Wireless Tracking Analysis in Location Fingerprinting

Eddie C.L. Chan  
Department of Computing  
The Hong Kong Polytechnic University  
Hung Hom, Kowloon, Hong Kong  
Email: csclchan@comp.polyu.edu.hk

George Baciu  
Department of Computing  
The Hong Kong Polytechnic University  
Hung Hom, Kowloon, Hong Kong  
Email: csgeorge@comp.polyu.edu.hk

S.C. Mak  
Department of Computing  
The Hong Kong Polytechnic University  
Hung Hom, Kowloon, Hong Kong  
Email: csscmak@comp.polyu.edu.hk

Abstract—Wireless tracking analysis is useful for deploying the efficient indoor positioning system. Location Fingerprinting (LF) method uses a training dataset of Wi-Fi received signal strength (RSS) at different location to track the target. Fuzzy logic modeling can be applied to evaluate the behavior of wireless received signal strength (RSS). Previous analytical models based on LF are not sufficient for modeling spatial factors of wireless coverage. Spatial analytical model is useful for analysis of how the wireless infrastructure affecting the accuracy of positioning. The main concept of fuzzy logic is to reflect the reality of our world of experience, which is uncertain and fuzzy. In this paper, we develop a multi-layer fuzzy modeling for the wireless coverage in the huge and open area. Large scale site surveying has been used to collect RSS in 9.34 hectare campus area. The color fuzzy model allows us to visualize the spatial distribution of wireless RSS. Base on the fuzzy analytical model, we analyze the effect of existence of human’s presence and large obstacle, the accuracy and efficiency of tracking system.

I. INTRODUCTION

Wireless tracking technology such as Global Positioning System (GPS) technology is the most effective in relatively open and flat outdoor environments but is much less effective in non-line-of-sight (NLOS) environments such as hilly, mountainous, or built-up areas. In recent years, positioning systems for indoor areas using the existing wireless local area network (WLAN) infrastructure have been suggested. Location Fingerprinting (LF) method uses a training dataset of Wi-Fi received signal strength (RSS) at different location to track the target.[1] Moreover, there is a lack of analytical models that can be used as a framework for designing and deploying the positioning systems.

The analytical model should have spatial elements to visualize the RSS distribution, evaluate and predict precision performance of indoor positioning systems. Such models can be used to improve the design of positioning systems, for example by eliminating some fingerprints and reducing the size of the location fingerprint database. The analytical model can help to optimize the tracking system by designing the distribution of access points. For example, there might be too many access points, packed too closely together in a region which might lead to overlaps of signals, cause interference and potential security risks. Some access points can even be removed to achieve the same signal stability level. On the other hand, there might be fewer access points in a region. This leads to weak signals and cause unusable wireless connections.

The rest of the paper is organized as follows: Section II describes the related work on modeling indoor positioning system and method of positioning system. Section III discusses the measurement setup of large scale site RSS surveying in 9.34 hectare campus area. Section IV presents the methodology of new analytical model to derive the RSS distribution based on the 2D multi-layer fuzzy map. Section V presents the analysis result of obstacles, human bodies and Wi-Fi APs location. Section VI discuss the accuracy and computational complexity of the proposed solution. Finally, conclusion and future work of the paper are presented in Sections VII respectively. The contribution of the paper is to provide a spatial analytical model for LF. The usage of model can increase the efficiency usage of wireless infrastructure and improve the accuracy positioning system.

II. RELATED WORK

Most existing analysis of LF method [3] and [6] base on accuracy performance of positioning system and proximity graph, such as Voronoi diagram, clustering graph. Some research works [5], [6] and [7] ignores the radio signal properties and some assumes the distribution of the RSS is in Gaussian and pair wise. In [8], the distribution of the RSS has been proven to be not usually Gaussian, it is often left-skewed and the standard deviation varies according to the signal level. It is clear from their measurements that signals from multiple access points (APs) are mostly independent and the interference from other APs using the same frequency does not have a significant impact on the RSS pattern. More than one cluster may represent one location because of the multimodal distribution of the RSS. In such a case, using a simple Euclidean distance as to determine the location may classify some patterns into a wrong location easily.

Considering the algorithm for positioning of objects in space, recent research [2] aims to combine the benefits of the Radio Frequency(RF) propagation loss model and fingerprinting method. As a result, a hybrid algorithm has been suggested for indoor positioning using wireless LAN. The propagation loss method is known to perform poorly compared to LF. But LF requires an extensive training dataset and cannot adapt
well to configuration changes or receiver breakdown. If there is no training dataset, their hybrid method coincides with RF propagation loss model method. As the spatial granularity of training dataset increases, the regression function approaches the one (implicitly) used by fingerprinting and the result of the algorithm become close to that of the fingerprinting method. But even in this case, unlike in fingerprinting, the hybrid method retains characteristics, i.e., it makes intelligent use of missing values, produces an error bound, and can be made dynamic.

The fuzzy visualization map concepts widely applied in other fields, such as temperature, rainfall and atmosphere. Moreover, these concepts have not been applied in the modeling of wireless positioning system. The problem of modeling relationships among wireless infrastructure, large obstacles and spatial distribution of RSS has been studied in this paper using the concepts of multi-layer fuzzy model and 2D propagation based simulation model. The fuzzy map models helps to find the where the RSS is denser and cluster different RSS in different color layer.

III. MEASUREMENT SETUP

In this section, we will base on [1], [2], [3], [9], [10], [11] and [12] method for measuring RSS setup. A standard laptop computer equipped with an Intel WLAN card and client manager software was used to collect samples of RSS from APs inside the Hong Kong Polytechnic University (PolyU) campus. The WLAN card is chipset inside the laptop. There are basically 26 buildings from Core A to Core Z and 7 extra buildings with Wi-Fi access. Each core building is covered by at least 13 APs. Table I shows the distribution of access points in each building in PolyU campus. The approximate total area of the campus is 9.34 hectare. The radio frequency channels of IEEE 802.11b are in the 2.4 GHz band which is shared by other equipment in the industrial, scientific, and medical (ISM) band such as Bluetooth. The number of non-overlapping channels for 802.11b is three. We observe that the RSS value reported by the WLAN card is an average value over a sampling period and in integral steps of 1 dBm. The received signal sensitivity of the WLAN card also limits the range of the RSS to be between -93 dBm and -15 dBm. Nevertheless, the highest typical value of the RSS is approximately -30 dBm at one meter from any AP.

A. SAMPLING SCHEDULE

The sampling schedule is to collect the RSS data every 5 seconds. The vector of RSS data at each location forms the location fingerprint with around 20 RSS elements in the vector. Total 27 locations of measurement are chosen in the campus. Figure 1 shows the 27 locations site plan. The radio channels used for each AP are channel 11, 6, and 1, respectively. The sampling will be taken with two periods of time, (7:30am-9:30am) (leisure) and (4:30pm - 6:30pm) (busy)). From [3], the presence or absence of people in the building significantly affects the RSS values. The data were collected four times with four different directions, North, South, East and West. The duration of sampling was 2 weeks with total 12 days (from Mon to Sat). In meanwhile, temperature, weather, sampling time and humidity were recorded. The total number of RSS samples would be 12 days X 4 directions X 27 buildings X 20 APs X 2 Times = 51840.

B. SURVEYING TOOL FOR RSS RETRIEVAL

The surveying tool is a Wi-Fi enabled device to detect RSS readings. A Place Lab is a piece of software produced by Intel research group at placelab.org [13]. Place Lab is a JAVA based radio beacons auditing tools, such as 802.11 access points, GSM cell phone towers, and fixed Bluetooth devices that already exist in large numbers around us in the environment. Its role is to detect wireless access points on

<table>
<thead>
<tr>
<th>Area</th>
<th>Location</th>
<th>No. of Access Point</th>
</tr>
</thead>
<tbody>
<tr>
<td>Library</td>
<td>1/F (rectangular wing only)</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2/F (rectangular wing only)</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>3/F (rectangular wing only)</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>4/F (rectangular wing only)</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>5/F (rectangular wing only)</td>
<td>2</td>
</tr>
<tr>
<td>Sports Complex</td>
<td>Staff Canteen</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>G/F Open Area</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1/F Open Area</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2/F &amp; 3/F Open Area</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Main Gymnasium</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Fong Shu Chuen Hall</td>
<td>1</td>
</tr>
<tr>
<td>Communal Building On Jubilee Sports</td>
<td>G/F Kwong Centre</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2/F Student Lobby</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>3/F Student Canteen</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>4/F Student Quiet</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Room &amp; Lobby</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>4/F Student Restaurant</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>5/F Staff Club</td>
<td>1</td>
</tr>
<tr>
<td>Jockey Club Auditorium</td>
<td>Stage</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Podium</td>
<td>1</td>
</tr>
<tr>
<td>Lecture Theatres</td>
<td>FJ 3/F</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Li Ka Shing Tower G/F</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>PQ 3/F</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>ST 1/F/F</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>TU 1/F</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>TU P/F</td>
<td>1</td>
</tr>
<tr>
<td>Chiang ChenStudio Theatre</td>
<td>AG 1/F</td>
<td>1</td>
</tr>
<tr>
<td>Li Ka Shing Tower</td>
<td>2/MF, Student Computer Cluster</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Service Counter</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>16/F Senate Room</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>18/F Council Chamber</td>
<td>1</td>
</tr>
<tr>
<td>Podium Area</td>
<td>A,B,C,D,E,FJ,H,M,PQ,T,L</td>
<td>13</td>
</tr>
</tbody>
</table>

TABLE I  
NUMBER OF ACCESS POINTS(APs) ON EACH BUILDING IN POLYU CAMPUS
declared and undeclared and provide information about these wireless networks. The software performs a periodic scan of local area detecting any wireless signals transmitted on the 802.11b standard. The time interval of these scan can be manually set by the user.

For 27 locations on the floor, shown as small red dots, we collected environmental readings at these locations over two-week period of time. At each testing location, we picked a frequency 2.4GHz and calculated their average amplitude respectively over the recording period. Note that the RSS is the received signal from a beacon packet, while the spectrum energy is the ambient RF energy corresponding to a specific frequency range.

IV. FUZZY MODELING FOR RSS DISTRIBUTION

Fuzzy modeling can be applied to evaluate the properties of wireless received signal strength (RSS). In this section, we describe and evaluate how the preliminary multi-layer fuzzy model can be applied.

A. Propagation-based Approximation

The propagation-based algorithm will be discussed in this section. The following definitions are needed: \( AP = \{ AP_1, ..., AP_n \} \) is a finite set of access points and \( n \) is the total number of APs. \( D = \{ d_1, ..., d_m \} \) is a set of locations where \( D \subset \mathbb{R}^n \), \( R = \{ r_1, ..., r_k \} \) is the received signal strength (RSS) vector of each access point \( AP_i \) respect to \( d_i \). The vector \( r_i \) is the “fingerprint” of the location \( d_i \).

In each access point, \( AP_i \), we take a mean value of RSS, \( r_i \). By using propagation-based theorem [14], we can calculate the RSS by using the follow algorithm:

\[
    r_j(d_j,k) = r_0(d_0) - 10\alpha \log_{10}(d_{j,k}) - \text{wallLoss}
\]  

where \( r_0 \) is the initial RSS at the reference distance of \( d_0 \) is 1 meter (this is 41.5 dBm for line-of-sight propagation (LOS) and for 37.3 dBm non-line-of-sight propagation (NLOS) for some report measurement)[10].

Besides, the variable \( \alpha \) denotes the path loss exponent (clutter density factor), which for indoor location at a carrier frequency of 2.4 GHz is reported to be 2 for LOS propagation and 3.3 for NLOS propagation [14]. Under other circumstances, the indoor path loss exponent \( \alpha \) can be between 1 and 6. \( \text{wallLoss} \) is the sum of the losses introduced by each wall on the line segment drawn at Euclidean distance \( d_{j,k} \). We calculate the \( \text{wallLoss} \) and \( \alpha \) clutter density factor for \( AP_i \). The RSS between each building is approximated by propagation-based theorem.

B. RSS Normalization

In each \( d_i \), we find \( |r_i| = \sqrt{r_1 + r_2 + ... + r_n} \). Let \( R_i = |r_i| \). In order to use the same scale, we need to normalize \( R_i \). The normalization algorithm show as follow,

\[
    x_i = \frac{R_i - R_{\min}}{R_{\max} - R_{\min}}
\]

where \( R_{\max} \) is the maximum overall signal strength, \( R_{\min} \) is the minimum overall signal strength, \( R_i \) is the original signal strength, \( x_{\max} \) is the maximum normalized signal strength and \( x_{\min} \) is the minimum normalized signal strength, \( x_i \) is the value to be normalized.

The range of the signal strengths are normalized from 0 to 1.

\[
    x_i = \frac{R_i - R_{\min}}{R_{\max} - R_{\min}}
\]  

C. Fuzzy Membership Function

The normalization distribution is used to represent the fuzzy membership functions.

![The RSS fuzzy membership graph](image)
where \( p(x) \) is the probability function, \( x \) is the normalized RSS, \( \sigma \) is the standard deviation of normalized signal normalized strength in a region, \( \mu \) is the mean of signal strength in a region. The wireless network is fully covered for the whole campus.

The membership function of term set, \( \mu(\text{SignalStrengthDensity}) = \{\text{Red, Green, Blue}\} \).

The red color means the signal strength density is high, green color means the signal strength is medium and the blue color means the signal strength density is low. The fuzzy set interval of blue is \([0, 0.5]\), \([0, 1]\) is green and \([0.5, 1]\) is red in color.

For the blue region, \( \sigma = 0.5, \mu = 0 \).

\[
\mu_{\text{Blue}}(0 < x < 0.5) = \frac{2}{\sqrt{2\pi}} e^{-2x^2}
\]

For the green region, we substitute \( \sigma = 0.5, \mu = 0.5 \).

\[
\mu_{\text{Green}}(0 < x < 1) = \frac{2}{\sqrt{2\pi}} e^{-2(x-\frac{1}{2})^2}
\]

For the red region, \( \sigma = 0.5, \mu = 1 \).

\[
\mu_{\text{Red}}(0.5 < x < 1) = \frac{2}{\sqrt{2\pi}} e^{-2(x-1)^2}
\]

Figure 3 shows the fuzzy membership function. X-axis represents the normalized signal strength from 0 to 1 (from -93dBm to -15dBm). The width of membership function depends on the standard deviation of the RSS. The overlap area will be displayed with the mixture of the color.

D. Fuzzy Spatio-temporal Cluster

The following mathematical notations are need: \( \mathbf{R} = \{r_1,...,r_k\} \) is a finite set of the RSS vector in the training set. \( \mathbf{B} = \{b_1,...,b_k\} \) is a finite set of the RSS vector belonging to a particular color region, where \( \mathbf{B} \subset \mathbf{R} \) and \( \mathbf{B} \subset [l,u] \), \( u \) is the upper bound of fuzzy interval and \( l \) is the lower bound of fuzzy set interval. \( \mathbf{D} = \{d_1,...,d_a\} \) is a set of locations where \( \mathbf{D} \subset \mathbb{R}^2 \). \( T = \{t_1,...,t_n\} \) is a set of time series. \( \mathcal{A} = \{a_1,...,a_n\} \) is a set of spatio-temporal region(fuzzy color region) where \( \mathcal{A} = \{a|a = td, t \in T, d \in \mathbf{D}\} \). There exists a mapping \( f : \mathbf{B} \rightarrow \mathcal{A} \). The many-to-one mapping will always find between \( \mathbf{B} \) and spatial temporal region. Such that if \( f(d,t) \in \mathbf{B} \). At time interval \( t \in [t_{i-1},t_i] \), \( f(d,t) = f(d,t_{i-1}) \) such that area of fuzzy color region in \( \mathbb{R}^2 \) will be

\[
\int \int_{x=x_1,y=y_1}^{x=x_2,y=y_2} f(x,y,t_{i-1}) dx dy
\]

E. Fuzzy Positioning Approach

The fundamental method of positioning algorithm is the K-Nearest Neighbor (K-NN) algorithm. The Radar project in Microsoft Research was the first to study WLAN location determination. The Rader project used a K-NN algorithm and was able to achieve an accuracy of 2.95 meters with 50% probability. The issue of reducing computational overhead of positioning system has not been addressed in recent research. In this paper, we apply each fuzzy color layer region in figure 4 as the preliminary clusters.

Instead of using the KNN algorithm in every \( r_i \), we estimate the location \( \mathbf{d} \) by given fingerprint vector \( \mathbf{r} \) to find Euclidean distance between \( \mathbf{r} \) and cluster color region \( \mathbf{B}_j \). Determination of finding the nearest \( \mathbf{B}_j \) in the training data set which have the small Euclidean distance \( |r - \mathbf{B}_j| \).

\[
\mathbf{B} = \sum_{i=1}^{n} \frac{1}{|r - \mathbf{B}_i|}
\]

V. RSS Behavior Study by Fuzzy Modeling Approach

In this section, we will experiment the pattern of RSS in PolyU campus. The RSS pattern can be affected the accuracy of the location estimation in the system. We base on the user’s presence factors and the signal strength distribution pattern to analyze the result.

A. Behavior Study on the Human’s presence

Human’s body consists of 70% water. And the water can absorb the signal strength. [10] Different clutter density, such as water will absorb the signal strength more than the air. The effect of the user’s presence can affect the mean of the RSS value, and then it could be affect the accuracy of location estimation in the system. Therefore, we would discuss the effects of user’s presence in this part. As the previous section mentioned, we collected the RSS data in 2 periods, one is in the morning leisure period (7.30am-9.30am) and the other is in the busy evening period (4:30pm - 6:30pm). We would like to observe the difference between two periods.

In figure 7(a) and 7(b), they show the difference RSS pattern which the RSS collect in the two different time slot. We can
Fig. 4. Fuzzy RSS Distribution with the campus floor plan

see that there is significant change of the RSS value. In the busy period, the red color intensity is lower in figure 7(a) & 7(b). In figure 5(a) & 5(b), the red color histogram shows that in leisure period the red color intensity increase from 0.5 to 0.77 in busy to leisure period. This means RSS becomes weak in busy period. We can see that if the location is full of people, then the signal strength will have a great effect. As a result, the accuracy of location estimation is not precise. However, the ideal location is only for the research only.

B. Effect of LOS on RSS

Figure 4 shows the effect of LOS in two major clusters of RSS. The two major centers of high intensity locate at F core and S core respectively. The distance between F core to S core buildings is around 600m apart. In table 1, the locations of APs are equally distributed. The RSS should be covered evenly. Moreover, between M core (Lee Ka Shing Tower) to R core buildings (Shirley Chan Tower), the RSS distribution is relatively low. The heights of two buildings in M core and R core are around 80m and 70m respectively. The distance between M core to R core is around 200m apart. Figure 8 shows from M core to R core building experience low signal strength propagation The RSS transmission path between buildings can be LOS, partial LOS or shadow where NLOS propagation is possible. For LOS conditions, RSS should fit into log-normal distribution. A multi-story building in a campus area will experience lower signal strengths within tall buildings due to the absence of LOS propagation.

VI. DISCUSSION

In this section, we discuss the accuracy issues on the proposed analytical model according to the RSS features. They are: 1) RSS variation, 2) RSS path loss exponent. Lastly, the computational complexity of the proposed solution will be discussed.

A. Effect of the RSS Variation on Accuracy

Accuracy of the tracking system is highly dependable on the RSS variation. If the standard deviation of RSS increases, the accuracy of the tracking system will be lower. Therefore, in order to maintain high accuracy, the suggest standard deviation of RSS should be under than 4dBm. (In other report, the standard deviation assumed as 2.13dBm. [1]) However, the standard deviation depends on the real environment, for example, because the human body can absorb the signal strength, so if the environment is in busy traffic state, the standard deviation will be large.

B. Effect of the RSS Path Loss Exponent on Accuracy

Accuracy of the tracking system is highly dependable on the RSS path loss exponent. The path loss exponent can be varied between 1 and 6 according to the RSS absorbing medium. Path loss exponent represents the attenuation rate of the RSS. When the path loss exponent increases, the accuracy of the tracking system becomes higher. Therefore, high path loss exponent is easy to track the target.

C. Computational Complexity of Positioning Algorithm

Originally, computational complexity of positioning algorithm depends on the total number of sampling RSS vector. We propose to use the fuzzy color layer region as the preliminary clusters. Instead of using the KNN algorithm in every sampling RSS vector, we estimate the location d by given fingerprint vector r to find Euclidean distance between r and cluster...
color region $B_i$. $O(m)$ is the computational complexity of the positioning algorithm, where $m$ is the number of color region. $m$ is much smaller than the total number of sampling RSS vector, such that the proposed equation has lower computational complexity.

VII. Conclusion

In this paper, we provide a spatial analytical model for wireless tracking. The usage of model can increase the efficiency usage of wireless infrastructure. The issue of reducing computational overhead of positioning system has not been addressed in recent research. We use the fuzzy color layer as a cluster, instead of finding every Euclidean distance of Location Fingerprint to reduce the search time. Extracting the useful color RSS region and eliminating unnecessary location fingerprints from a training data set can reduce computational time spent in the online phase for LF based positioning systems significantly. The dynamic spatio-temporal 3D model
can be implemented to visualize the real-time behavior of RSS in the future. In order to achieve the real-time model, wireless sensor network will be applied to sample the RSS, instead of using manual survey.

Fig. 8. RSS Distribution between M to R core buildings

ACKNOWLEDGMENT

The authors would like to thank the Information Technology Service Center in The Hong Kong Polytechnic University to provide information about wireless infrastructure in The Hong Kong Polytechnic University.

REFERENCES


