{n_r} (x_{r,d}(t) - m_{r,d})^2 \tag{3.4}$$

$$S_{B,d} = \frac{1}{R} \sum_{r=1}^{R} (m_{r,d} - m_d)^2 \tag{3.5}$$

where, $m_{r,d}$ represents the sample mean of d-th AP at the r-th location.

$$m_{r,d} = \frac{1}{n_r} \sum_{t=1}^{n_r} x_{r,d}(t) \tag{3.6}$$

The within-class scatter value represents the RSS separation of a fixed location and it can be regarded as the noise of this environment. The between-class scatter value represents the separetion of $m_{r,d}$ at all locations and it can be regarded as the contained signal information. Then the total scatter value $S_{T,d}$ follows that

$$\begin{aligned}
S_{T,d} &= \frac{1}{R \cdot n_r} \sum_{r=1}^{R} \sum_{t=1}^{n_r} (x_{r,d}(t) - m_{r,d} + m_{r,d} - m_d)^2 && (3.7) \\
&= \frac{1}{R \cdot n_r} \sum_{r=1}^{R} \sum_{t=1}^{n_r} (x_{r,d}(t) - m_{r,d})^2 + \frac{1}{R \cdot n_r} \sum_{r=1}^{R} \sum_{t=1}^{n_r} (m_{r,d} - m_d)^2 && (3.8) \\
&= S_{W,d} + \frac{1}{R} \sum_{r=1}^{R} (m_{r,d} - m_d)^2 && (3.9) \\
&= S_{W,d} + S_{B,d} && (3.10) \\
&= noise + signal && (3.11)
\end{aligned}$$

Eq. 3.7 means that the total scatter is the sum of the within-class scatter and the between-class scatter. We select the important AP which has the maximum signal to noise ratio. Thus, we define the criterion function at d-th AP

$$J_d = \frac{signal}{noise} = \frac{S_{B,d}}{S_{W,d}} \tag{3.12}$$

For all APs, the measured RSSs at different locations are collected at same time. So we assume that the noise at all locations are the same. That means the within-class scatter value $S_{W,d}$ for all locations are the same. Thus, we treat the between-class scatter value $S_{B,d}$ as the quantified importance of d-th AP.

Let η_d denote the quantified importance for the d-th AP. The larger η_d is, the more important for d-th AP is. The importance of d-th AP should be capable to distinguish the character of locations distinctly in the signal space. Therefore, the variance of the d-th AP can be a quantified metric to characterize the relevance of location prediction since it explicitly shows the separation of RSS over the whole localization area. Thus, we define η_d, which represents the contained information of the d-th AP.

$$\eta_d = S_{B,d} = \frac{1}{R}\sum_{r=1}^{R}(m_{r,d} - m_d)^2 \tag{3.13}$$

Eq. 3.13 calculates the separation of each location in the d-th AP, which indicates how fast the power loss increases with the distance in the target environment. If η_d is very small, RSS is hardly used to extract the location imformation because the signal strength does not change with varying distances. On the other hand, RSS changing at different locations is evident with a large value of η_d. That explains why the separation of RSS over the whole localization area can be used to indecate the amount of information to estimate the location. The large variance indecates the more importance because the greater variability of the RSS is observed over the target environment. We select D importance APs by Eq. 3.13, which ranks APs in descending order of the value of η_d.

Next, we give an experiment in real environment to verify our proposal. Fig. 3.1 shows the fifth floor of BL building in NTU and the triangles represent the measured locations. We measure RSSs in this area by a laptop with Windows XP operating system and NetStumbler network software. We collect 100 samples per location at different time periods. Our measurements show that 15 APs are stable in this floor. Then, we use Eq. 3.13 to calculate the importance of each AP. AP_1 AP_2 AP_3 denote the best 3 APs and AP_4 AP_5 AP_6 denote the worst 3 APs.

Fig. 3.2(a) and Fig. 3.2(b) show the RSS measurement space for various locations in the fifth floor of BL building. These figure clearly present the asymmetric contribution of the best 3 APs (AP_1, AP_2, AP_3) and the worst 3 APs

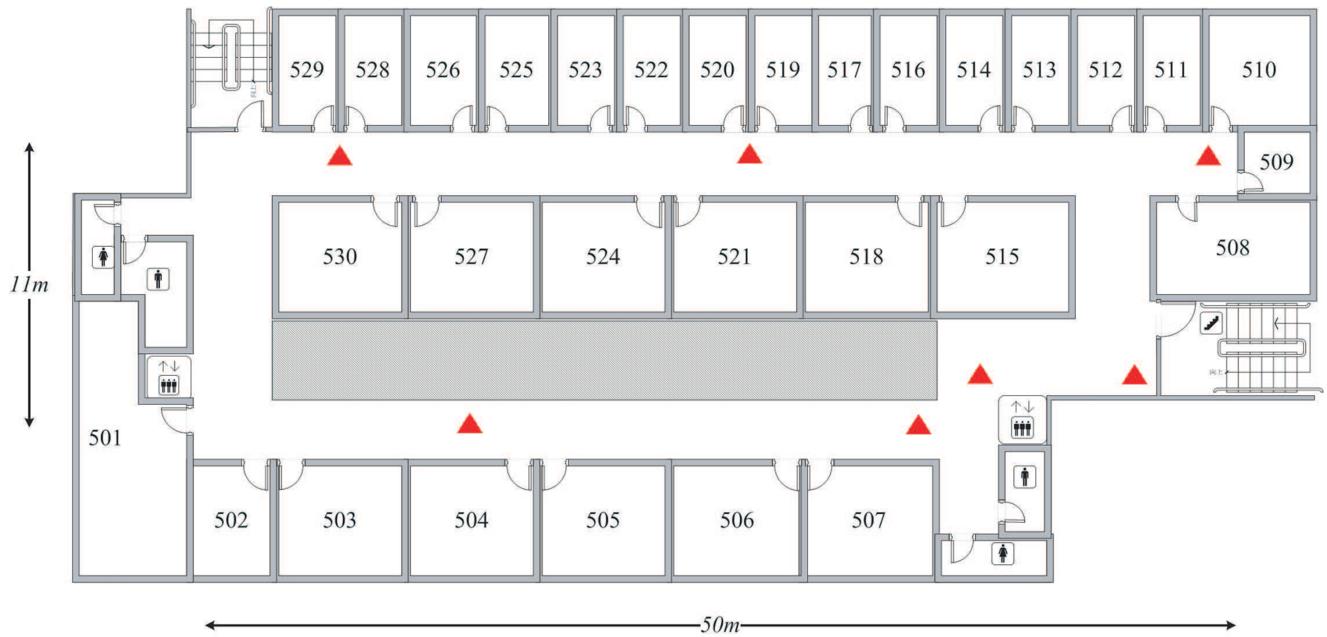
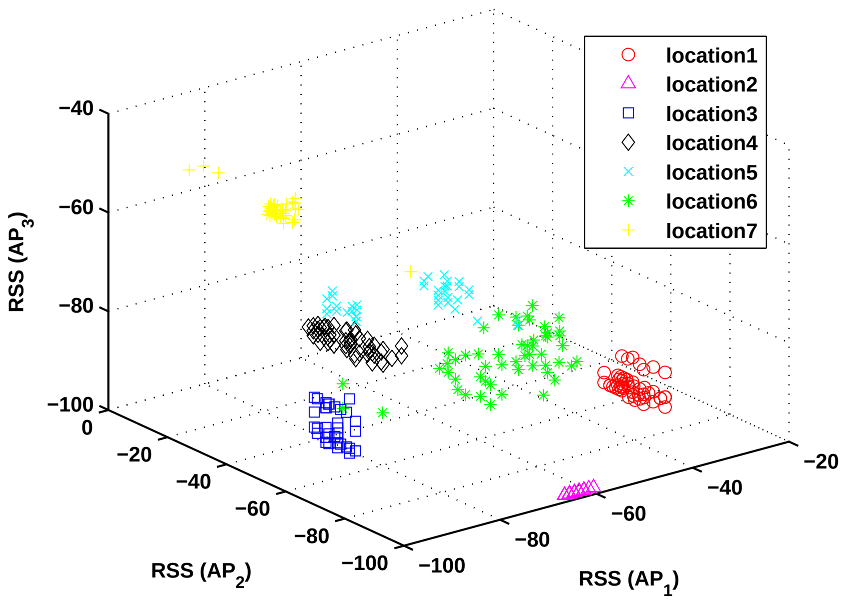
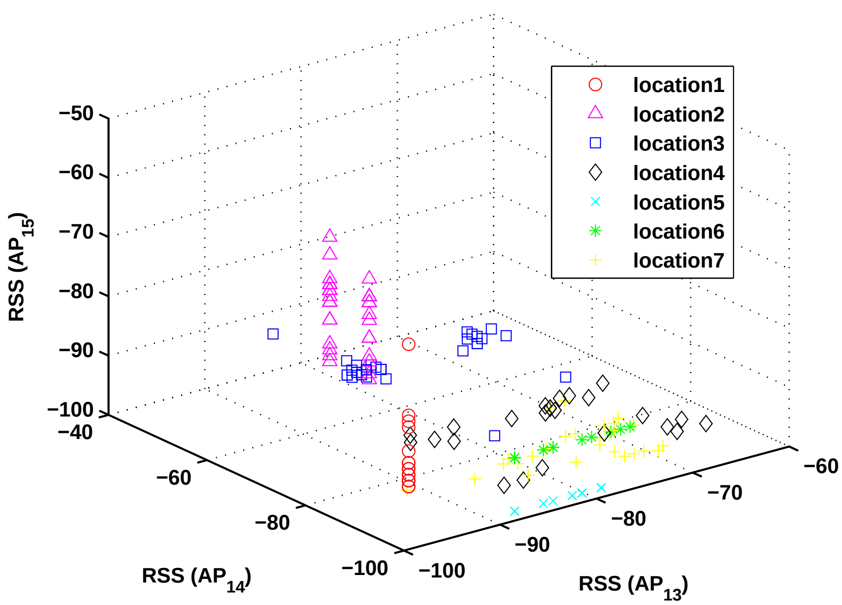


Figure 3.1: BL building.

(AP_{13}, AP_{14}, AP_{15}). As can be seen, the best 3 APs are more important than the worst 3 APs because the best 3 APs present beter RSS discrimination for the changing distance in this indoor environment.

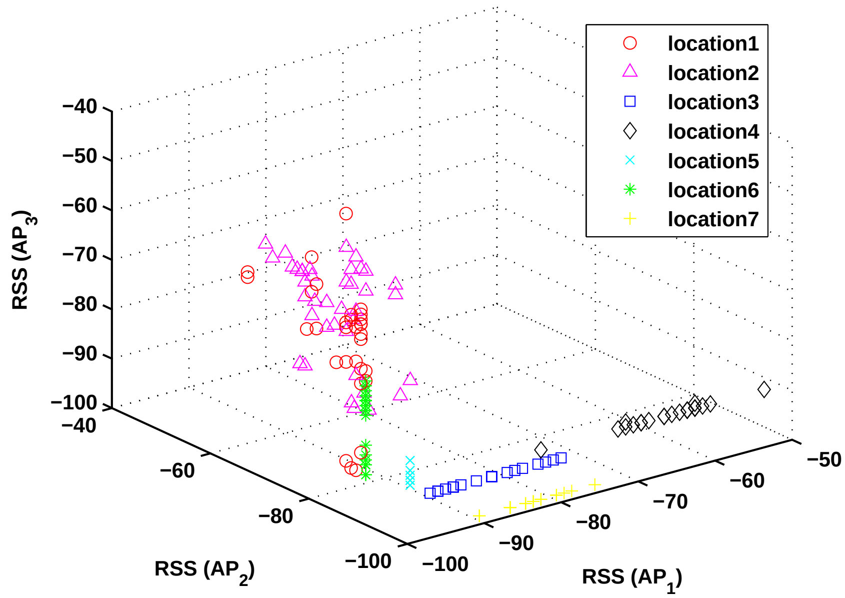


(a) The best 3 APs

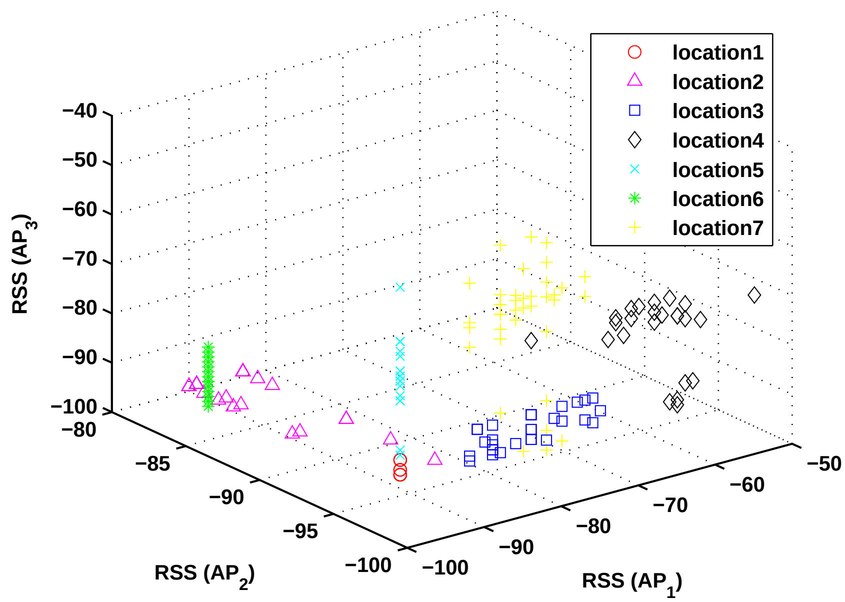


(b) The worst 3 APs

Figure 3.2: The RSS measurement space for various locations in the BL building.



(a) MaxMean



(b) InfoGain

Figure 3.3: The RSS measurement space for various locations by using (a) MaxMean and (b) InfoGain methods to select the most important three APs.

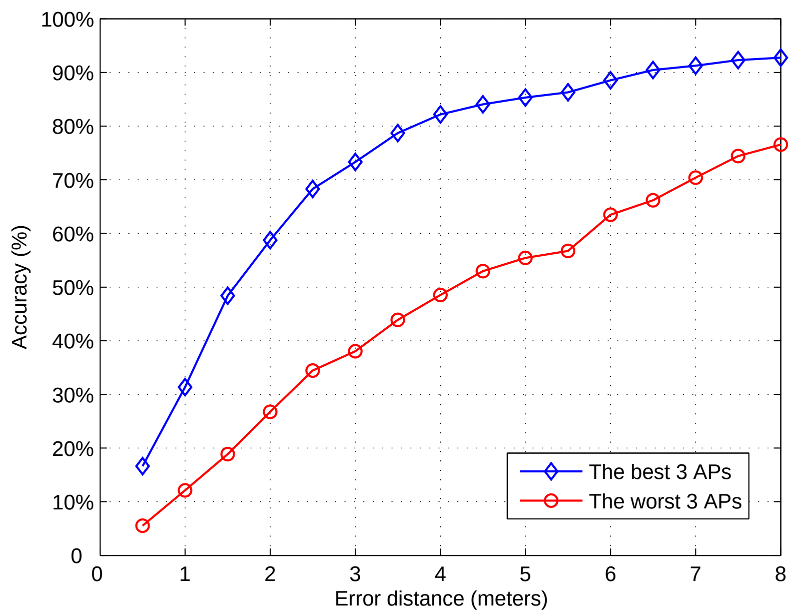


Figure 3.4: The accuracy of the best APs versus the worst APs.

Chapter 4

Proposed positioning algorithm

4.1 Overview of proposed positioning system

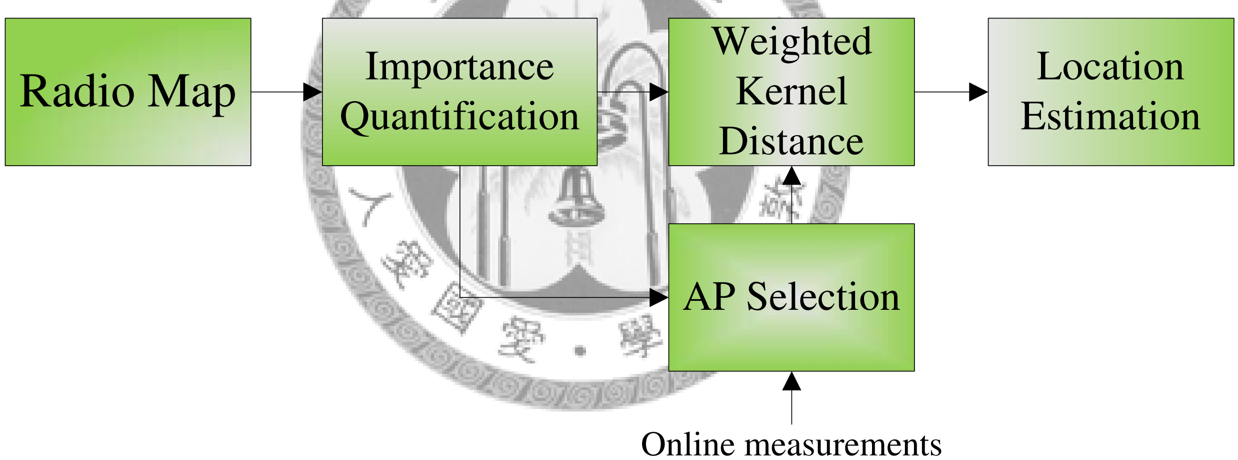


Figure 4.1: Overview of proposed positioning system.

The proposed positioning system contains some stages, as shown in Fig. 4.1. In the first stage, a site survey performed in the target environment is required to collect the RSS. The RSS are collected at sampling locations to build the radio map. A radio map thus provides a model of RSS in a development area. We usually collect a sequence of RSS, each sequence contains many samples per location to observe its temporal variation. The more reference locations means the higher density in the radio map at the expense of more collecting effort. After constructing the radio map, a WLAN client's is estimated by inspecting currently measured RSS.

The importance quantification stage is a method to quantify the importance of each AP. The importance is a value which is quantified by the signal discrimination between distinct locations. Depending on these importance, we select the more important D APs from the radio map in the AP selection stage. Furthermore, using such importance value, we can calculate the discriminative gain through a quasi entropy function. The discriminative gain is a weight of a important AP which is added into the kernel distance function in the weighted kernel distance stage. Finally, we use the mentioned kernel weight to calculate the final estimation result. We clearly introduce our positioning algorithm in the next section.

4.2 Incorporating the Importance into Location Estimation

This section illustrates how the different importance is incorporated into the location estimation. If the amount of information of each AP is quantified, we can gracefully incorporate such physical property into the calculation of positioning algorithm. Since the location estimation is performed by a kernel function, the different importance should be embedded into the kernel distance function as

$$\widehat{K}(X, X_r(t)) = \exp\{\sum_{d=1}^{D} \frac{-w_d^2}{2\widetilde{\sigma}_{r,d}^2} \cdot [x_d - x_{r,d}(t)]^2\} \qquad (4.1)$$

where w_d represent the weight at d-th AP, $\widetilde{\sigma}_{r,d}^2$ is an adjustable kernel width, and D represents the number of selected APs. In Eq. 4.1 we incorporate a novel parameter w_d, which represents the unequal contribution of each AP. Therefore, Eq. 4.1 can be regarded as a weighted kernel function where the effect of APs is differentiated. In other words, we assign larger weights to the relevant APs and smaller weights to the irrelevant APs. This way, RSS from different APs are fused with different importance and the location computation can be dominated by the more relevant ones. Afterwards, the location can be estimated by

$$\hat{l} = \sum_{r=1}^{R} l_r \left(\frac{1}{n_r} \sum_{t=1}^{n_r} K(X, X_r(t)) \right) \tag{4.2}$$

From Eq. 4.1 and Eq. 4.2 the comtribution from each distance member $X - X - r(t)$ is fused with different weights w_d in the transformed kernel to estimate the user's location $\hat{l}$. The higher the weight is, the bigger belief we give to this component which dominates the computation. To our knowledge, such physical property has not been exploited in designing a location system. When the weights are all equal ($w_d = 1, d = 1 \cdots D$), this method is reduced to the traditional kernel positioning.

4.3 Weights determination

This section investigates how to determine the weights required in Eq. 4.2. One practical problem we discover during the experiments is that there exists large differences of η_d between different signals. That is, some η_d may present several hundred times larger than the other. To provide a graceful quantitative metric in calculating the spatial likelihood of the measured signals, we utilize a quasi entropy function $f(\cdot)$ to determine the weight w_d as follows:

$$\begin{aligned} w_d &= \alpha + f(\eta_d) \\ &= \alpha + \frac{-(1-\eta_d^*)\log(1-\eta_d^*)}{\beta} \end{aligned} \tag{4.3}$$

where β is the maximum value of the numerator to make the value of $f(\eta_d)$ smaller than 1 ($\beta = \max(-(1-\eta_d^*)\log(1-\eta_d^*))$, $d = 1, 2 \cdots D$), and η_d^* is the normalized value of η_d as

$$\eta_d^* = \eta_d / \sum_{d=1}^{D} \eta_d \tag{4.4}$$

In Eq. 4.3, $1-\eta_d^*$ can be viewed as a numerical value of probability and $f(\cdot)$ is smaller with the definition of entropy function. It can be observed that $f(\eta_d)$ increases with η_d and ranges between 0 and 1 ($0 \leq f(\eta_d) \leq 1$). That means that the changing scale of weights is constrained in a resonable range. The parameter α is a constant which controls the bias gain. This value is adjusted to make the minimum gain larger than α. Thus, we can control the weights as $\alpha \leq w_d \leq \alpha + 1$ at the same time to avoid an abrupt change on the weights.

Chapter 5

Experimental Setup And Results

5.1 Experimental Setup

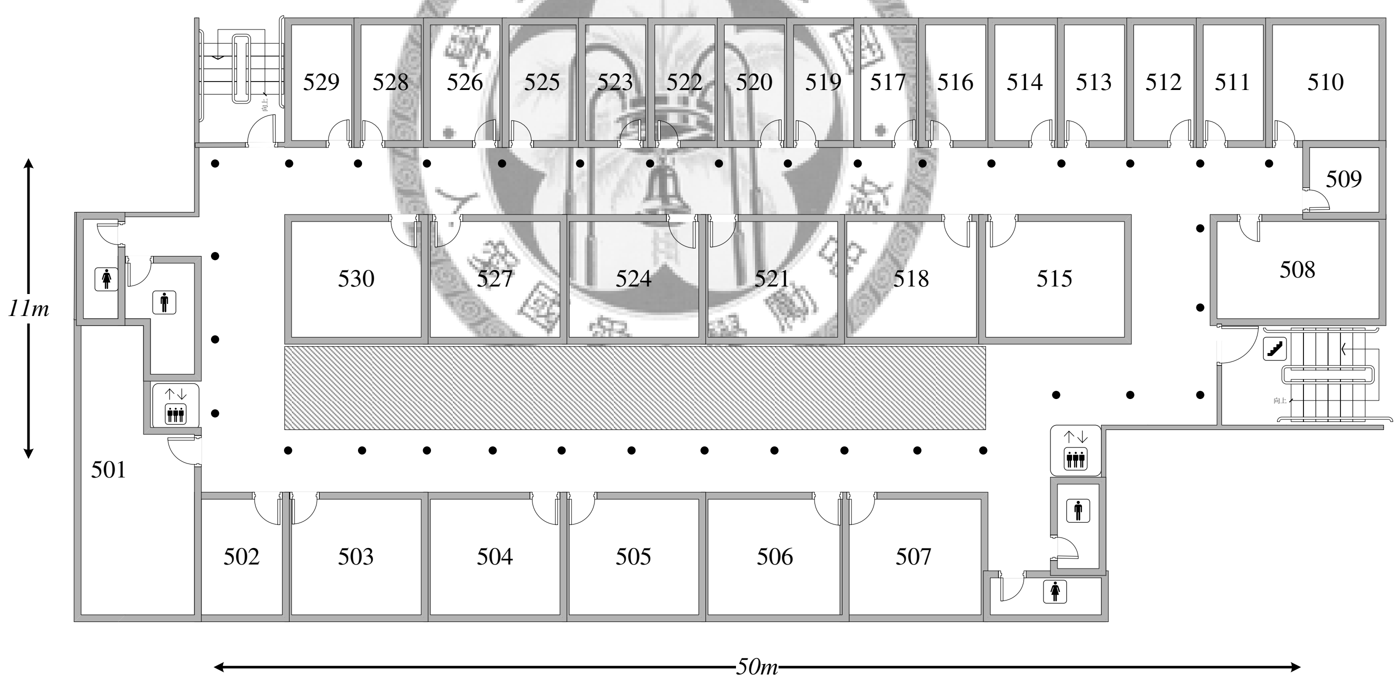


Figure 5.1: The fifth floor plane of the BL building, where we had performed the experiments. The dots represent the reference locations.

The proposed algorithm is evaluated on a realistic indoor environments. The measurements are collected on the fifth floor of BL building in NTU, as shown in 5.1. We collect WLAN data in this area by a laptop with Windows XP operating system and NetStumbler network software. The dimensions of this test-bed are 52 meters times 18 meters. 35 reference locations are selected with a 3.2 meters space (R=35). We collect 50 samples per location at different time periods for training data and testing data, respectively ($n_r = 50$). Our measurements show that over 30 APs can be detected in this floor and 15 APs are stable. We utilize MaxMean and InfoGain criterion to select the important APs from the 15 stable APs for comparison. The kernel width is set a constant and the bias weight α is 1 in the experiments. Finally, the positioning error is defined as the Euclidean distance between the estimated result and the true coordinate.

5.2 Experimental results

5.2.1 AP selection

First, we compare the effects of AP selection methods. The performance is evaluated in terms of the positioning accuracy, which is defined as the cumulative percentage of estimations within specified errors.

Fig. 5.2 shows the accuracy comparison between MaxMean, InfoGain and the proposed method. This figure clearly shows that our approach outperforms the traditional methods under the same AP numbers. Using 5 APs, our approach exceeds 70% while those of MaxMean and InfoGain both are worse than 50%. That means that our approach has the advantage of using the fewest APs to achieve the same level accuracy.

Fig. 5.3 reports the mean and standard deviation (std) of error on the number

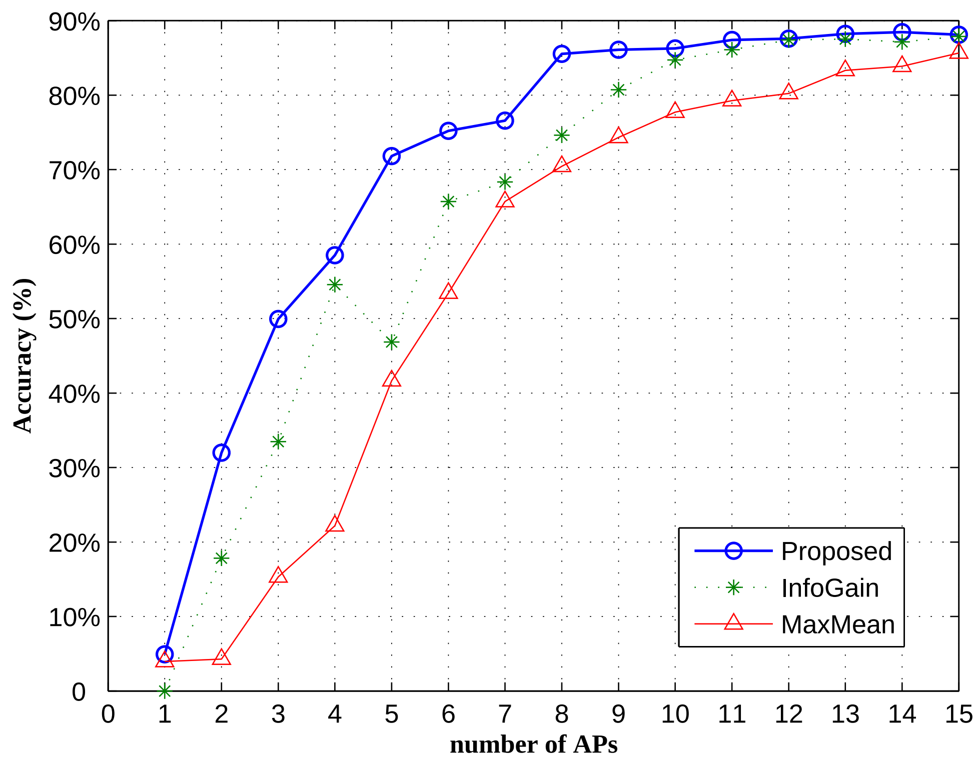
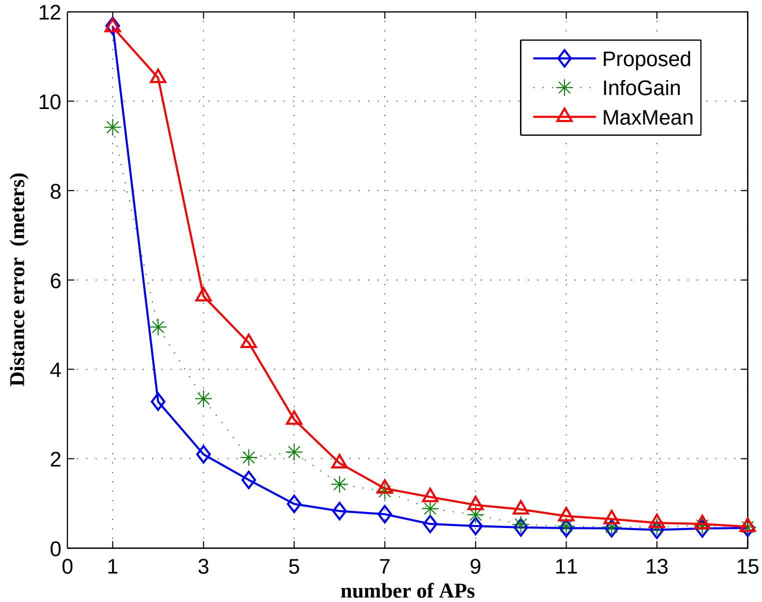
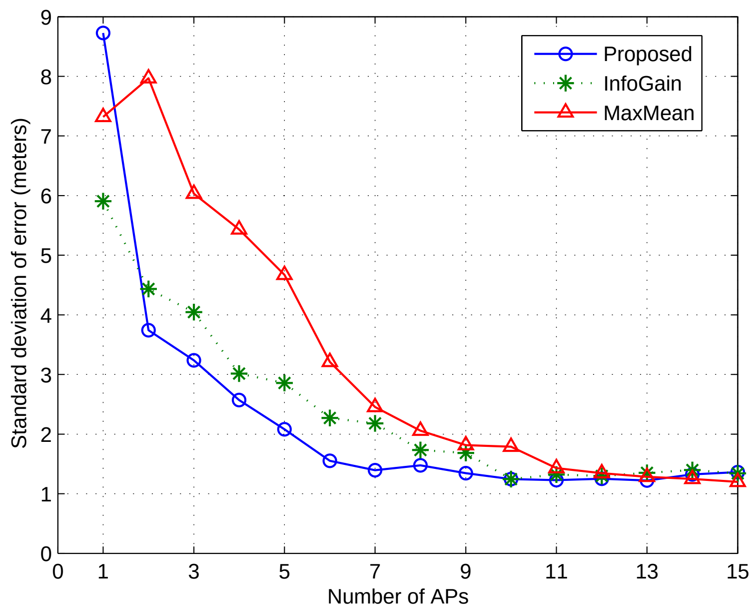


Figure 5.2: The accuracy of error distance within 1 meter versus number of APs.



(a)



(b)

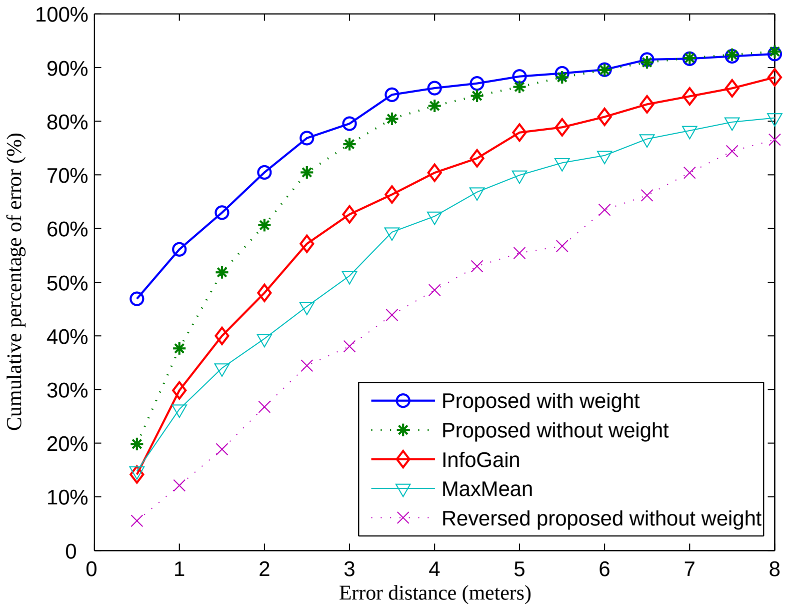
Figure 5.3: (a) Mean and (b) Standard Deviation of the estimated error versus number of APs.

of involved APs. Fig. 5.3(a) clearly shows that the error distance of our method is beter than other methods. Also, the error distance in five APs are similar to more than five APs. Thus, selecting more than five APs can increase the computational complexity but the decreasing error distance is limited. As seen from Fig. 5.3(b), the increase in number of APs reduce the std of error because more location-related information is utilized. More importantly, Fig. 5.3(b) again shows that our approach outperform traditional approaches. The decrease of error of our method is larger than that of the existing methods.

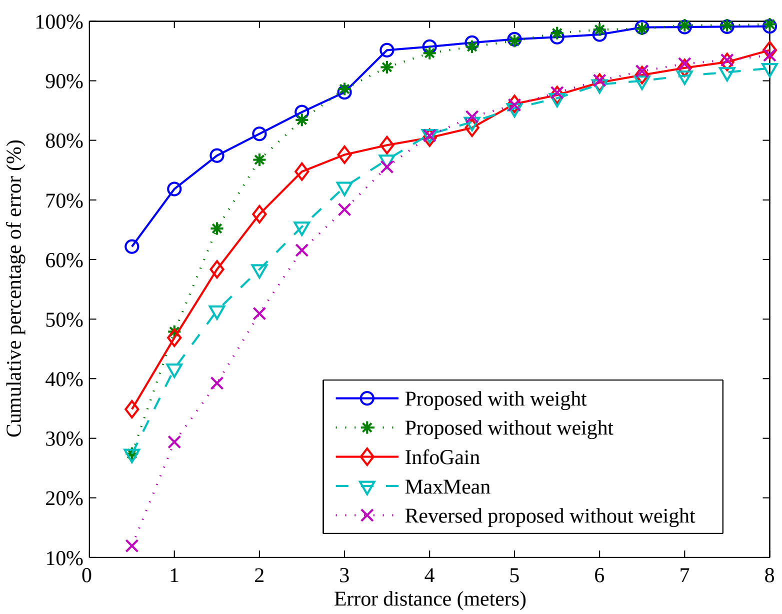
5.2.2 Performance evaluation by incorporating the importance of APs

In this subsection, we evaluate the performance after incorporating the different importance of APs. Fig. 5.4(a) and Fig. 5.4(b) shows the accuracy versus error distance under 3APs and 5APs, respectively. In these figures, the reversed proposed method ranks APs in reverse order of the way the proposed algorithm does. These figures clearly indicate superior performance based on our positioning algorithm, as compared to existing methods. After incorporating the different weights, the accuracy is further greatly improved. In the Fig. 5.4(a), the accuracy within two meters is improved from 60% to 70% if each kernel distance is fused with different importance. Also, the accuracy within one meter is improved from 48% to 71% by incorporating the weight to the kernel distance in Fig. 5.4(b). On the contrary, the reversed proposed method performs the worst.

Fig. 5.5(a) and Fig. 5.5(b) reports the mean and std of error while the embedded weights are determined from the 3APs and 5APs by proposed method, respectively. It is interesting from Fig. 5.5 that the accuracy is not improved if

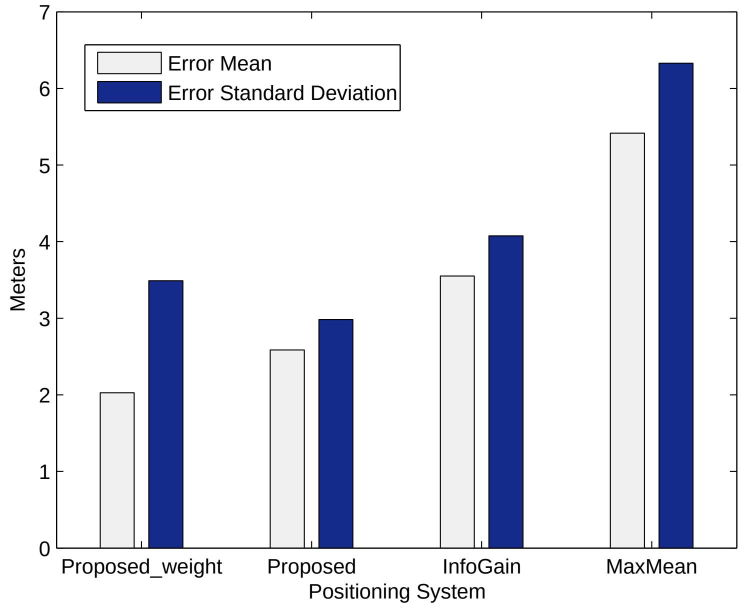


(a) 3APs

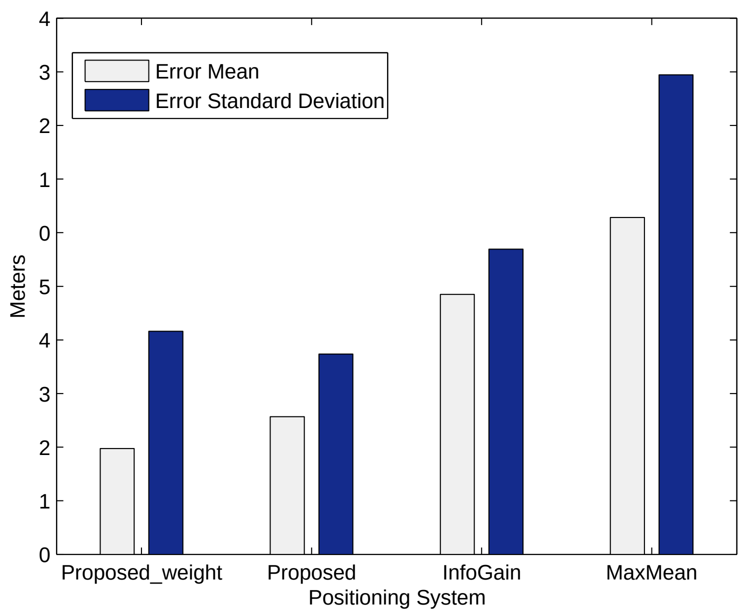


(b) 5APs

Figure 5.4: The cumulative percentage of error for different AP selection methods and the effect of weights incorporation.



(a) 3APs



(b) 5APs

Figure 5.5: Mean and standard deviation of error while the embedded weights are determined from the selected APs by proposed method.

the APs are selected by MaxMean. This is because MaxMean is not a suitable measurement of AP relevancies. Once the APs are not carefully selected, the weightings may not produce a better result. Fig. 5.5 indicates that the performance of InfoGain slightly improves and only the proposed method achieves a 2 m error mean.

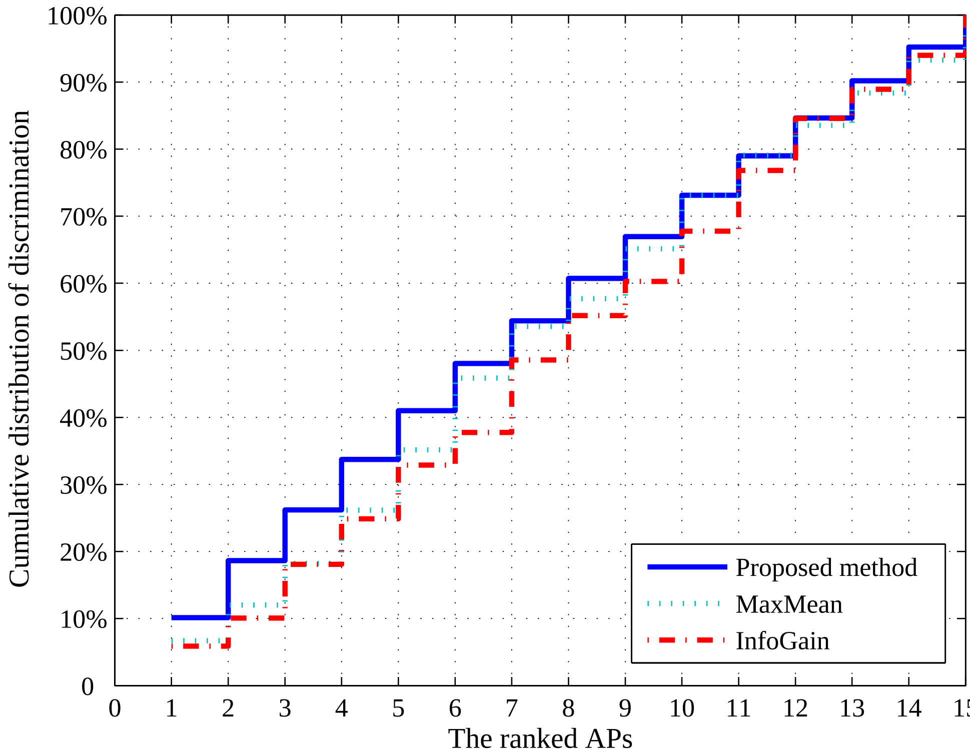


Figure 5.6: The cumulative percentage of η_d obtained from MaxMean, InfoGain and the proposed method.

Finally, Fig. 5.6 shows the cumulative percentage of η_d obtained from MaxMean, InfoGain and the proposed method. From Fig. 5.6, the slope in our method is steeper than the other methods. This observation agrees fairly well with the results in Fig. 5.5. That means that η_d is a better indication of the quantified importance.

5.3 Test in different points

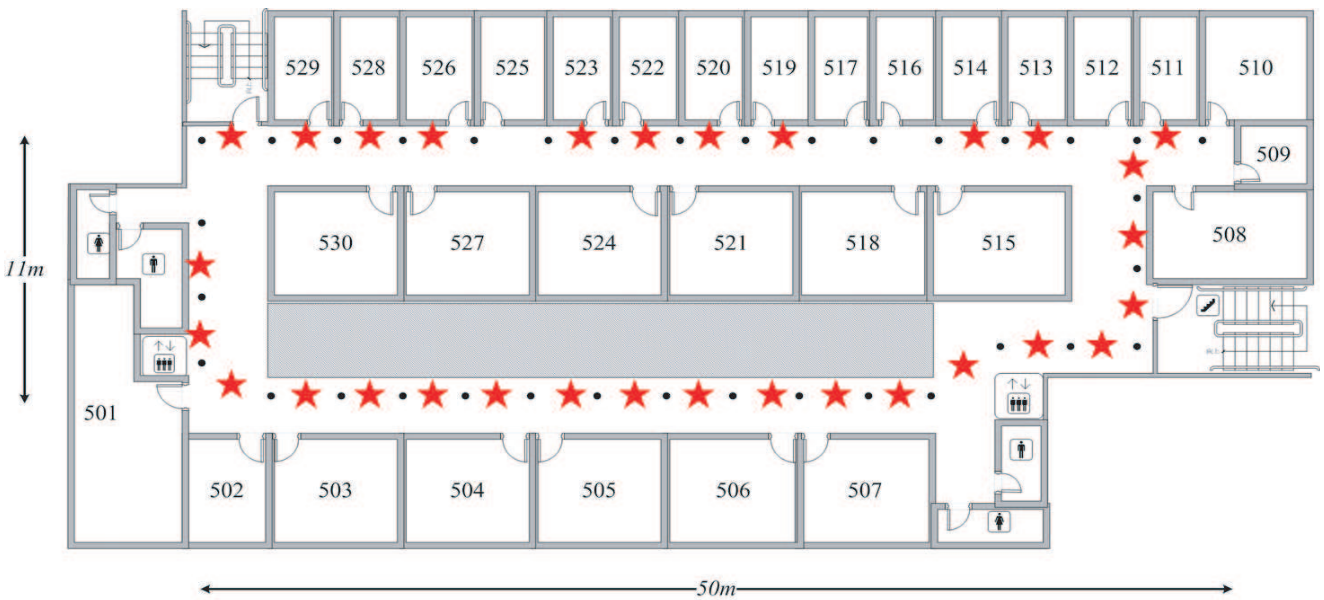


Figure 5.7: The fifth floor plane of the BL building, where we had performed the experiments. The red stars represent the testing locations.

In this section, we measure the additional testing points which are different from training points, as shown in Fig. 5.7. In Fig. 5.7, the red stars represent the testing points and the dots represent the original training points. We follow the same procedure of Section 5.1 to measure the RSS in this area. We collect 100 samples per training point at different time periods for training data. The testing samples are collected at 30 testing points. The testing points are selected between the training points. A total of 50 samples are collected per testing point at a rate of one sample/sec. We also compare proposed method with MaxMean and InfoGain criterion to select the important APs from the 15 stable APs. The kernel width is set a constant and the bias weight α is 1 in the experiments. Finally, the positioning error is defined as the Euclidean distance between the estimated result and the true coordinate.

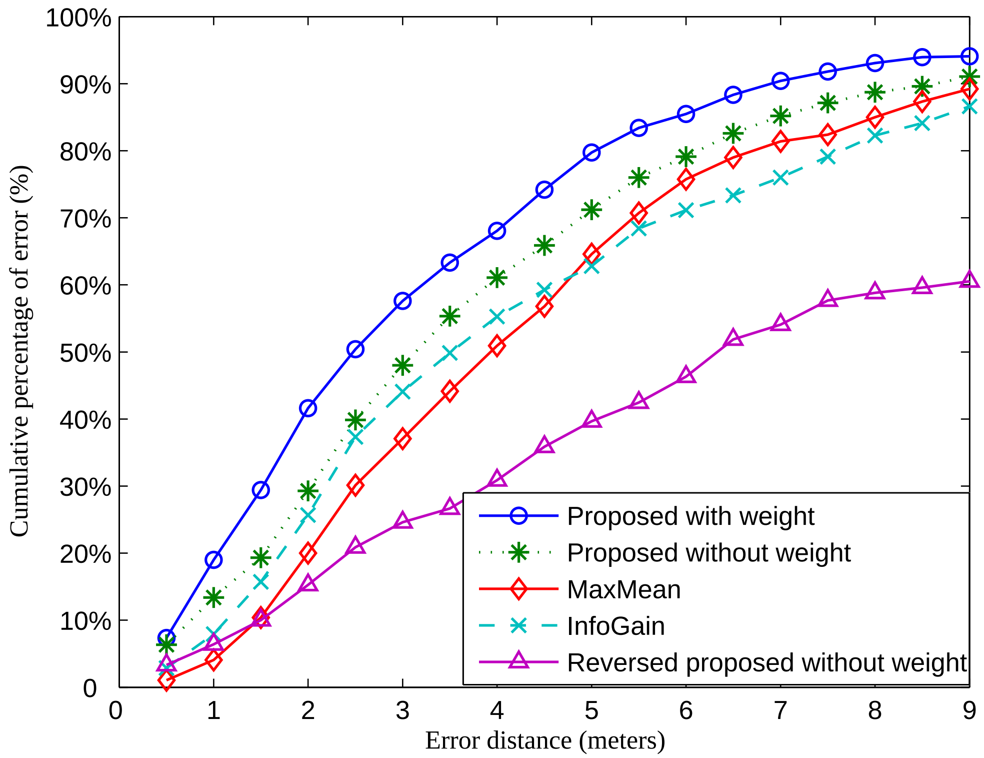


Figure 5.8: The cumulative percentage of error for different AP selection methods and the effect of weights incorporation.

In this subsection, we evaluate the performance of our proposed system and also compare the AP selection method with MaxMean and InfoGain by the different testing points. Because all the training points and testing points are different, the positioning systems need more information to improve the positioning accuracy. Thus, selecting the more important APs may increase the accuracy clearly. The performance is evaluated in terms of the positioning accuracy, which is de-

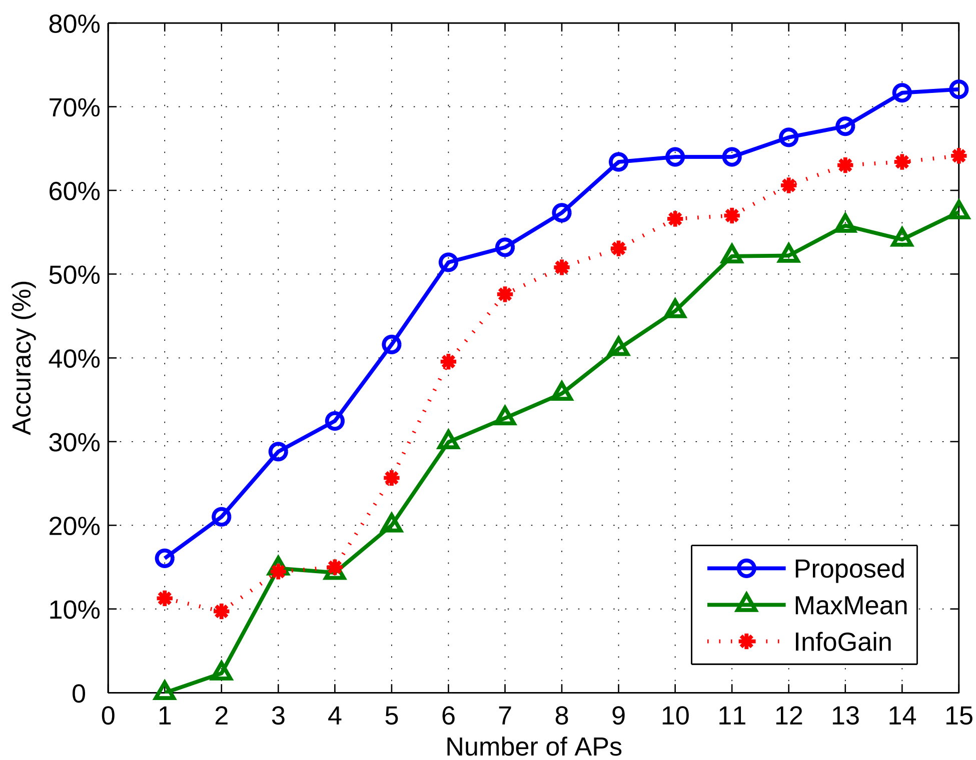


Figure 5.9: The accuracy of error distance within two meters versus number of APs.

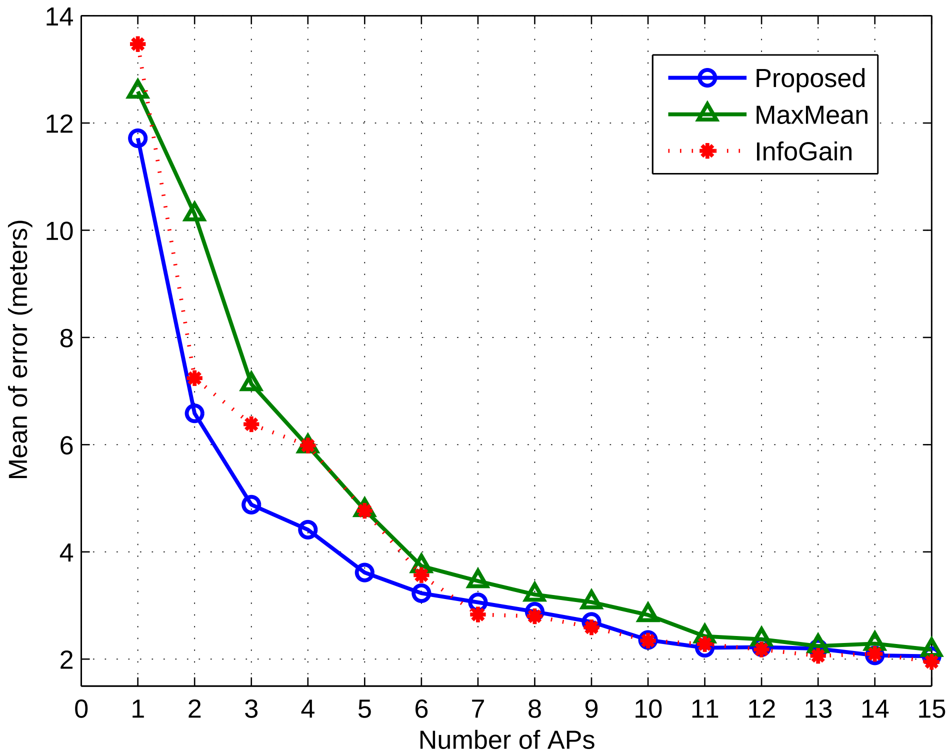


Figure 5.10: Mean of the estimated error versus number of APs.

fined as the cumulative percentage of estimations within specified errors. First, we compare our proposed AP selection method with MaxMean, InfoGain, and reversed proposed method, as shown in Fig. 5.8. In Fig. 5.8, the accuracy of error distance within two meters in proposed method is 48% while those of MaxMean, InfoGain, reversed proposed method are 30%, 43%, and 20% respectively. Fig. 5.9 shows the accuracy of error distance within two meters versus number of APs. This figure clearly shows that our approach outperforms the traditional methods under the same AP numbers. Using 6 APs, our approach exceeds 50% while those of MaxMean and InfoGain both are worse than 40%. Furthermore, Fig. 5.10 shows mean of the estimated error versus number of APs. That means that our approach has the advantage of using the fewest APs to achieve the same level accuracy.

Moreover, the improvement of incorporating the kernel weight is shown in Fig. 5.8. The accuracy within three meters is improved from 48% to 58% if each kernel distance is fused with different importance. More result is presented in Fig. 5.11. As seem to Fig. 5.11, the proposed method has more robustness in different number of APs. In this figure, our proposed system can further improve the error distance, as compared to traditional MaxMean and InfoGain.

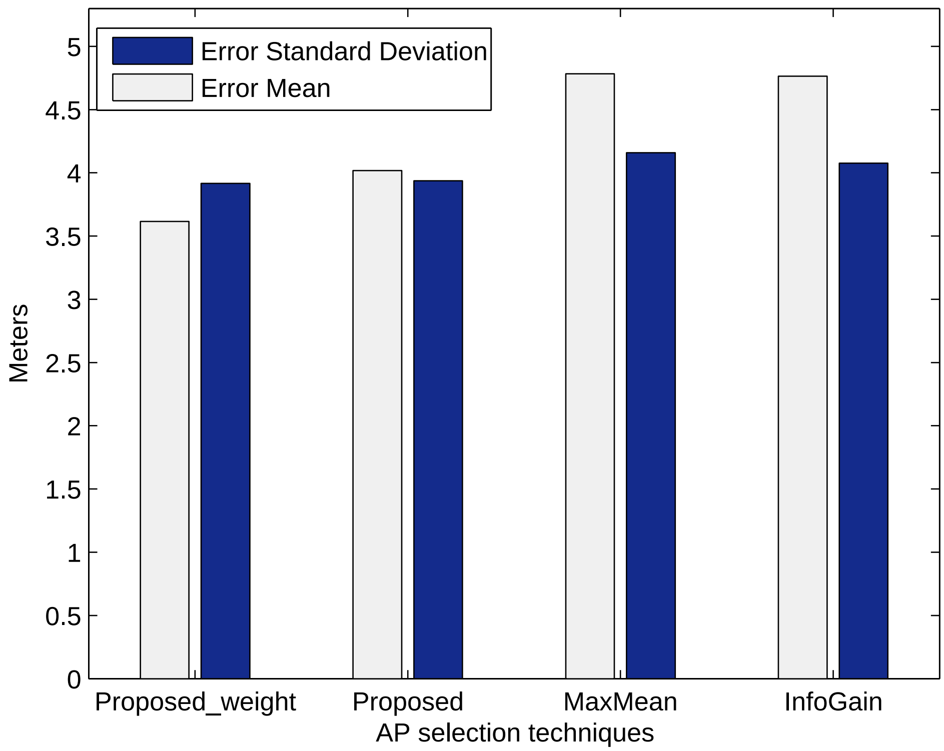


Figure 5.11: Mean of error while the embedded weights are determined from the selected APs by proposed method.

Chapter 6

Applications

The main contribution of this article is to provide an efficient method to quantify the importance of APs. Then, using these quantified values to select the most important APs for positioning. The positioning accuracy is further improved by using such method to select important APs. However, such technique dose not know the AP's location. A significant question is where are the important APs. Our method can solve not only an AP selection problem but also an AP placement problem. That is, AP selection is an efficient application in AP placement. The problem of AP placement is how to find the important AP's position. Clearly, the values of η_d are efficient measurement of this problem. Recalling to Sec. 3.2, the η_d are defined by

$$\eta_d = \frac{1}{R}\sum_{r=1}^{R}(m_{r,d} - m_d)^2 \tag{6.1}$$

,where the Eq. 6.1 calculates the signal separation of the d-th AP. We denote the d-th AP's position by l_d. Our method assumes that we know all AP's positions. Then, we calculate η_d for all APs and select the larger η_d for positioning AP. Thus, the selected AP's positions can be obtain by l_d.

For example, if we want to construct a positioning system in a WLAN environment and set five APs for positioning AP. We can set some steps of this example. First, we place many APs around this environment and we know all AP's positions. Second, we measure the RSS in this area and build the radio map. Third, we calculate the η_d of each AP and select the largest five η_d for positioning AP. Then, the corresponding AP's positions can obtain by selected η_d.

This application of AP selection is an efficient method for constructing a positioning system. In a WLAN environment, the only one different of each AP is position so to select the important AP means place the AP in the discriminative positions. The results of this application can refer to Ch. 5.

Chapter 7

Conclusions

The main finding of this work is to investigate the different contribution of each AP to the location estimation in WLANs. We argue that the importance should be embedded in the positioning algorithm. However, traditional approaches treat each RSS in an equal way. This article shows that it would be more profitable to take such properties into consideration while designing a location system. The main contribution is two parts. First, a novel mechanism is proposed to measure the degrees of the AP relevancies. The importance of each AP is quantified by the signal discrimination between distinct locations. We utilize such numerical importance to select important APs to avoid unnecessary calculations. Second, the importance is further embedded into our positioning system. We provide a weighted kernel function where the effect of APs is differentiated. That is, the larger weights are assigned to the more important APs. Moreover, we develop a quasientropy function to avoid an abrupt change on the weights. Our positioning system is developed in a real-world WLAN environment, where the realistic measurement of receive signal strength (RSS) is collected. Experimental results show that the positioning accuracy is significantly improved by taking the different importance into consideration.

Bibliography

[1] J. Hightower and G. Borriello, "Location systems for ubiquitous computing," *IEEE Computer Magazine*, vol. 34, no. 8, pp. 57–66, 2001. 1

[2] T. S. Rappaport, J. H. Reed, and D. Woerner, "Position location using wireless communications on highways of the future," *IEEE Communications Magazine*, vol. 34, no. 10, pp. 33–41, 1996. 1

[3] S. Tekinay, "Wireless geolocation systems and services," *IEEE Communications Magazine*, vol. 36, no. 4, pp. 28–28, 1998. 1

[4] K. Mikkel, Baun, "A taxonomy for radio location fingerprinting," *Lecture Notes in Computer Science*, vol. 4718, pp. 139–156, 2007. 1, 7

[5] S. Gezici, "A survey on wireless position estimation," *Wireless Personal Communications*, vol. 44, no. 3, pp. 263–282, 2008. 1

[6] P. Bahl and V. N. Padmanabhan, "RADAR: An in-building RF-based user location and tracking system," in *INFOCOM*, 2000, pp. 775–784. 1, 7

[7] F. Gustafsson and F. Gunnarsson, "Mobile positioning using wireless networks: possibilities and fundamental limitations based on available wireless network measurements," *IEEE Signal Processing Magazine*, vol. 22, no. 4, pp. 41–53, 2005. 1

[8] A. Kushki, K. Plataniotis, and A. Venetsanopoulos, "Sensor selection for mitigation of rss-based attacks in wireless local area network positioning," 31 2008-April 4 2008, pp. 2065–2068. 2

[9] M. Zhang, S. Zhang, and J. Cao, "Fusing received signal strength from multiple access points for WLAN user location estimation," *Internet Computing in Science and Engineering*, pp. 173–180, 2008. 2

[10] J. Kwon, B. Dundar, and P. Varaiya, "Hybrid algorithm for indoor positioning using wireless LAN," *Vehicular Technology Conference*, vol. 7, pp. 4625–4629, 2004. 2

[11] K. W. Kolodziej and J. Hjelm, *Local Positioning Systems: LBS Applications and Services*. CRC Taylor & Francis, 2006. 7

[12] X. Li and K. Pahlavan, "Super-resolution TOA estimation with diversity for indoor geolocation," *IEEE Transactions on Wireless Communications*, vol. 3, no. 1, pp. 224–234, 2004. 7

[13] K. Pahlavan, X. Li, and J. Makela, "Indoor geolocation science and technology," *IEEE Communications Magazine*, vol. 40, no. 2, pp. 112–118, 2002. 7

[14] G. Sun, J. Chen, W. Guo, and K. Liu, "Signal processing techniques in network-aided positioning: a survey of state-of-the-art positioning designs," *IEEE Signal Processing Magazine*, vol. 22, no. 4, pp. 12–23, 2005. 7

[15] A. Ankur, K. Parakram, and S. Huzur, "LOCATOR: location estimation system for wireless LANs," *Wireless Mobile Applications And Services On WLAN Hotspots*, pp. 102 – 109, 2004. 7

[16] P. Castro and R. Munz, "Managing context data for smart spaces," *Personal Communications, IEEE [see also IEEE Wireless Communications]*, vol. 7, no. 5, pp. 44–46, 2000. 9

[17] T. Roos, P. Myllymaki, H. Tirri, P. Misikangas, and J. Sievanen, "A probabilistic approach to WLAN user location estimation," *Wireless Information Networks*, vol. 9, no. 3, pp. 155–164, 2002. 9

[18] A. Kushki, N. Plataniotis, Konstantinos, and N. Venetsanopoulos, Anastasios, "Kernel-based positioning in wireless local area networks," *IEEE Transactions on Mobile Computing*, vol. 6, no. 6, pp. 689–705, 2007. 9, 14, 15

[19] Z. li Wu, C. hung Li, J.-Y. Ng, and K. R. Leung, "Location estimation via support vector regression," *IEEE Transactions on Mobile Computing*, vol. 6, no. 3, pp. 311–321, 2007. 9

[20] P. Krishnan, A. Krishnakumar, W.-H. Ju, C. Mallows, and S. Gamt, "A system for LEASE: Location estimation assisted by stationery emitters for indoor RF wireless networks," *INFOCOM*, vol. 2, pp. 1001–1011, 2004.

[21] R. Battiti, T. L. Nhat, and A. Villani, "Location-aware computing: a neural network model for determining location in wireless LANs," Technical Report DIT-02-0083, Department of Information and Communication Technology, University of Trento, Italy, Tech. Rep., 2002.

[22] A. M. Edgar, C. Raul, and F. Jesus, "Estimating user location in a WLAN using backpropagation neural networks," *Lecture Notes in Computer Science*, vol. 3315, pp. 737–746, 2004.

[23] C. Nerguizian, C. Despins, and S. Affes, "Geolocation in mines with an impulse response fingerprinting technique and neural networks," *IEEE Transactions on Wireless Communication*, vol. 5, no. 3, pp. 603–611, 2006.

[24] M. Brunato and R. Battiti, "Statistical learning theory for location fingerprinting in wireless LANs," *Computer Networks*, vol. 47, no. 6, pp. 825–845, 2005.

[25] S.-H. Fang and T.-N. Lin, "Indoor location system based on discriminant-adaptive neural network in ieee 802.11 environments," *IEEE Transactions on Knowledge and Data Engineering*, vol. 19, no. 11, pp. 1973–1978, 2008.

[26] C. Nerguizian, C. Despins, and S. Affes, "Geolocation in mines with an impulse response fingerprinting technique and neural networks," in *Vehicular Technology Conference*, 2004, pp. 3589–3594.

[27] T. King, S. Kopf, T. Haenselmann, C. Lubberger, and W. Effelsberg, "Compass: A probabilistic indoor positioning system based on 802.11 and digital compasses," in *Wireless Network Testbeds, Experimental evaluation and Characterization*, 2006.

[28] S. Golden and S. Bateman, “Sensor measurements for wi-fi location with emphasis on time-of-arrival ranging,” *IEEE Transactions on Mobile Computing*, vol. 6, no. 10, pp. 1185–1198, 2007.

[29] J. Yin, Q. Yang, and L. M. Ni, “Learning adaptive temporal radio maps for signal-strength-based location estimation,” *IEEE Transactions on Mobile Computing*, vol. 7, no. 7, pp. 869–883, July 2008. 14

[30] C. Patterson, R. Muntz, and C. Pancake, “Challenges in location-aware computing,” *Pervasive Computing, IEEE*, vol. 2, no. 2, pp. 80–89, 2003.

[31] S.-H. Fang, T.-N. Lin, and P.-C. Lin, “Location fingerprinting in a decorrelated space,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 5, pp. 685–691, 2008.

[32] V. Otsason, A. Varshavsky, A. LaMarca, and E. de Lara, “Accurate GSM indoor localization,” *Lecture Notes in Computer Science*, vol. 3660, pp. 141–158, 2005.

[33] X. Chai and Q. Yang, “Reducing the calibration effort for probabilistic indoor location estimation,” *IEEE Trans. on Mobile Computing*, vol. 6, no. 6, pp. 649–662, 2007.

[34] ——, “Reducing the calibration effort for location estimation using unlabeled samples,” *Pervasive Computing and Communications*, pp. 95–104, 2005.

[35] L. F. M. de Moraes and B. A. A. Nunes, “Calibration-free WLAN location system based on dynamic mapping of signal strength,” in *Mobility management and wireless access*, 2006, pp. 92–99.

[36] K. Mikkel, Baun, T. Georg, and C. Linnhoff-Popien, “Zone-based rss reporting for location fingerprinting,” *Lecture Notes in Computer Science*, vol. 4480, pp. 316–333, 2007.

[37] T. King, T. Haenselmann, and W. Effelsberg, “On-demand fingerprint selection for 802.11-based positioning systems,” in *a World of Wireless, Mobile and Multimedia Networks*, 2008.

[38] M. Youssef, A. Agrawala, and A. U. Shankar, "WLAN location determination via clustering and probability distributions," in *Pervasive Computing and Communications*, 2003, pp. 143–150. 14

[39] Y. Chen, J. Yin, X. Chai, and Q. Yang, "Power-efficient access-point selection for indoor location estimation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, no. 7, pp. 877–888, 2006. 14, 15

[40] A. Mahtab Hossain, H. N. Van, Y. Jin, and W.-S. Soh, "Indoor localization using multiple wireless technologies," *Mobile Adhoc and Sensor Systems*, pp. 1–8, 2007.

[41] T. Roos, P. Myllymäki, H. Tirri, P. Misikangas, and J. Sievänen, "A probabilistic approach to wlan user location estimation," *International Journal of Wireless Information Networks*, vol. 9, no. 3, pp. 155–164, July 2002. 13

[42] K. Kaemarungsi and P. Krishnamurthy, "Modeling of indoor positioning systems based on location fingerprinting," *INFOCOM*, vol. 2, pp. 1012–1022, 2004. 15

[43] A. J. Weiss, "On the accuracy of a cellular location system based on RSS measurements," *IEEE Transactions on Vehicular Technology*, vol. 52, no. 6, pp. 1508–1518, 2003.