# WIFI Fingerprinting Indoor Localization System based on Spatio-temporal (S-T) Metrics

Julie Yixuan Zhu, Jialing Xu, Anny Xijia Zheng, Jiaju He, Chaoyi Wu, Victor O.K. Li
Department of Electrical and Electronics Engineering
the University of Hong Kong
Hong Kong
Email:yxzhu@eee.hku.hk

***Abstract*—Indoor localization has greatly leveraged applications regarding to location based service (LBS), which witnessed ever-increasing impact on human life. Among the existing localization solutions, WIFI-based received signal strength index (RSSI) fingerprinting is widely used due to desirable features such as universal availability, privacy protection, and low deployment cost. However, to build a robust, accurate RSSI fingerprinting localization system regardless of application occasions confronts two challenges. The first challenge is to construct a fine-grained and up-to-date RSSI map with reasonable labor cost in the training phase, and the second challenge is to deploy effective algorithm in the localization phase. This article illustrates the design and deployment of our indoor localization system targeting at the above mentioned problems. The overall solution is based on five spatio-temporal (S-T) metrics, to improve localization accuracy. Localization performance is evaluated in three indoor scenes at different scales, which show good accuracy with a median error of 1-2m under office environment, and 3-4m accuracy with no less than 70% probability when the environment is extremely crowded and noisy.**

***Index Terms*—Indoor localization; RSSI; Location based service (LBS); WIFI; S-T metrics.**

## I. Introduction

WIFI-based on RSSI fingerprinting has been widely used due to desirable features such as universal availability, privacy protection, and low deployment cost. Existing works [1], [2] have empirically evaluated the performance of RSSI fingerprinting, inferring its performance better than other WIFI-based method, such as WIFI triangulation. What's more, the limits of localization accuracy are unlikely to be exceeded without better models, due to the environmental change and noises.

Basically, there are two challenges limit the accuracy of indoor localization. The first challenge is to construct a fine-grained and up-to-date RSSI map with reasonable labor cost in the training phase, and the second challenge is to deploy effective algorithm in the localization phase. In this article, an improved localization system is proposed. The training phase is based on an S-T similarity model to achieve a fine-grained and up-to-date RSSI map [3], and location can be achieved by matching live RSSIs with the map.

The proposed full-solution RSSI fingerprinting indoor localization system, shown as Fig. 1, contains three major techniques:

1) An android based smartphone app for localization and RSSI survey.

2) Data management system to interpolate and update a fine-grained RSSI map.

3) A localization engine based on J2EE7 platform.

The merits of our proposed system are two-folds. First, it allows asymmetric computational complexity for the RSSI map training phase and the positioning phase. The RSSI map training phase is relatively complicated yet can be outsourced to high performance computation clusters. Thus both the accuracy and time efciency can be guaranteed for the positioning phase which is relatively simple. Second, taking advantage of the S-T correlation, the RSSI map can be interpolated and updated with fairly high quality and low sampling density. This system is efcient for the volunteer enabled survey where the site survey locations are randomly selected.

The rest of this article is organized as follows. Section II introduces the algorithm in both the training phase and the localization phase. Section III illustrates the system deployment. Section IV evaluates the localization performance in three application scenes at different scales, which show good localization accuracy with a median error of 1-2m. And Section V concludes the article.

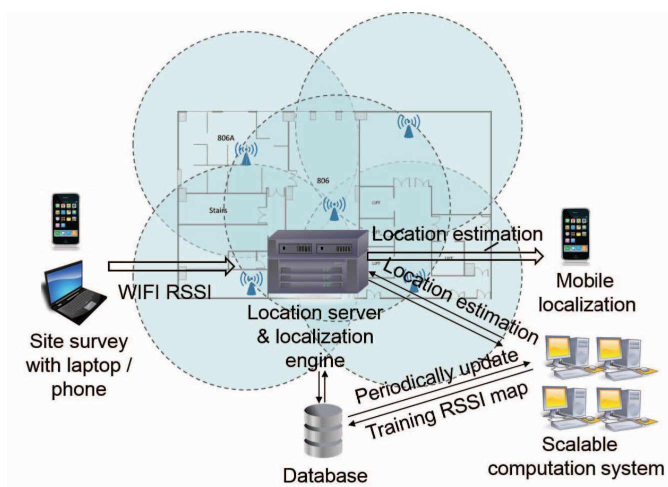


Fig. 1: Overview of the total localization solution.

978-1-4673-8054-6/14/$31.00 ©2014 IEEE 

## II. ALGORITHM DESCRIPTION

### A. S-T based RSSI Map Construction

This section is to introduce the S-T based RSSI map construction method, which is based on our proposed S-T similarity model [3]. Basic idea of this model is to make use of S-T characteristic metrics to interpolate and update a fine-grained RSSI map. In this way, a volunteer-enabled WIFI signatures survey system can be achieved.

We use $S_{i,j,t_k}$ to represent the collected signal strength at cell $i$, from the $j$th AP, at discrete timestamp $t_k,\ where\ i = 1, 2, \ldots, M;\ j = 1, 2, \ldots, N_i;\ k = 1, 2, 3...$

Here, $N_i$ indicates the number of APs detected at cell $i$, and $t_k$ represents the $k$th sampling timestamp. Thus the data stream can be expressed in a three-dimensional data format. We assume the whole tested area is divided to $M$ identical cells, and $l_i$ represents the $i$th cell, i.e., the center of the $i$th cell.

The S-T characteristic metrics for constructing RSSI map are defined as:

*1) Spatial distance:* The spatial distance of location $i_1$ and $i_2$ can be represented as $Spatial\_dist(i_1, i_2) = \parallel l_{i_1} - l_{i_2} \parallel,\ where\ i_1, i_2 \in 1, 2, \ldots, M$

*2) Signal similarity:* This metric is defined to measure the similarity of two specific APs $j_1$ and $j_2$ at the whole testbed area, with timestamp $t_k$ bounded by a time window $[t_{begin}, t_{end}]$ at all the locations can be represented as $Sig\_similarity(j_1, j_2) = \dfrac{\sqrt{\sum_{i=1}^{M} \sum_{k=begin}^{end} \left(S_{i,j_1,t_k} - S_{i,j_2,t_k}\right)^2}}{number\ of\ t_k\ in\ [t_{begin}, t_{end}]},\ where\ i \in 1, 2, \ldots, M;\ j_1 \in 1, 2, \ldots, N_{i_1};\ and\ j_2 \in 1, 2, \ldots, N_{i_2};\ t_k \in [t_{begin}, t_{end}]$

*3) Similarity likelihood:* This metric is to calculate the likelihood of two RSSI distributions located at a specific cell $i$, computed by histogram statistics.

*4) RSSI vector distance:* The RSSI vector distance is a useful metric to differentiate different cells, which is defined as the average Euclidian distance of two RSSI streams at two selected locations $i_1$ and $i_2$, i.e., $RSSI\_vector\_dist(i_1, i_2) = \sqrt{\sum_{j=1}^{N_{i_2}} (\bar{S}_{i_1,j} - \bar{S}_{i_2,j})^2},\ where\ i_1, i_2 \in 1, 2, \ldots, M;\ j \in 1, 2, \ldots, N_{i_2};\ t_{k_1}, t_{k_2} \in [t_{begin}, t_{end}]$

*5) S-T reliability:* In order to filter untrusted RSSI values, we define the metric of reliability based on the S-T correlation of a certain signal, that is, $\mathbf{Re} = (Re_1, ..., Re_j, ..., Re_{N_{i_2}}),\ where\ i_1, i_2 \in 1, 2, \ldots, M; j \in 1, 2, \ldots, N_{i_2}; t \in [t_{begin}, t_{end}]$. The $j$th element $Re_j$ are defined as the S-T reliabiltiy for differentiating two locations, based on the signal from the $j$th AP. In this formula, $i_2$ is the reference point for comparision.

$$Re_j = \begin{cases} Const, & \dfrac{\parallel \bar{S}_{i_1,j} - \bar{S}_{i_2,j} \parallel}{\sigma(i_2, j)} > 90\%\_CI \\ 1, & \text{otherwise} \end{cases}$$

$where\ 90\%\_CI\ represents\ the\ 90\%\ confidence\ interval.$

The training of RSSI map is based on the neighborhood RSSI signatures as well as the five metrics, to interpolate RSSIs at non-sampled locations.

### B. RSSI Fingerprinting

Basically, the methodology used by many previous designs is based on formulating the Euclidean distance or the likelihood as the target function of some selected significant factors and trying to solve the equation as an optimization problem [4]. With our investigation in practice, this principle causes over-sensitivity which is inevitably to the “noise” in real world. The “noise” means the unpredictable cause of the fluctuation of RSSI, including the impacts of changing temperature, humid, the surrounding crowd density, and the influence of the user’s own body and receiver’s antenna direction. In our evaluation of many existing solutions under noisy scenarios, most of them cannot deal with the noise and are impossible to figure out the expected most possible location correctly. Thus our localization algorithm focuses on the “most impossible” case and design filter pipelines ahead of the commonly deployed localization algorithms, such as the Euclidean distance based algorithm and the likelihood based algorithm.

The system utilizes virtual longitude/Latitude coordinate scheme for positioning instead of any other arbitrary 2-D coordinate schemes. Unlike previous solutions, the real dimensions of the site layout are not under consideration. That is what the virtual means, and like in the global coordinate system, the x-axis, the longitude, is set within the range from -180 to +180, while the y-axis, the latitude, is within the range from -90 to +90. In this way, the system can do easy data exchange with many industrial geographic information systems (GIS), such like Open GIS and Arc GIS, and get strong support on data manipulation and visualization from various geographic tools, such as JTS and Openlayers.

To achieve good trade-off between accuracy and complexity when working with various positioning algorithms and strategies, a second coordinate scheme based on a grid model is introduced. In this model, the physical site layout is first sliced into cells, forming a grid array. Each cell is a rectangle area with given dimension, and is identified as [x, y] where x and y indicates its vertical and horizontal orders.

To seek the most impossible candidate cells and further the most possible location, the system make use of pipeline design pattern, called filter chain. In the pipeline, single strategy or algorithm acts as an individual filter, trying to filter off most impossible cells. All the selected filters are queued in a line following specified order. This design is open, and various strategies and algorithms can be plugged in and arranged in optimized order. In the currently algorithm, we choose the following filters pipelined as below:

1) Filter off received WIFI AP Mac addresses which are not in survey database;

2) Filter off the cells where there is most impossible to hear the received WIFI AP RSSI values;

3) Filter off the isolated cells with significant long distance from others;

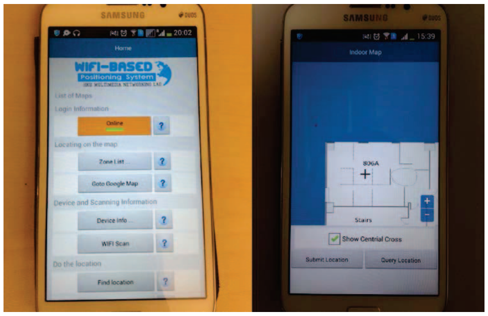

Fig. 2: User interface of Android app.

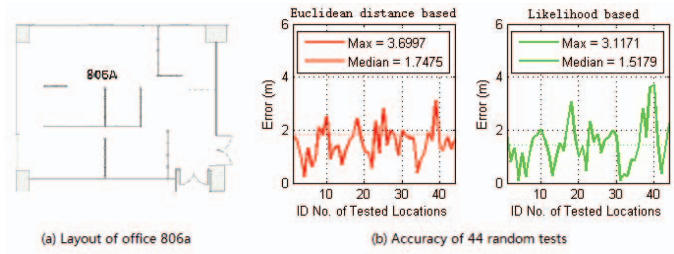


Fig. 4: (a) Floor layout 1 - Chow Yei Ching Building at the Univeristy of Hong Kong (12m*8m). (b) Accuracy test measured by 44 random selected locations.

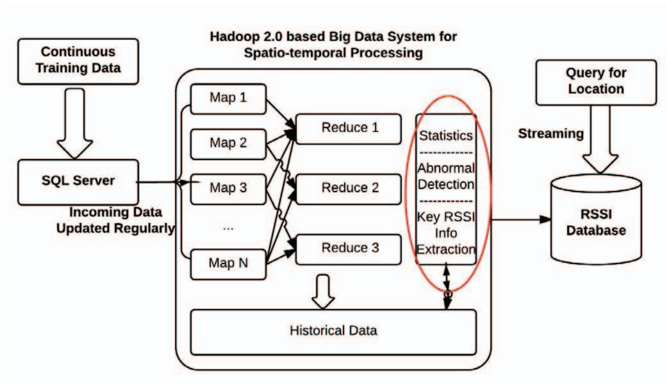


Fig. 3: Design flow of RSSI map training phase.

4) Within the area enclosed by the left candidate cells, figure out the most possible location by 2-D spatial operation on the set of possible coverage area for individual WIFI APs.

## III. System deployment

This section illustrates the overall architecture design of the indoor localization system, with new strategies and features involved.

The system is based on the traditional Client/Server (C/S) paradigm, and includes two parts, a central locating server and a client-side Android app, working together on the RSSI map database construction and the localization service. The server is based upon J2EE stack, and takes the tiered structure (shown as in Fig. 1). The data exchange between the server and the client-side APP take the use of standard Representational state transfer (REST) architecture and JavaScript Object Notation (JSON). By following the industrial standards, third-party client applications can be easily adapted to use the service in future.

On the client side, the design of client-side Android app takes the architecture of separating user interface (UI) + service. The backend service encapsulates the core logic and functions while running at background on user's smartphone, providing unique services when invoked by frontend UI. Unlike the traditional SDK, providing standalone service can help third-party APP vendor focus on UI design and integration. Fig. 2 illustrates the user interface of Android app.

Upon this system structure, the idea of crowdsourcing is adopted in the site survey. Unlike previous solutions, there is no need for operating three separate processes in order: site survey, RSSI map generation, and providing localization service. In our solution, the client-side app undertakes the jobs of both site survey and localization, with which users (ones with specified permission) can submit RSSI signatures to the server anywhere and anytime. The server generates up-to-date signal map incrementally at real time with the continuous arriving survey fingerprints. In this way, the dataset is always updated to reflect the latest site condition without costly complete site survey. Finally, to deal with the computational complexity of our proposed S-T metrics based RSSI map construction, we outsource the RSSI map training phase to a Hadoop-2.0 based cluster system to handle large numbers of incoming RSSI streams based on Mapreduce. The cluster also measures the RSSI statistics, detect the abnormalities, and extract key RSSI information for RSSI map interpolation. Then it sends the well-trained RSSI map back to the database. Fig. 3 shows the design flow of our proposed RSSI map training phase.

## IV. Localization performance in three indoor scenes

We evaluate the feasibility and accuracy of our localization system in three application scenes, at different scales, as shown in Fig. 4-6. The first application scene is the office 806A of Chow Yei Ching Building at the University of Hong Kong, at a 12m*8m scale. With 44 random tests at different locations, our system achieves good accuracy with a median error of 1.51m, and a max error of 3.12m based on the likelihood localization method. And the results based on the likelihood method shows better performance than the Euclidean distance based localization method, which achieves a median error of 1.75m, and a max error of 3.70m.

To verify the universality of our system, we furthermore deploy our system into another two scenarios, i.e., the HK Science Park (40m*20m), and the HKU Centennial Campus (80m*40m). Localization performance is evaluated in these two scenes also show good accuracy with a median error of 1-2m under office/hall environment, and 3-4m accuracy with no less than 70% probability when the environment is extremely crowded and noisy.

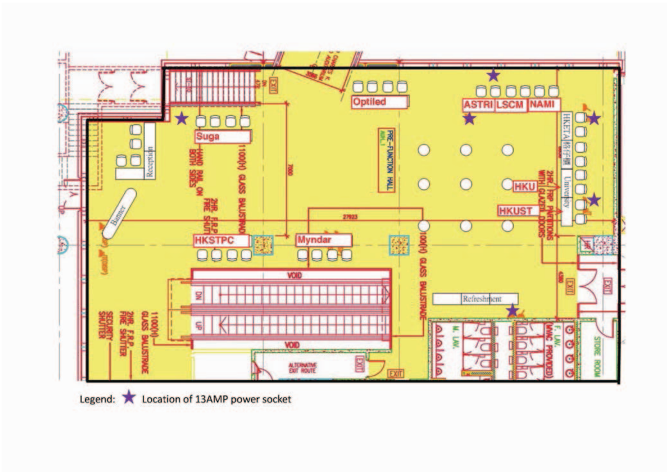


Fig. 5: Floor layout 2 - Exposition Hall at Hong Kong Science Park (40m*20m).

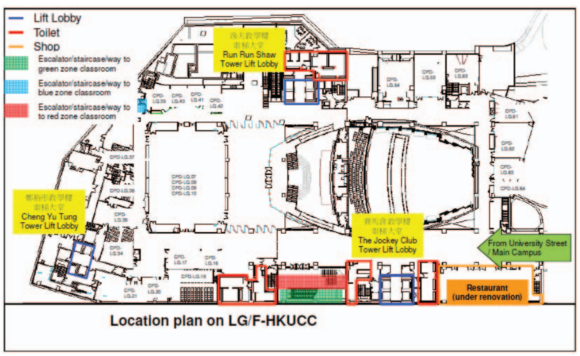


Fig. 6: Floor layout 3 - Centennial Campus at the Univeristy of Hong Kong (80m*40m).

## V. Conclusion and future work

This article gives a systematic description of our total-solution indoor localization system based on S-T metrics, which contains three components: 1) an android based smartphone app for localization and RSSI survey, 2) a data management system to interpolate and update a fine-grained RSSI map, and 3) a positioning engine based on J2EE7 platform. The system focuses on two challenges. The first challenge is to construct a trustworthy RSSI map. Five S-T metrics are adopted in the training phase to achieve a fine-grained and up-to-date RSSI map. The second challenge is to add pipelined filters in the localization phase, to rule out most impossible cells. Evaluation is conducted in three indoor scenes at different scales, which shows 1-2m median error and is bounded by 3-4m max error.

Future work will focus on the real-time efficiency of WIFI fingerprinting. Based on our existing works, we have found the scan of WIFI Mac APs by the smartphone takes a significant proportion of time, which leads to localization latency. And the proposed solution will try to selectively scan the WIFI signatures of the most “important” APs, which requires further research on building models to extract key APs of the indoor environment.

## References

[1] M. Quan, E. Navarro, and B. Peuker, Wi- localization using rssi ngerprinting, 2010.
[2] E. Elnahrawy, X. Li, and R. P. Martin, The limits of localization using signal strength: A comparative study, in Sensor and Ad Hoc Communications and Networks, 2004. IEEE SECON 2004. 2004 First Annual IEEE Communication.
[3] Y. Zhu, A. Zheng, J. Xu, and V.O.K. Li, Spatio-temporal (S-T) Similarity Model for Constructing WIFI-based RSSI Fingerprinting Map for Indoor Localization, submitted to IPIN2014.
[4] V. Honkavirta, T. Perala, S. Ali-Loytty, and R. Piche, A comparative survey of wlan location fingerprinting methods, in Positioning, Navigation and Communication, 2009. WPNC 2009. 6th Workshop on. IEEE, 2009, pp. 243251.