國立交通大學

網路工程研究所

碩 士 論 文

透過無線網路基地台相互監測之適應性樣本比對定位

Using Adaptive Radio Maps for Pattern-Matching Localization via Inter-Beacon Measurement

研 究 生：許藍尹

指導教授：曾煜棋 教授

中 華 民 國 九 十 九 年 六 月

透過無線網路基地台相互監測之適應性樣本比對定位
Using Adaptive Radio Maps for Pattern-Matching Localization via Inter-Beacon Measurement

研 究 生：許藍尹　　Student：Lan-Yin Hsu

指導教授：曾煜棋 教授　　Advisor：Prof. Yu-Chee Tseng

國 立 交 通 大 學
網 路 工 程 研 究 所
碩 士 論 文

A Thesis

Submitted to Institute of Network Engineering

College of Computer Science

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Master

in

Computer Science

June 2010

Hsinchu, Taiwan, Republic of China

中華民國九十九年六月

Using Adaptive Radio Maps for Pattern-Matching Localization via Inter-Beacon Measurement

Student: **Lan-Yin Hsu**

Advisor: **Prof. Yu-Chee Tseng**

Department of Computer Science

National Chiao-Tung University

1001 Ta Hsueh Road, Hsinchu, Taiwan 300, R.O.C.

透過無線網路基地台相互監測之適應性樣本比對定位

學生：許藍尹　　　　　　　　　　指導教授：曾煜棋 教授

國立交通大學網路工程研究所碩士班

摘　　要

近幾年以室內定位為基礎的位置感知服務(Location-Based Service)興起，顯示出室內定位的重要性。在室內定位系統中，又有許多研究著重在樣本比對演算法上。 然而，樣本比對定位系統的準確度會根據即時收到的訊號強度(RSS)和訓練的訊號分佈圖之間的比對來決定， 而且RSS可能會因為環境改變並和原來訓練的訊號分佈圖有所偏差。在這篇論文中，我們想要在既有WiFi環境下， 用環境中的基地台來即時監測環境變化，並透過這些基地台相互監測狀況，找出符合當時環境的訊號分佈圖來做樣本比對定位， 以提高定位系統精準度。這篇論文最主要和別人不同的地方是我們不需要透過額外硬體設備來監控環境變化。另外，我們從模擬和實驗中可以得知我們提出的兩個方法： 基於迴歸線和基於分群的方法都比傳統樣本比對定位在環境動態改變的情況下來得好。

關鍵字： K-means分群演算法、行動計算、迴歸模式、樣本比對定位、感測網路

Using Adaptive Radio Maps for Pattern-Matching Localization via Inter-Beacon Measurement

Student: Lan-Yin Hsu Advisor: Prof. Yu-Chee Tseng

Department of Computer Science

National Chiao Tung University

ABSTRACT

A growing number of location-based applications are probably based on indoor positioning system. In the indoor positioning, there is much effort focusing on the pattern-matching technique. However, the accuracy of pattern-matching positioning system depends on the comparison between current received signal strength (RSS) and the training radio map. The RSS might be different from those RSS in the training radio map due to the environmental dynamics. In this paper, we deployed the beacons in our environment, which can real-time monitor the environmental dynamics and we proposed a novel technique which can adapt the radio map to the environmental dynamics and predict the current radio map in the positioning phase based on the inter-beacon measurement. The main difference between other works is that we do not need the extra hardware to monitor the environmental dynamics. Our simulation and experiment show that both the regression-based and clustering-based methodology are better than original pattern-matching localization in the unstable environment.

Keywords: k-means clustering, mobile computing, regression model, pattern-matching localization, sensor network.

Acknowledgement

I have to thank many people for the assistance they have provided me during my research. Firstly, I have to express thanks to my advisor, Prof. Yu-Chee Tseng, for his guidance in my research. I have learned from him not only the skill of doing research and the writing skill of thesis but also the attitude toward everything. And I also appreciate my dissertation committee members: Prof. Wen-Tsuen Chen, Prof. Yeh-Ching Chung, Prof. Ruay-Shiung Chang, Prof. Hsi-Ya Chang. The information and advice they gave and their expertise and help is invaluable to me.

Besides, I am also grateful to my seniors, Dr. Sheng-Po Kuo, Chi-Chung Lo, and Che-Pin Chang. They gave me the invaluable information and help, so I have learned more from them. And I also thank to HSCC members who work and play with me, I will treasure the memory which they accompanied with me in HsinChu. Finally, I sincerely thank my family and my boy friend, Curis, for their love, encouragement, and support. With their company, I would be full of energy to finish the thesis.

Lan-Yin Hsu

Department of Computer Science

National Chiao Tung University

June 2010

Contents

List of Figures

Chapter 1

Introduction

Positioning system is one of the most popular issues for wireless technology and mobile computing used in commercial and research interest. A central task for developing this system is to enable mobile devices to attain its location and provide location-based service (LBS) [4], such as navigation and tracking. At present, GPS is a technology that is often used for car navigation, tracking, and other location-related services. But it cannot be used when the user is in the indoor environment due to the line of sight (LOS) problem, i.e., GPS signals cannot pass through buildings or other sheltered objects.

There is much research focusing on the development of wireless positioning in an indoor environment, which is based on using Radio Frequency (RF) to locate the mobile user (the person who wears mobile device and wants to be localized). The wireless positioning system is a system that utilizes RF as input and outputs the location of mobile user. A promising approach based on the pattern-matching technique has achieved meter-level accuracy [1, 5, 6, 10]. The pattern-matching technique consists of two phases: training phase and positioning phase. In the training phase, the operator which enables wireless (e.g. mobile devices and notebooks) is used to receive the signal strength from all the beacons in the environment at the fixed points and save those signal strength to the remote database. While in the positioning phase, the mobile device receives real time signal strength and sends to the server. The server compares the signal pattern with those patterns in the remote database, and then chooses the best-matched location for the mobile user.

However, the received signal strength (RSS) would fluctuate due to environmental dynamics, such as people blocking, temperature, humidity and temporal variation in the reality. As a result, positioning accuracy based on the pattern-matching technique would reduce rapidly. In the past, online determination of the training data has been proposed in LANDMARC system [8], LEMT system [11], and sensor-assisted system [2] to improve the reliability of radio maps and cope with the time-varying phenomenon of radio environment.

The LANDMARC system [8] uses reference tags to dynamically construct and update radio map. It could alleviate the effect caused by the fluctuation of signal strength in RF. The system computes the distances between the RSS vectors received from the tracking tags and those from different reference tags to get the user's location. The LEMT system [11] utilizes reference points to detect real-time RSS samples, and then uses regression analysis to adapt the measured radio map to estimated radio map. Finally the system attains the estimated location by nearest-neighbor method. The sensor-assisted system [2] produces radio maps through the mobile user real-time receives data and uses extra sensors to detect environmental dynamics. The LANDMARC system [8], LEMT system [11], and sensor-assisted system [2] were effective in the dynamic environment. However, they required additional hardware to detect the environmental dynamics.

The purpose of this paper was to propose a system which can monitor the environmental dynamics and do not need the extra hardware. We propose the scheme of real-time predicting radio maps and adapting radio map by using Inter-Beacon measuring system (the system with the beacons can receive signal strength from each other) to choose or produce the best-matched radio map that conform to environmental dynamics without any extra cost. We utilize the inter-

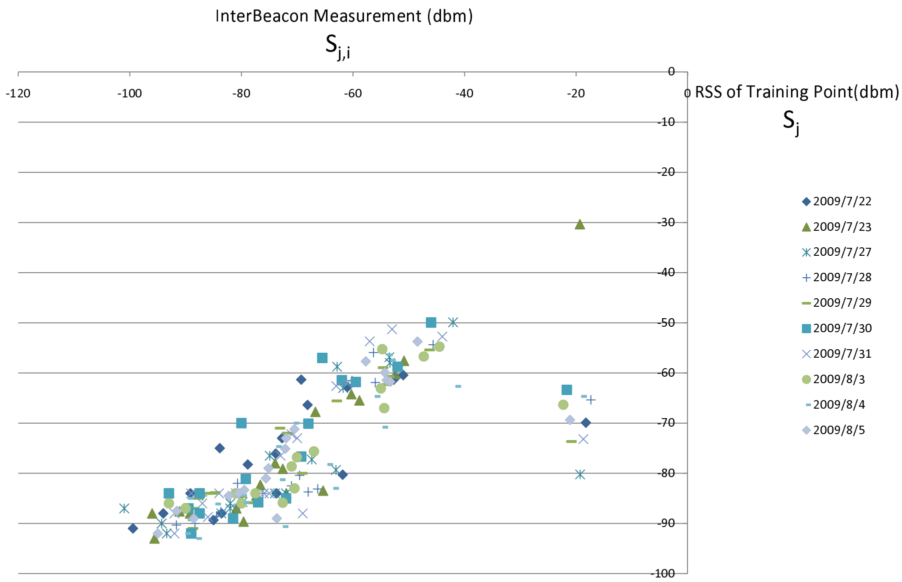


Figure 1.1: Tendency toward positive correlation.

beacon measuring the environmental dynamics, and save those inter-beacon measured values with signal patterns which received by mobile device to remote database for the positioning phase. Our scheme proposed two methodologies of finding the radio map that is almost in accordance with the current environment. Firstly, we can real-time measure the environmental dynamics and apply the appropriate radio maps to the current environment. Secondly, we can real-time predict the appropriate radio maps by the regression model. The regression line had a tendency toward positive correlation between the inter-beacon measured value and the RSS of signal pattern. Fig. 1.1 shows the observation. According to the observation, we can utilize the regression line to predict the radio maps in accordance to the current environment when given the inter-beacon measured value.

The rest of the paper is organized as follows. Chapter 2 describes the backgrounds of the pattern-matching localization. We explain our approach in Chapter 3. Chapter 4 and Chapter 5 demonstrates the performance of our system and the Chapter 6 concludes the paper.

Chapter 2

Backgrounds

Positioning systems can be classified as AoA-based, ToA-based, TDoA-based, and radio-based ones. In this paper, we are more interested in radio-based systems. Such system can be classified into two categories: *Radio-Propagation Model* [3] and *Empirical-Fit Model* [9].

In the first *Radio-Propagation Model* [3], it tries to use the perceived RSS vector and a multi-lateration mechanism to calculate the device's location. The path loss of a beacon b_j at a location l is modeled by:

$$P_r(\ell, b_j) = P_t(b_j) - 10\eta \log_{10}(\frac{d}{d_0}) + N(0, \sigma_f), \tag{2.1}$$

Where P_t is the transmit power of b_j, η is an environment-dependent constant, d_0 is a reference distance, d is the distance between b_j and l, and $N(0, \sigma_f)$ is a zero-mean normal distribution random variable with a standard deviation σ_f to represent background noise.

Suppose that the location of b_j is known, formula (2.1) helps us to measure the distance d. If at least three beacons can be heard, then through a multi-lateration algorithm, the location of the device can be predicted. Unfortunately, this model is less useful in indoor environments due to two reasons: (i) formula (2.1) is not so accurate in indoor environments and (ii) it is hard to model the locations of beacons in a multi-floor building.

The second *Empirical-Fit Model* [9] is also known as pattern-matching localization. It tries to empirically capture the RSS patterns at different locations for RSS-matching purpose. Given a set of beacons $\mathcal{B} = \{b_1, b_2, \ldots, b_m\}$ and a set of training locations $\mathcal{L} = \{\ell_1, \ell_2, \ldots, \ell_n\}$ in a sensing field $\mathcal{F} \subseteq \mathbb{R}^2$. Each beacon in $\mathcal{F}$ is capable of periodically transmitting radio signal. The system works in two phases. In the *training phase*, we measure the RSS vectors from beacons at each training location ℓ_i for a time period and create a *feature vector* $\boldsymbol{\upsilon}_i = [\upsilon_{i,1}, \upsilon_{i,2}, \ldots, \upsilon_{i,m}]$ for ℓ_i, where $\upsilon_{i,j} \in \mathbb{R}$ is the average RSS from b_j, $j = 1..m$. Those feature vectors are collected in a set $\mathcal{V} = \{\boldsymbol{\upsilon}_1, \boldsymbol{\upsilon}_2, \ldots, \boldsymbol{\upsilon}_n\}$, or called *radio maps*. In the *positioning phase*, a mobile user measures RSS vector $\boldsymbol{s} = [s_1, s_2, \ldots, s_m]$ and compares $\boldsymbol{s}$ against $\mathcal{V}$. The best matched one or ones in $\mathcal{V}$ are used to predict the mobile user's current location. For example, in [1], the distance between s and each l_i is defined as:

$$h(\ell_i) = \|\boldsymbol{s}, \boldsymbol{\upsilon}_i\| = \sum_{j=1}^{n} \sqrt{(s_j - \upsilon_{i,j})^2}. \tag{2.2}$$

The ℓ_i with the smallest distance is considered the predicted location.

Chapter 3

Adaptive Radio Maps via Inter-Beacon Measurement

An inherent limitation with the pattern-matching localization method is the signal instability problem. In this paper, we propose to use an *inter-beacon measurement* to alleviate this problem. We observe that most beacons used in practice have both transmitting and receiving capabilities (for example, WiFi and ZigBee stations are widely used as beacons). If we allow these beacons to measure neighboring beacons' RSSs, the result may help calibrate our radio maps. Consider the example in Fig. 3.1. Suppose that b_i and b_j are two beacons and l is a training

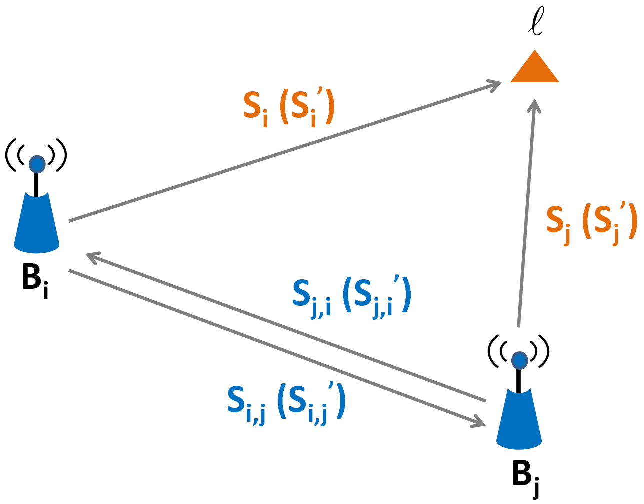


Figure 3.1: An example of applying inter-beacon measurement.

location. At the training phase, let S_i and S_j be the RSSs measured at l. During the positioning phase, suppose that a device arrives at l and measures the RSSs of b_i and b_j to be S_i' and S_j', respectively. If, unfortunately, $S_i^{'}$ and $S_j^{'}$ deviate too much from S_i and S_j, it will be very difficult to determine whether the device is at l or not. The main idea of our inter-beacon measurement method is to add two tags, $S_{i,j}$ and $S_{j,i}$, during the training phase to represent the RSS of b_i seen by b_j and the RSS of b_j seen by b_i, respectively. Then, during the positioning phase, in addition to measuring $S_i^{'}$ and $S_j^{'}$, we will also collect $S_{i,j}^{'}$ (the RSS of b_i seen by b_j) and $S_{j,i}^{'}$ (the RSS of b_j seen by b_i). It is expected that using the set $\{S_i, S_j, S_{i,j}, S_{j,i}\}$ collected from the training phase and the set $\{S_i^{'}, S_j^{'}, S_{i,j}^{'}, S_{j,i}^{'}\}$ collected from the positioning phase, we may have more clues to tell that the mobile device is now at location l.

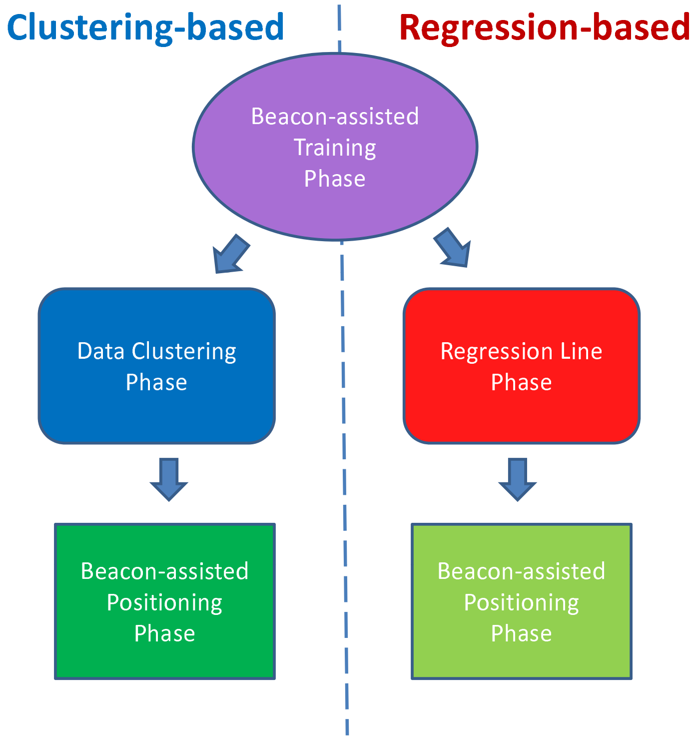


Figure 3.2: Flow chart of the proposed two solutions.

We are given a set of m beacons $\mathcal{B} = \{b_1, b_2, \ldots, b_m\}$ and a set of n training locations $\mathcal{L} = \{\ell_1, \ell_2, \ldots, \ell_n\}$. Below, we propose two solutions. The first solution uses the inter-beacon measurement to cluster the training data into multiple radio maps. To position a device, we first use the current inter-beacon measurement to pick an appropriate radio map, from which the closest location is predicted. The second solution is similar, but we further use the current inter-beacon measurement to interpolate the current radio map. Both solutions consist of three phases, as illustrated in Fig. 3.2. These two solutions both share the same first phase. But they differ in their second and third phases.

3.1 Solution 1: Clustering-based methodology

Beacon-assisted Training Phase: In this phase, at each training location, we will collect RSSs from all potential beacons at different times, which we call the *beacon-to-device vectors*. In addition, whenever a beacon-to-device vector is collected we will ask beacons to switch to receive mode and monitor each other's signals, which we call *beacon-to-beacon vector*. Such beacon-to-beacon vectors are to reflect the environmental factors when the corresponding beacon-to-device vectors are collected. Specifically, for each training location ℓ_i, we will collect multiple beacon-to-device and beacon-to-beacon vector pairs at different times to represent the diversity of the environment. Each beacon-to-device vector has the format $\upsilon_i^{(x)} = [\upsilon_{i,j}^{(x)}]_{j=1\ldots m}$, where x is the timestamp when the vector is measured and $\upsilon_{i,j}^{(x)}$ is the RSS of the signal emitted by beacon b_j measured by the device. Also, at time x, we will establish a beacon-to-beacon vector $\mu_i^{(x)} = [\mu_{i,j,k}^{(x)}]_{j=1\ldots m, k=1\ldots m, j\neq k}$, where $\mu_{i,j,k}^{(x)}$ is the RSS of the signal emitted by beacon b_j measured by beacon b_k. (Note that in practice, the measurements at time x can be the average of several samples around time x.) The vector pair at time x is written as ($\upsilon_i^{(x)}$, $\mu_i^{(x)}$).Then we collect all the above vector pairs together into a training data set $\mathcal{T} = \{(\upsilon_i^{(x)}, \mu_i^{(x)}) \mid \forall l_i \in \mathcal{L}, \forall x\}$, called the complete radio map.

Data Clustering Phase: As mentioned above, for each vector pair, the former is the RSSs observed by the device and the later is to reflect the corresponding environment factor. Therefore, in this phase, we will partition the training set $\mathcal{T}$ into several subsets according to their environment factors. We apply the following modified k-means clustering algorithm [5, 7] to achieve this goal.

1. Collect all beacon-to-beacon vectors into a set $\mu = \{\mu_i^{(x)} \mid \forall i, \forall x\}$.

2. Partition μ into k subsets using the k-means algorithm. Let the result be μ_1, μ_2,..., μ_k. (Note that in the k-means algorithm, when computing the similarity of two vectors $\mu_p^{(x)}$ and $\mu_q^{(y)}$, we define their distance to be

$$d(\mu_p^{(x)}, \mu_q^{(y)}) = \sqrt{\sum_{\forall j \forall k} (\mu_{p,j,k}^{(x)} - \mu_{q,j,k}^{(y)})^2}. \tag{3.1}$$

The k-means algorithm puts vectors of smaller distances together.)

3. Partition $\mathcal{T}$ into k subsets $\mathcal{T}_1$, $\mathcal{T}_2$,...,$\mathcal{T}_k$ according to the result of step 2. Specially, for each $i = 1...k$ and each $\mu_j^{(x)} \in \mu_i$, we include the corresponding $(\upsilon_j^{(x)}, \mu_j^{(x)})$ into set $\mathcal{T}_i$. For each subset $\mathcal{T}_i$, we let its feature environment vector be $\omega_i = [\omega_{i,j,k}]_{j=1...m,k=1...m,j\neq k}$, where $\omega_{i,j,k} = \dfrac{\sum_{\forall i \forall x} \mu_{i,j,k}^{(x)}}{|\mathcal{T}_i|}$.

4. The above process may not properly partition the training data of a location ℓ_i into subsets $\mathcal{T}_1$, $\mathcal{T}_2$,...,$\mathcal{T}_k$. That is, some subsets $\mathcal{T}_j$ may not contain any training data of ℓ_i. If this happens to any ℓ_i and $\mathcal{T}_j$, we compare the feature vector ω_j against $\mu_i^{(x)}$ for all x. Let $\mu_i^{(x)}$ be the one such that $d(\mu_i^{(x)}, \omega_i)$ is the smallest. We add the pair $(\upsilon_i^{(x)}, \mu_i^{(x)})$ into $\mathcal{T}_j$. This action has two effects. First, this pair $(\upsilon_i^{(x)}, \mu_i^{(x)})$ thus appear in more than one subset. Second, when we use $\mathcal{T}_j$ as a radio map for localization (see the third phase), it is ensured that each ℓ_i has at least one training data in $\mathcal{T}_j$.

Beacon-assisted Positioning Phase: When a device needs to determine its current location, it measures RSSs from all beacons, denoted by vector $\tilde{\upsilon} = [\tilde{\upsilon}_1, \tilde{\upsilon}_2, ..., \tilde{\upsilon}_m]$, where $\tilde{\upsilon}_i$ is the RSS of beacon b_i, $i = 1...m$. Then it submits $\tilde{\upsilon}$ to the location server, which determines the device's location as follows.

1. The server firstly asks all beacons to measure each other's RSS. Let the current beacon-to-beacon vector be $\tilde{\mu} = [\tilde{\mu}_{j,k}]_{j=1...m,k=1...m,j\neq k}$, where $\tilde{\mu}_{j,k}$ is the RSS of b_j measured by

b_k.

2. The server compares $\tilde{\mu}$ against the feature vector ω_i of each $\mathcal{T}_i$. Let ω_i be the one such that $d(\tilde{\mu}, \omega_i)$ is the smallest. Then the location server can determine mobile user's location by using formula (2.2) to compare its real-time signal pattern $\tilde{\upsilon}$ against $\mathcal{T}_i$ (with smallest $h(\mathcal{T}_i)$).

Note that, firstly, all the measurements we list above are considered as the average of a short-term measurement. Secondly, in order to avoid the short-term interference and high computation cost, the beacon-to-beacon vector can be collected periodically and then the server can determine the most appropriate sub-database.

3.2 Solution 2: Regression-based methodology

Beacon-assisted Training Phase: (This is the same as Solution 1.)

Regression Lines Phase: In Fig. 1.1, we can observe that the beacon-to-beacon vector and beacon-to-device vector have a tendency toward positive correlation. Therefore, in this phase, we will further use the current inter-beacon measurement to interpolate the current radio map. We apply the following regression analysis to achieve this goal.

1. As mentioned above, the training data set is $\mathcal{T} = \{(\upsilon_i^{(x)}, \mu_i^{(x)}) \mid \forall l_i \in \mathcal{L}, \forall x\}$. Considering each location l_i, each beacon b_j, and transmitter beacon b_k, we assume that there is a predictive relationship like $\upsilon_{i,k} = a \times \mu_{i,j,k} + b$. Let $\mathbf{p}_j = [a_j, b_j]^T$. We can put these x-times measurements together in linear regression analysis $A \times \mathbf{p} = C$:

$$\underbrace{\begin{bmatrix} 1 & \upsilon_{i,k}^{(1)} \\ \vdots & \vdots \\ 1 & \upsilon_{i,k}^{(x)} \end{bmatrix}}_{A} \times \underbrace{\begin{bmatrix} a_j \\ b_j \end{bmatrix}}_{\mathbf{p}_\mathrm{j}} = \underbrace{\begin{bmatrix} \mu_{i,j,k}^{(1)} \\ \vdots \\ \mu_{i,j,k}^{(x)} \end{bmatrix}}_{C} \qquad (3.2)$$

Then the value of **p** can be measured by the least-squares analysis:

$$\mathbf{p} = [a, b]^T = (A^T A)^{-1} A^T C. \tag{3.3}$$

Note that the total number of beacon b_j are $m-1$ and $j \neq k$. In this case, each beacon b_j will associate with its coefficients set $\mathbf{p}_j$.

2. However, some regression lines are not suitable for predicting the RSS, such as the vertical lines, horizontal lines. Therefore, if the degrees between the regression line and x-axis is larger than 80 degrees or smaller than 10 degrees, those data of regression lines will be deleted.

3. For each location l_i and transmitter beacon b_k, we have $m-1$ coefficients set $\mathbf{p}_j$. We think the correlation coefficient is a good parametric to decide which $\mathbf{p}_j$ to be used. Here, we use Pearson's correlation coefficient. The correlation coefficient between two random variables X and Y with expected values $\overline{X}$ and $\overline{Y}$ and standard deviations σ_X and σ_Y is defined as:

$$\rho = \frac{cov(X, Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \overline{X})(Y - \overline{Y})]}{\sigma_X \sigma_Y} \tag{3.4}$$

where E is the expected value operator and cov means covariance. Therefore, we could define ρ_j:

$$\rho_j = \frac{cov(\upsilon_{i,k}, \mu_{i,j,k})}{\sigma_{\upsilon_{i,k}} \sigma_{\mu_{i,j,k}}}. \tag{3.5}$$

4. In the end, we claim the coefficients set $\mathbf{p}_j$ by largest ρ_j and it must be more than 0.

Note that there may not exist the coefficients set $\mathbf{p}_j$ that its' correlation coefficient $\rho_j > 0$. (Because the location l_i, all the beacon b_j, and transmitter beacon b_k would have lower relationship or sometimes the signal patterns and the inter-beacon measurement have the relationship of a small group, not a line.) Therefore, if we can not find the coefficients set $\mathbf{p}_j$, we say that is *unpredictable*.

The Beacon-assisted Positioning Phase: When a device needs to determine its current location, it measures RSSs from all beacons, denoted by vector $\tilde{\upsilon} = [\tilde{\upsilon}_1, \tilde{\upsilon}_2, ..., \tilde{\upsilon}_m]$, where $\tilde{\upsilon}_i$ is the RSS of beacon b_i, $i = 1...m$. Then it submits $\tilde{\upsilon}$ to the location server, which determines the device's location as follows.

1. The server firstly asks all beacons to measure each other's RSS. Let the current beacon-to-beacon vector be $\tilde{\mu} = [\tilde{\mu}_{j,k}]_{j=1...m,k=1...m,j\neq k}$, where $\tilde{\mu}_{j,k}$ is the RSS of b_j measured by b_k.

2. The server predicts the expected radio map, denoted by $\mathcal{V}' = \{\boldsymbol{\upsilon}'_1, \boldsymbol{\upsilon}'_2, \ldots, \boldsymbol{\upsilon}'_n\}$. For each location l_i, the server predicts the expected radio map by the current beacon-to-beacon vector $\tilde{\mu}$ and coefficients set $\mathbf{p}_j$ (replace into $\upsilon'_{i,k} = a \times \tilde{\mu}_{j,k} + b$). Specifically, that is given location l_i and transmitter beacon b_k, we could determine the expected RSS $\upsilon'_i = [\upsilon'_{i,1}, ..., \upsilon'_{i,k}, ..., \upsilon'_{i,m}]$ by coefficients set $\mathbf{p}_j$, where $k = 1...m$.

3. If $\upsilon'_{i,k}$ is unpredictable, the server compares $\tilde{\mu}$ against the feature vector ω_i of each $\mathcal{T}_i$. Let ω_i be the one such that $d(\tilde{\mu}, \omega_i)$ is the smallest, where $i = 1...k$. In $\mathcal{T}_i$, the server determines $\upsilon'_{i,k}$ from all the training data set at location l_i, denoted by:
$$\upsilon'_{i,k} = \frac{\sum_{\phi=1}^{n} \upsilon^{\phi}_{i,k}}{n}. \tag{3.6}$$

4. Finally, the location server can determine mobile user's location by using formula (2.2) to compare its real-time signal pattern $\tilde{\upsilon}$ against $\mathcal{V}'$.

Note that, avoid the short-term interference and high computation cost, the beacon-to-beacon vector can be collected periodically and then the server can predict the expected radio map, $\mathcal{V}'$, periodically.

Chapter 4

Simulation Results

The objective of simulation is to demonstrate how much accuracy of our localization system surrounded with environment dynamics. We implemented the *Ideal NNSS*, *NNSS*, *Clustering-based methodology* and *Regression-based methodology* for comparison. *Ideal NNSS* is an ideal situation, that is, the environmental condition of the training phase is the same as in the positioning phase. While *NNSS* is the situation considered with environmental change, the environmental condition of the training phase is different in the positioning phase. Based on the simulation results, we discuss (1) What do the noisy environment affect the system. (2) How many beacons do we need to minimize error distance. (3) How irregular do the beacons deploy would have effect on the error distance. (4) What if the environment with the wall, how heavily the accuracy of our scheme would be affected. (5) How long do our system cost to do localization.

4.1 Environmental setting

Assume our simulation environment is a sensing field of size 50m * 50 m, and for each location we train 3 samples, and we set up grid size of 1m apart from each location. The radio power is set 15db to ensure that the interbeacon and device or interbeacon and interbeacon could be reachable to each other. To simulate the movement of the user, we use random waypoint mobility model with moving speed 1m/s and sampling period is 1s. In the *Ideal NNSS* case, the environmental situation is set to dynamic change by Gaussian distribution with 0.6 standard deviation. Consider the interference of walls, the setting of walls and the mobility path in our simulation shows in Fig. 4.1.

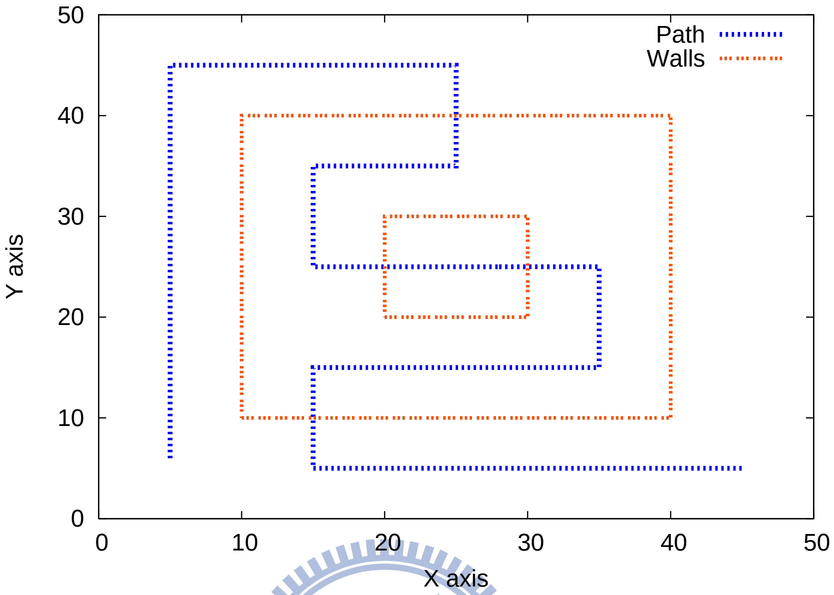
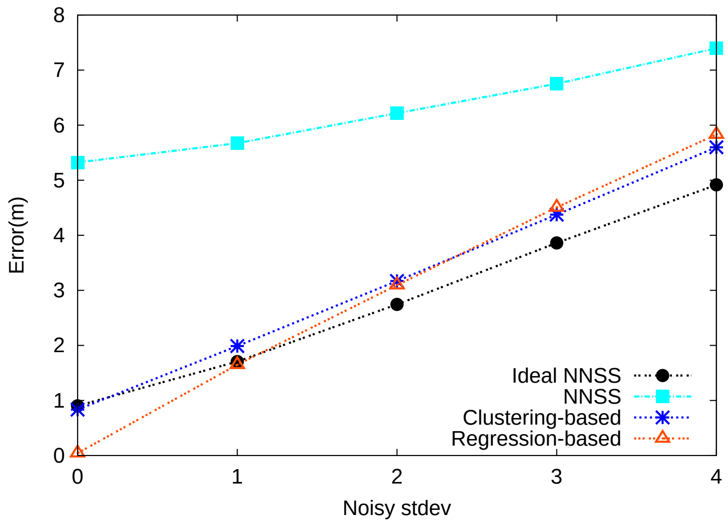


Figure 4.1: The simulated setting.

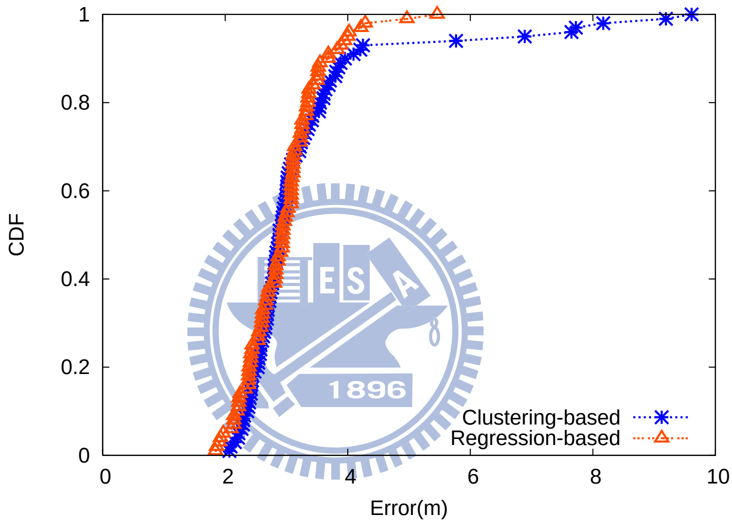
4.2 Impact of environmental noise

Environmental noise affects the parameter η of path-loss model, and also affects the RSS in the RF-based localization. As a result, we want to use the beacon measures real-time environmental dynamics to detect how greatly does the environmental noise affect. In this set of simulation, we vary the noisy factor from 2 to 4, and the environmental dynamics change through Gaussian distribution.

We found that in the *Ideal NNSS*, whose error distance is increasing when the noise factor increased; whereas in the *NNSS*, whose largest error distance is 7.4m. Compare to our methodologies, we can reduce the error distance to 5m with the environmental dynamics and just have more than 1m of error compared to the *Ideal NNSS*. In fact, we can achieve definite accuracy when the environment is changeful. The simulation results are shown in Fig. 4.2(a). For



(a) Noisy environment with different STDEV



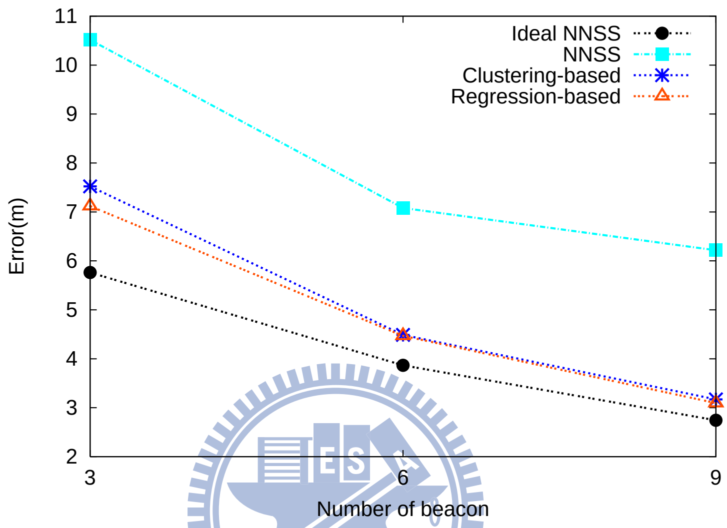
(b) CDF of Clustering-based and Regression-based (STDEV=2)

Figure 4.2: Impact of environmental noise

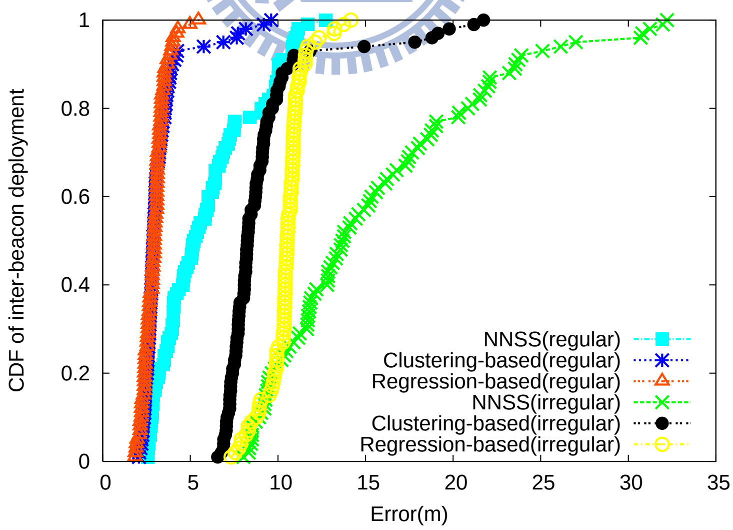
noisy standard deviation 0, our *Regression-based methodology* is better than *Clustering-based methodology*, and their performance would properly the same when the noisy standard deviation becomes larger. But in Fig. 4.2(b), we can see the CDF of *Regression-based methodology* and *Clustering-based methodology*. From the figure of CDF, the *Regression-based methodology* is better than *Clustering-based methodology* because *Regression-based methodology* is good at handling worst case, for example, *Regression-based methodology* could predict the adaptive radio map that does not exist in our database.

4.3 Impact of number and deployment of beacons

The number of beacon in the environment would affect the accuracy of pattern-matching localization. In this set of simulation, we set up 3, 6, 9 beacons in the environment, we want to see how many beacons do we need, and how much accuracy would increase if the number of beacons increase.



(a) Impact of number of beacon.



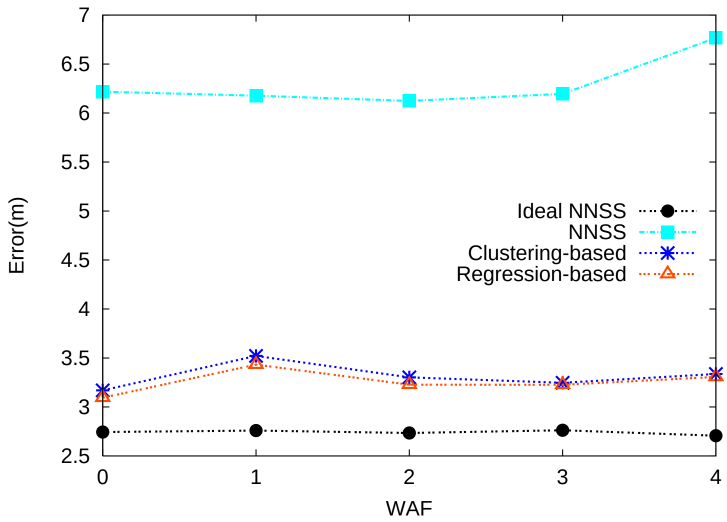
(b) Impact of deployment of beacons

Figure 4.3: Impact of number and deployment of beacons.

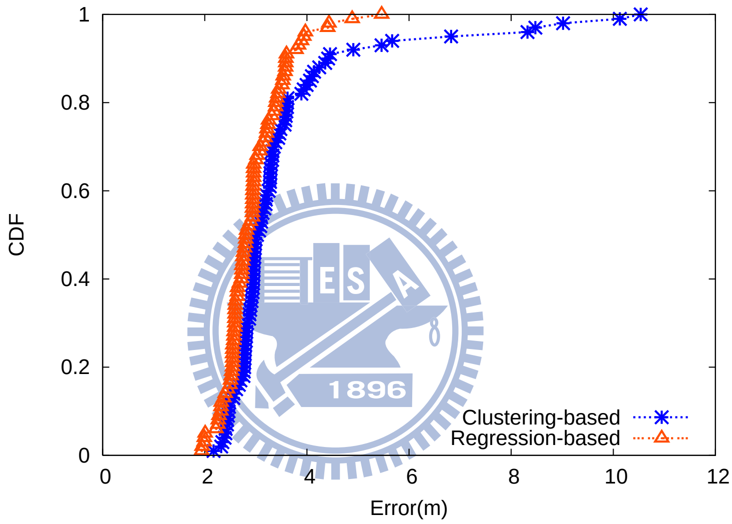
Simulation results shown in Fig. 4.3(a). For *Regression-based methodology* and *Clustering-based methodology*, we found that the performance would increase when the number of beacons increased. On the other hand, we set up the deployment of beacons as regularity and irregularity. We want to see how heavily would the location of beacons affect the accuracy in our system. The result is shown in Fig. 4.3(b). From the figure, we found that in *Regression-based methodology* and *Clustering-based methodology*, we have the better performance than the *NNSS* no matter what kind of deployment. But in irregular deployment, we also found that *Clustering-based methodology* is slightly better than *Regression-based methodology*. So the irregular deployment of beacons would have much effect on *Regression-based methodology*.

4.4 Impact of interference of wall

From Fig. 4.1 we can see the setting of wall in our simulation, and we want to see if the setting would have impact on the accuracy. We have two setting about walls. Firstly, we deploy all phases in *Regression-based methodology* and *Clustering-based methodology* with same setting of wall as in Fig. 4.1. The result is shown in Fig. 4.4(a). The performance of *Regression-based methodology* and *Clustering-based methodology* are similar, and they are better than the *NNSS*. Secondly, we train the radio maps with walls and without walls in the beacon-assisted training phase, and in the beacon-assisted positioning phase we run the environment with walls and without walls, half for each other. The reason is that if the environment with the interference of walls , *Regression-based methodology* and *Clustering-based methodology* would choose the adaptive radio map for pattern-matching localization. The result of CDF is shown in Fig. 4.4(b). *Regression-based methodology* is slightly better than *Clustering-based methodology*, because the handling of worst case in *Regression-based methodology* is better than in *Clustering-based methodology*.



(a) Impact of interference of wall.



(b) Impact of random walls.

Figure 4.4: Impact of interference of wall.

4.5 Impact of cost time

As described before, in order to reduce the online positioning time, we proposed the *Clustering-based methodology*. As a result, the cost time is the measurement that we want to decrease. We add the *NNSS with all Radio Maps* for comparison. *NNSS with all Radio Maps* means the *NNSS* not only includes one radio map but also many radio maps for pattern-matching localization. The result is shown in Fig. 4.5. The building time is the time we have to load the adaptive radio map to our system, while the positioning time is the total time our positioning system reports the location to the user who runs the overall route (Fig. 4.1).

	Ideal NNSS	NNSS	Clustering-based	Regression-based	NNSS with all Radio Maps
Building Time (sec)	0.3654	0.37005	1.5104	2.8332	3.8611
Positioning Time (sec)	2.88985	2.86435	3.0522	1.3547	31.48025
Total Cost Time (sec)	3.25525	3.2344	4.5626	4.1879	35.34135

Figure 4.5: The cost time of one route.

From the Fig. 4.5 we can see that *NNSS with all Radio Maps* has the worst performance. *Regression-based methodology* is slightly faster than *Clustering-based methodology*, even though the building time in *Regression-based methodology* is longer than in *Clustering-based methodology*. *Regression-based methodology* and *Clustering-based methodology* are about one second slower than the *Ideal NNSS* and *NNSS*.

Chapter 5

Experimental Results

In this section, we will demonstrate our experimental environment and evaluate the performance of *Regression-based methodology* and *Clustering-based methodology*. And we also compare with *NNSS*.

5.1 Environmental setting

Our environment is shown in Fig. 5.1. The engineering building at second floor is our environment, and its height and width is 41 meters and 68 meters, separately. We trained the half environment for 20 days and there are about 60 training locations for each day. From Fig. 5.1, the blue points are training locations, while the red points are beacons.

5.2 Experimental results

The experimental results is shown in Fig. 5.2. *Clustering-based methodology* is better than *Regression-based methodology* in practical situation. Because the RSS of beacons in the environment may not be received by operator every day, it would have the effect on the *Regression-based methodology* which is worse than *Clustering-based methodology*. But both *Clustering-based methodology* and *Regression-based methodology* are better than *NNSS* in practical.

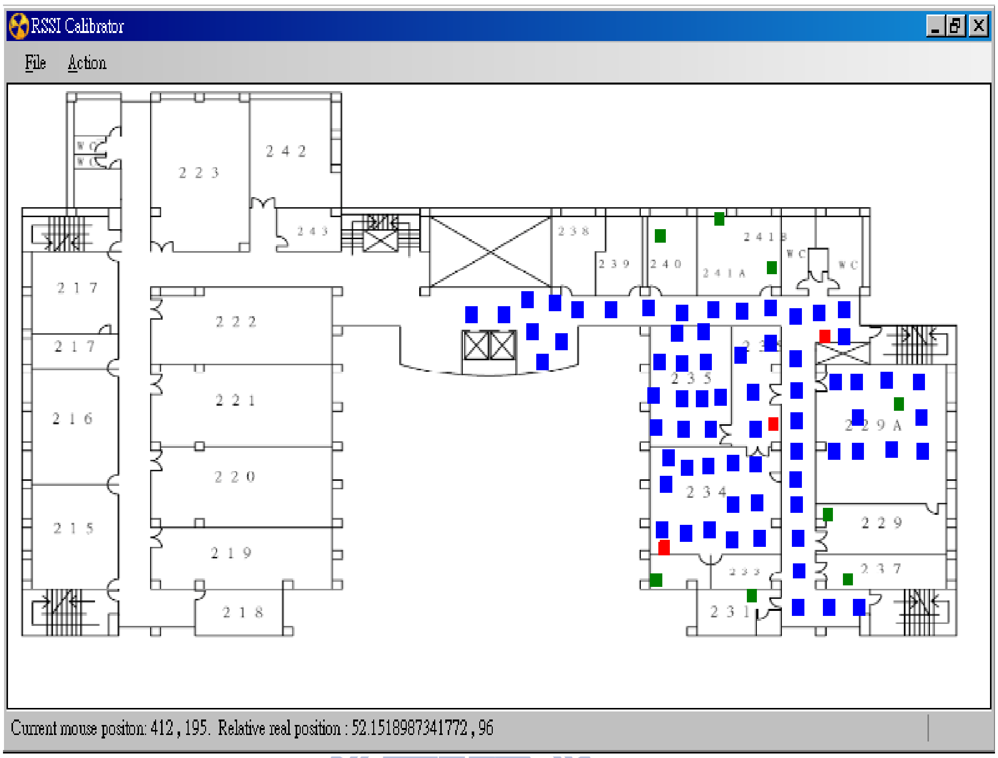


Figure 5.1: The environmental setting.

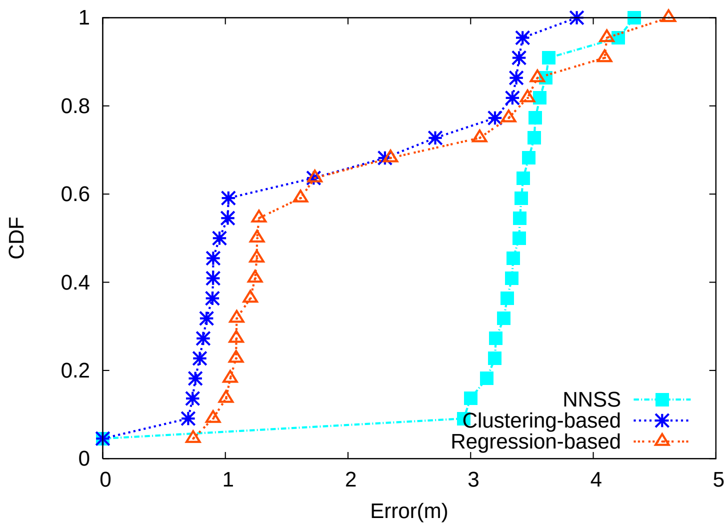


Figure 5.2: The CDF of experimental results.

Chapter 6

Conclusions

This work aims at the changing of environmental dynamics that might adversely affects the positioning accuracy in pattern-matching localization system. Unlike the previous works, the main advantage of inter-beacon measurement focuses on the WiFi beacons and localization system. There is no need for extra hardware. To reduce the adverse effect of the computing cost in the positioning phase, we use k-means clustering algorithm to partition the location database. Additionally, we can predict the radio map which is not original in our database by interpolating the regression model. Our simulation results have shown that the performance of *Regression-based methodology* and *Clustering-based methodology* is similar, but *Regression-based methodology* is better at handling the worst case than *Clustering-based methodology*. For the aspect of cost time, *Regression-based methodology* is faster than *Clustering-based methodology* even though the building time of *Regression-based methodology* is longer than *Clustering-based methodology*. In practical situation, *Regression-based methodology* is worse than *Clustering-based methodology* because the RSS of beacons is not continuously received. As a whole, the error distance of our methodologies is reduced compared to traditional non-adaptive localization system. Besides, we found that our system only needs fewer beacons to provide good performance.

Bibliography

[1] P. Bahl and V. N. Padmanabhan. RADAR: An In-building RF-based User Location and Tracking System. In *IEEE INFOCOM*, volume 2, pages 775–784, 2000.

[2] Y.-C. Chen, J.-R. Chiang, H. hua Chu, P. Huang, and A. W. Tsui. Sensor-assisted Wi-fi Indoor Location System for Adapting to Environmental Dynamics. In *Proc. 8th ACM Intl' Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, pages 118–125, 2005.

[3] K. Kaemarungsi and P. Krishnamurthy. Modeling of Indoor Positioning Systems Based on Location Fingerprinting. In *IEEE INFOCOM*, volume 2, pages 1012–1022, 2004.

[4] S.-P. Kuo, S.-C. Lin, B.-J. Wu, Y.-C. Tseng, and C.-C. Shen. GeoAds: A Middleware Architecture for Music Service with Location-Aware Advertisement. In *IEEE Int'l Conf. on Mobile Ad-hoc and Sensor Systems (MASS)*, 2007.

[5] S.-P. Kuo, B.-J. Wu, W.-C. Peng, and Y.-C. Tseng. Cluster-Enhanced Techniques for Pattern-Matching Localization Systems. In *IEEE Int'l Conf. on Mobile Ad-hoc and Sensor Systems (MASS)*, 2007.

[6] J. Letchner, D. Fox, and A. LaMarca. Large-Scale Localization from Wireless Signal Strength. In *Proc. of the Nat'l Conf. on Artificial Intelligence (AAAI)*, pages 15–20, 2005.

[7] J. MacQueen. Some Methods for Classification and Analysis of Multivariate Observations. In *Proc. 5th Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 281–297, 1967.

[8] L. Ni, Y. Liu, Y. C. Lau, and A. Patil. Landmarc: indoor location sensing using active rfid. pages 407–415, March 2003.

[9] J. J. Pan, J. T. Kwok, Q. Yang, and Y. Chen. Multidimensional Vector Regression for Accurate and Low-Cost Location Estimation in Pervasive Computing. *IEEE Trans. on Knowledge and Data Engineering*, 18(9):1181–1193, 2006.

[10] T. Roos, P. Myllymäki, H. Tirri, P. Misikangas, and J. Sievänen. A Probabilistic Approach to WLAN User Location Estimation. *Int'l Journal of Wireless Information Networks*, 9(3):155–164, 2002.

[11] J. Yin, Q. Yang, and L. M. Ni. Learning adaptive temporal radio maps for signal-strength-based location estimation. *IEEE Transactions on Mobile Computing*, 7(7):869–883, 2008.

Curriculum Vita

Lan-Yin Hsu

Contact Information
Department of Computer Science
National Chiao Tung University
1001 Ta Hsueh Road, Hsinchu, Taiwan 300
Email: lyhsu@cs.nctu.edu.tw

Education

M.S.: Computer Science, National Chiao Tung University (2008.9 ∼ 2010.6)
B.S.: Computer Science, National Cheng Chi University (2004.9 ∼ 2008.6)

Awards

1. 2008年 第十屆WeWin創業大賽 入圍.
2. 2008年 U活智慧王RFID TENDENCY CUP 入圍.
3. 2009年 第九屆旺宏金矽獎 優勝獎(GEO定位王).
4. 2009年 第九屆旺宏金矽獎 優勝獎(智慧型展場導覽系統).
5. 2009年 全國RFID校園創意競賽 佳作.
6. 2009年 MobileHeros通訊大賽-Android手機軟體設計競賽 最佳校園創意獎及企業特別獎（高通獎）.

Publication Lists

Chi-Chung Alan Lo, Lan-Yin Hsu, and Yu-Chee Tseng, ”利用基地台間相互監測之適應性樣本定位系統”, *Wireless Ad Hoc and Sensor Networks (WASN)*, 2009.