

Spatio-temporal (S-T) Similarity Model for Constructing WIFI-based RSSI Fingerprinting Map for Indoor Localization

Julie Yixuan Zhu*, Anny Xijia Zheng^{†‡}, Jialing Xu* and Victor O.K. Li*

**Department of Electrical and Electronic Engineering, the University of Hong Kong, Hong Kong*
 {yxzhu, jlXu, vli}@eee.hku.hk

[†]*Claude E. Shannon Communication and Network Group, Research Laboratory of Electronics, MIT, USA*
 xjzheng@mit.edu

Abstract—WIFI-based received signal strength indicator (RSSI) fingerprinting is widely used for indoor localization due to desirable features such as universal availability, privacy protection, and low deployment cost. The key of RSSI fingerprinting is to construct a trustworthy RSSI map, which contains the measurements of received access point (AP) signal strengths at different calibration points. Location can be estimated by matching live RSSIs with the RSSI map. However, a fine-grained map requires much labor and time. This calls for developing efficient interpolation and approximation methods. Besides, due to environmental changes, the RSSI map requires periodical updates to guarantee localization accuracy. In this paper, we propose a spatio-temporal (S-T) similarity model which uses the S-T correlation to construct a fine-grained and up-to-date RSSI map. Five S-T correlation metrics are proposed, i.e., the spatial distance, signal similarity, similarity likelihood, RSSI vector distance, and the S-T reliability. This model is evaluated based on experiments in our indoor WIFI positioning system test bed. Results show improvements in both the interpolation accuracy (up to 7%) and localization accuracy (up to 32%), compared to four commonly used RSSI map construction methods, namely, linear interpolation, cubic interpolation, nearest neighbor interpolation, and compressive sensing.

Keywords—*S-T similarity mode; WIFI fingerprinting; RSSI map; Characteristic metrics; Interpolation.*

I. INTRODUCTION

Indoor localization is the key to indoor location based service (LBS), which is estimated to grow from a \$448.6 million market in 2013 to \$2.60 billion in 2018 [1]. Many indoor positioning solutions, such as WIFI fingerprinting, WIFI triangulation, radio frequency identification (RFID), inertial measurement and LED lighting, have been proposed. Among these solutions, WIFI-based RSSI fingerprinting is widely used due to desirable features such as universal availability, privacy protection, and low deployment cost. In addition, its performance has been proved to be superior to another WIFI-based method, namely, the WIFI triangulation [2]. Generally, collected WIFI signatures are sensitive to environmental changes, leading to localization inaccuracy. [3] evaluated the accuracy limits of current fingerprinting algorithms, showing strong evidence that these limitations are fundamental, and unlikely to be overcome without better models. Thus efficient

models with high localization accuracy but low computational complexity are desired, and constructing a trustworthy RSSI map is critical.

RSSI map contains the measurements of received access point (AP) signal strengths at different locations, which are used to estimate the present location of a received RSSI. Statistical metrics, such as the average, variance, signal coverage, etc, are calculated from RSSI fingerprints database and stored in the RSSI map for further matching. However, the RSSI map training phase takes large amounts of labor and time, both for RSSI sampling and for computing. Many researchers have discussed constructing efficient RSSI maps, both for saving costs and improving localization accuracy. For example, [4] deployed the color radiomap interpolation methods, and [5], [6] periodically updated the database with auxiliary automatic equipment. [7] proposed a compressive sensing method for constructing the RSSI map with a small set of sampling since WIFI signals are sparse in DCT domain. [8] achieved pre-classification of the RSSI database by support vector machine (SVM) based training. And [9] observed the RSS distribution has two peaks, and utilized the double-peak Gaussian distribution metrics for RSSI database construction. Yet few researchers focused on the S-T correlation of collected RSSIs, which has the potential of filling up the unsampled locations based on the S-T similarity of neighbors, thus reducing the sampling density. In this paper we propose an S-T similarity model in the training phase, to create a fine-grained, up-to-date RSSI map.

This model is evaluated based on experiments in our hadoop-2.0 based indoor positioning system test bed, periodically updated with volunteer surveyed datasets. Results show improvements in both the interpolation accuracy (up to 7%) and localization accuracy (up to 32%), compared to four commonly used RSSI map construction methods, namely, linear interpolation, cubic interpolation, nearest neighbor interpolation, and compressive sensing. Methods proposed in the positioning phase are based on two traditional WIFI fingerprinting algorithms, that is, deterministic estimation and probabilistic estimation [10], with a median error of 1.2-2.75m in the office environment.

The merits of our proposed model are twofold. First, it allows asymmetric computational complexity for the RSSI map training phase and the positioning phase. The RSSI

[‡]Research performed while the second author was an undergraduate research assistant at the University of Hong Kong.

map training phase is relatively complicated yet can be outsourced to high performance computation clusters. Thus both the accuracy and time efficiency can be guaranteed for the positioning phase which matches live RSSIs with the trained RSSI map. 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 model is efficient for the volunteer-enabled RSSI survey where the locations are randomly selected, compared to traditional interpolation methods and compressive sensing.

The rest of this paper is organized as follows. Section II illustrates the mathematical expression of the model, and defines five characteristic metrics. Section III focuses on the training phase and builds a trustworthy RSSI map based on the S-T similarity model. Section IV describes the positioning phase and WPS deployment. Section V introduces the performance evaluation. Finally, Section VI concludes the paper and suggests future work.

II. BASIC CHARACTERISTIC METRICS FOR THE S-T SIMILARITY MODEL

This section introduces the S-T similarity model and its mathematical expression. We use S_{i,j,t_k} to represent the collected signal strength at calibration point i , from the j th AP, at discrete timestamp t_k , where $i = 1, 2, \dots, M$; $j = 1, 2, \dots, N_i$; $k = 1, 2, 3, \dots$

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 coordinate of Cell i , i.e., the center of Cell i . Thus we can use $\{S_{i,j_1,t_1}, S_{i,j_2,t_1}, \dots, S_{i,j_1,t_2}, S_{i,j_2,t_2}, \dots, S_{i,j_1,t_k}, S_{i,j_2,t_k}, \dots\}$ to represent the RSSI stream from N_i APs at calibration point l_i , flowing into the system with sequential timestamp. Similarly, $\{S_{i_1,j,t_1}, S_{i_2,j,t_1}, \dots, S_{i_1,j,t_2}, S_{i_2,j,t_2}, \dots, S_{i_1,j,t_k}, S_{i_2,j,t_k}, \dots\}$ denotes another RSSI stream, representing the signal strengths for a specific AP at all calibration points. Based on these RSSI streams, we define five characteristic metrics in the S-T similarity model, to classify, observe, and filter RSSI values.

A. Metric 1: Spatial Distance

The spatial distance of location i_1 and i_2 can be represented as:

$$\text{Spatial_dist}(i_1, i_2) = \| l_{i_1} - l_{i_2} \|, \text{ where } i_1, i_2 \in 1, 2, \dots, M$$

B. Metric 2: Signal Similarity

The similarity of signal j_1 and j_2 , with timestamp t_k bounded by a time window $[t_{begin}, t_{end}]$ at all the locations can be represented as:

$$\begin{aligned} & \text{Sig_similarity}(j_1, j_2) \\ &= \frac{\sqrt{\sum_{i=1}^M \sum_{k=begin}^{end} (S_{i,j_1,t_k} - S_{i,j_2,t_k})^2}}{\text{number of } t_k \text{ in } [t_{begin}, t_{end}]}, \text{ where } i \in \end{aligned}$$

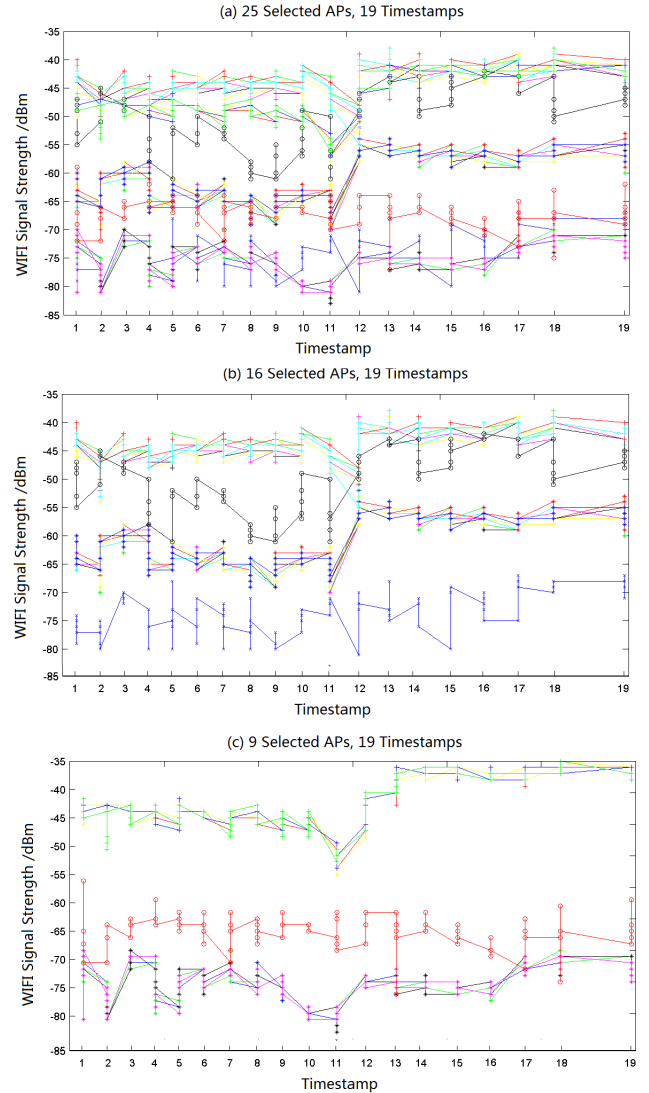


Fig. 1: Classification of signals from 25 APs into 7 clusters. (a) Signals from 25 APs collected with 19 timestamps. (b-c) 7 classified signal clusters. Their temporal distance between sampled RSSIs is not necessarily uniform.

$$1, 2, \dots, M; j_1 \in 1, 2, \dots, N_{i_1}; \text{ and } j_2 \in 1, 2, \dots, N_{i_2}; t_k \in [t_{begin}, t_{end}]$$

In the training phase, signal strength S_{i,j,t_k} at t_k may be comprised of a set of n samplings $\{S_{i,j,t_k}^1, S_{i,j,t_k}^2, \dots, S_{i,j,t_k}^n\}$, where $n = 1, 2, 3, \dots$. In this case S_{i,j,t_k} takes the average value of n samples. This metric is designed to cluster AP signatures which behave similarly or almost the same, thus to reduce duplication. We classify similar signals into the same cluster if the pairwise similarity metrics among a group of signals are within the smallest 10% signal similarity metrics. Fig. 1 illustrates the classification results, with signals from 25 APs categorized into 7 clusters.

C. Metric 3: Similarity Likelihood

For the each signal sequence from the j th AP at the i th calibration point, we utilize the histogram representation of S_{i,j,t_k} and let

$$f_m = \frac{\text{number of } S_{i,j,t_k} \text{ falling into the } m\text{th BIN}}{\text{number of } t_k \text{ in } [t_{begin}, t_{end}]}$$

denote the frequency of RSSI observations failling into the m th BIN among all the samples within the time window $[t_{begin}, t_{end}]$. Here the BINs are m_{max} disjoint RSSI intervals ranging from $[\min_{rssi}, \max_{rssi}]$, $m \in \mathbb{Z}$ and $m \in \{1, 2, \dots, m_{max}\}$. Thus the frequency vector $\mathbf{f} = (f_1, f_2, \dots, f_{m_{max}})$ will represent the RSSI probability distribution, which can be rewritten as the probability mass function (pmf):

$$f_m = p([S_{i,j,t_k}] = \text{center value of } m\text{th BIN} | i, j),$$

where $[S_{i,j,t_k}] = \frac{[S_{i,j,t_k} - \min_{rssi}]}{\text{BIN width}}$, $t_k \in [t_{begin}, t_{end}]$.

To get precise RSSI pmf , large RSSI training datasets are needed. Researchers have also proposed ways to fit the Gaussian-derived distribution function [11] with a relatively small RSSI training datasets by fetching the statistical mean $\bar{S}_{i,j}$ and standard deviation $\bar{\sigma}_{i,j}$.

We define two representation format for metric 3 to measure the similarity likelihood of two RSSI streams, $\{S_{i_1,j_1,t_1}, S_{i_1,j_2,t_1}, \dots, S_{i_1,j_1,t_2}, S_{i_1,j_2,t_2}, \dots, S_{i_1,j_1,t_{k_1}}, S_{i_1,j_2,t_{k_1}}, \dots\}$ and $\{S_{i_2,j_1,t_1}, S_{i_2,j_2,t_1}, \dots, S_{i_2,j_1,t_2}, S_{i_2,j_2,t_2}, \dots, S_{i_2,j_1,t_{k_2}}, S_{i_2,j_2,t_{k_2}}, \dots\}$, collected at location i_1 and i_2 :

$$\text{Likelihood_similarity}(i_1, i_2) = \prod_{j=1}^{N_{i_2}} p([\bar{S}_{i_1,j}] = [\bar{S}_{i_2,j}] | i_2, j), \text{ where } i_1, i_2 \in 1, 2, \dots, M; j \in 1, 2, \dots, N_{i_2}; t_{k_1}, t_{k_2} \in [t_{begin}, t_{end}]$$

$$\text{Log_likelihood_similarity}(i_1, i_2) = \sum_{j=1}^{N_{i_2}} \log(p([\bar{S}_{i_1,j}] = [\bar{S}_{i_2,j}] | i_2, j)), \text{ where } i_1, i_2 \in 1, 2, \dots, M; j \in 1, 2, \dots, N_{i_2}; t_{k_1}, t_{k_2} \in [t_{begin}, t_{end}]$$

Here $p([\bar{S}_{i_2,j}] = [\bar{S}_{i_1,j}] | i_2, j)$ represents the pmf of RSSI stream from the j th AP collected at location i_2 , compared to RSSI at location i_1 . Fig. 2 shows the observed relationship between the similarity likelihood and the spatial distance in terms of a selected calibration point, compared with all other calibration points. Statistical RSSI pmf are used for metric calculation. We observe the largest three values of similarity likelihood corresponding to low spatial distances.

D. Metric 4: RSSI Vector Distance

The spatial RSSI difference of two selected locations i_1 and i_2 can be represented as the average Euclidian distance of two RSSI streams, same as the above section, collected at the corresponding locations, i_1 and i_2 :

$$\text{RSSI_vector_dist}(i_1, i_2) = \sqrt{\sum_{j=1}^{N_{i_2}} (\bar{S}_{i_1,j} - \bar{S}_{i_2,j})^2}, \text{ where } i_1, i_2 \in 1, 2, \dots, M; j \in 1, 2, \dots, N_{i_2}; t_{k_1}, t_{k_2} \in [t_{begin}, t_{end}]$$

The RSSI vector distance is a useful metric to differentiate different calibration points. When comparing this metric in one selected calibration point with all the other points, we observe that the lowest five RSSI vector distance values correspond to low spatial distances. Fig. 3 shows an example.

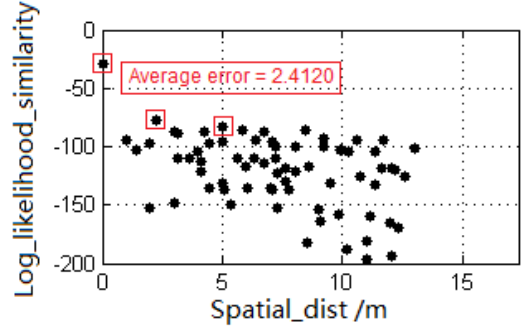


Fig. 2: An example of statistical relationship between $\text{Log_likelihood_similarity}$ and Spatial_dist at a Test Location. X-axis: the spatial distance with all the calibration points, Y-axis: the corresponding $\text{Log_likelihood_similarity}$, based on pmf .

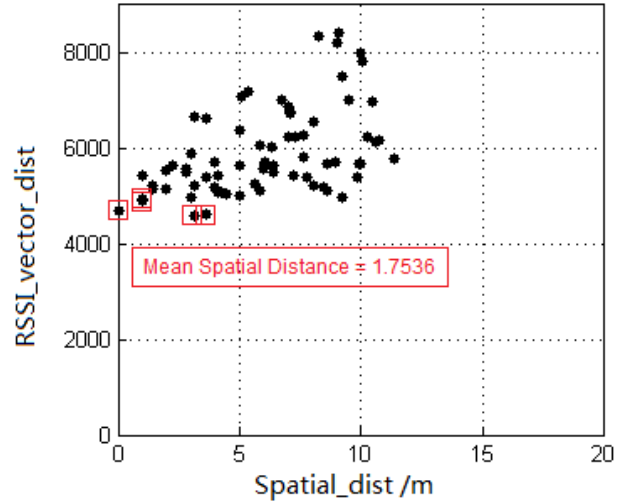


Fig. 3: An example of statistical relationship between RSSI_vector_dist and Spatial_dist at a Test Location. X-axis: the spatial distance with all the calibration points, Y-axis: the corresponding RSSI vector difference.

E. Metric 5: Signal S-T Reliability

Given a time window $[t_{begin}, t_{end}]$, the RSSI stream $\{S_{i_1,j,t_1}, S_{i_2,j,t_1}, \dots, S_{i_1,j,t_2}, S_{i_2,j,t_2}, \dots, S_{i_1,j,t_k}, S_{i_2,j,t_k}, \dots\}$ for a fixed AP can be illustrated as an averaged signal spatial distribution, as shown in Fig. 4. For the spatial distribution of an arbitrary AP, a dot appearing at one position means that this AP can be detected at this position. The sizes of the dots indicate the RSSI values, with larger dots representing stronger signal strengths. And the color is used to classify RSSI values.

In order to filter untrusted RSSI values, we define the metric of reliability based on the S-T correlation of a certain signal:

$$\mathbf{Re} = \text{Sig_st_re}(i_1, i_2) = (Re_1, \dots, Re_j, \dots, Re_{N_{i_2}}), \text{ where } i_1, i_2 \in 1, 2, \dots, M; j \in$$

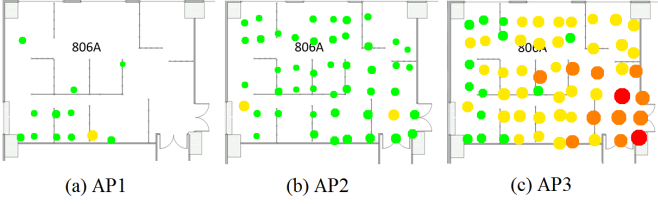


Fig. 4: The spatial distribution of 3 randomly selected APs inside an office.

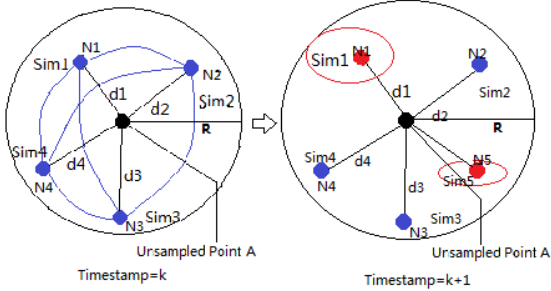


Fig. 5: Illustration of the S-T similarity model.

$1, 2, \dots, N_{i_2}; t \in [t_{begin}, t_{end}]$

The j th element Re_j is defined as the S-T reliability for differentiating two locations, based on the signal from the j th AP. In this formula, i_2 is the reference point for comparison.

$$Re_j = \begin{cases} Constant, & \frac{\|\bar{S}_{i_1,j} - \bar{S}_{i_2,j}\|}{\sigma(i_2,j)} > 90\%_{CI} \\ 1, & \text{otherwise} \end{cases}$$

where $90\%_{CI}$ represents the 90% confidence interval.

Basic idea of this metric is to add a constant weight to the signals falling out of the confidence interval, when comparing the RSSI vector difference and similarity likelihood of two locations. Thus two relatively remote locations have more probability to have higher RSSI vector difference and lower similarity likelihood.

F. S-T Similarity Model

The S-T similarity model for interpolating and updating the RSSI map is illustrated with Fig. 5 as an example. We assume that the state of an unsampled point A can be estimated by its neighbors $N(A, R, t_k) = \{N1, N2, N3, N4\}$ within the distance of R at time t_k . RSSI at A is spatio-temporally correlated with RSSI at each neighbor, for example, $N1$, via two factors: 1) the inverse spatial distance $\frac{1}{Spatial_dist(A, N1)}$, and 2) the similarity factor $Sim1$ between $N1$ and all the other neighbors $\{N(A, R, t_k) - N1\}$. We define the weight of correlation $W(A, N1)$ to be the product of the above mentioned factors:

$$W(A, N1) = \frac{1}{Spatial_dist(A, N1)} * Sim1, \text{ where } Sim1 = \frac{Likelihood_similarity(N1, Nx)}{Rssi_vector_dist(N1, Nx) * Sig_st_re(N1, Nx)}$$

and $Nx \in \{N(A, R, t_k) - N1\}$.

Finally the RSSI at unsampled point A can be represented as:

$$RSSI(A, t_k) = \frac{\sum_{Nx \in N(A)} RSSI(Nx, t_k) * W(A, Nx)}{\sum_{Nx \in N(A)} W(A, Nx)},$$

indicating that the uncertain RSSIs are most influenced by its nearest neighbor which shares the highest similarity with other neighbors.

This model also enables periodic updates of RSSI map once the update of RSSI is detected. Detection includes both the update at a sampled calibration point and at an unsampled calibration point.

III. RSSI MAP TRAINING

Algorithm 1 RSSI Map Creation

Require:

- $RSSI$ map at t_k ;
- k -element input RSSI stream at the same location $\{S_{i_1,j_1,t_1}, S_{i_1,j_2,t_1}, \dots, S_{i_1,j_1,t_2}, S_{i_1,j_2,t_2}, \dots, S_{i_1,j_1,t_{k_1}}, S_{i_1,j_2,t_{k_1}}, \dots\}$;
- k -element input RSSI stream from the same AP $\{S_{i_2,j_1,t_1}, S_{i_2,j_2,t_1}, \dots, S_{i_2,j_1,t_2}, S_{i_2,j_2,t_2}, \dots, S_{i_2,j_1,t_{k_2}}, S_{i_2,j_2,t_{k_2}}, \dots\}$;
- $i \in$ the calibration points $SET\{C\}, j \in 1, 2, \dots, N_i, k = 1, 2, 3, \dots$;
- the unsampled calibration points $SET\{A\}$;
- the interpolated calibration points $SET\{B\}$;

Ensure:

- 1: Classify the RSSI into groups based on the $Sig_similarity$ metric. Represent the classified RSSI stream by $\{S_{i_1,\hat{j}_1,t_1}, S_{i_1,\hat{j}_2,t_1}, \dots, S_{i_1,\hat{j}_1,t_2}, S_{i_1,\hat{j}_2,t_2}, \dots, S_{i_1,\hat{j}_1,t_{k_1}}, S_{i_1,\hat{j}_2,t_{k_1}}, \dots\}$, where $i \in SET\{C\}, j \in 1, 2, \dots, N_i, k = 1, 2, 3, \dots$;
- 2: **for** $timestamp = begin; timestamp \leq end; timestamp++$ **do**
- 3: **if** the new sampled point $i \in SET\{A\}$ **then**
- 4: Put this point in $SET\{C\}$ and drop it from $SET\{A\}, SET\{B\}$.
- 5: **end if**
- 6: Calculate Metric 1-5 defined in Section II for each $i \in SET\{C\}$;
- 7: **while** $SET\{A\}$ is not empty **do**
- 8: Find the neighbors of each unsampled point within a distance of R ;
- 9: Calculate the weight of correlation W between the point and each neighbor;
- 10: Interpolate the RSSI statistics with the S-T similarity model, and put the point into $SET\{B\}$.
- 11: **end while**
- 12: **end for**
- 13: **return** $RSSI$;

This section introduces how to create a trustworthy RSSI map based on the S-T similarity model. The training process faces two challenges. First, the system needs to decide the size of the RSSI streams used for creating the RSSI map, i.e., to choose the time window $[t_{begin}, t_{end}]$. With too short a

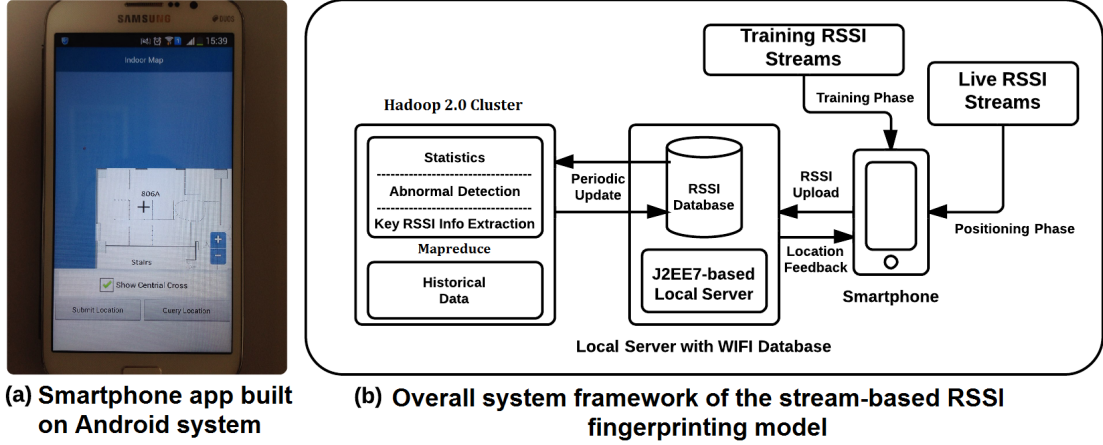


Fig. 6: Deployment framework of current WIFI positioning system.

window, the RSSI characteristic metrics may not be precise enough. But too long a window also causes performance degradation, since the historical RSSI may not reflect the current characteristics of WIFI signals due to environmental changes. Mostly we choose t_{end} to be the most-recent RSSI tuples, which are detected as an RSSI update at a certain calibration point. Then the problem goes to selecting the starting timestamp t_{begin} . Second, for an unsampled calibration point, the range containing its neighbors should be just the right size. Too large a range will increase computational complexity, while too small a range may skip the interpolation of some areas. We iterate and choose the time window and region size parameters which give the best performance.

Algorithm 1 illustrates the process of creating a trustworthy RSSI map, to construct a fine-grained and up-to-date RSSI map. Metrics 1-5 defined in Section II are measured. Performance evaluation of this RSSI map is shown in Section V.

IV. THE POSITIONING PHASE AND SYSTEM DEPLOYMENT

To verify the performance of RSSI map training, we implement the overall localization scheme on our WIFI positioning system (WPS) deployed at Chow Yei Ching Building of the University of Hong Kong. This section gives a brief introduction of the positioning phase based on two traditional methods: deterministic estimation and probabilistic estimation [10], as well as the system deployment.

1) *Deterministic estimation*: Given the measurement vector y and location x , the deterministic method assumes y to be a non-random vector. The main objective is to find the nearest calibration point l_i , i.e., $\hat{x}_{min_dist}(x,i) = l_i$, where $\|y - \tilde{S}_i\|$ gets the minimum among all the available i .

2) *Probabilistic estimation*: The probabilistic estimation assumes the measurement vector y to be a random vector. The idea is to compute conditional pdf $p(x|y)$ of location x give the measurement vector y . By which we can achieve the estimated location with the maximum-likelihood estimation $\hat{x}_{maximum_likelihood} = l_i$.

Fig. 6 illustrates the deployment framework of a Hadoop-2.0 based WPS. It consists of three subsystems, i.e., the smartphone app, the local server system, and a Hadoop-2.0 cluster for periodical update of RSSI database. The smartphone app is built on the Android system, serving as the user interface. It is designed with two functions: 1) submitting RSSI streams to the SQL database and 2) querying locations from the local server. The local server, built on J2EE7-based Java platform, is scalable and flexible to deploy localization algorithm by providing user-customized query APIs. The RSSI database stores the RSSI map and interacts with the smartphone uploaded RSSIs indexed by S-T information. It also interacts with the external Hadoop-2.0 cluster to handle large numbers of incoming RSSI streams based on Mapreduce. In this way the RSSI map can be periodically updated. 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. Corresponding to the RSSI training phase and the positioning phase, proposed S-T similarity model are implemented respectively in the cluster and the local server.

V. PERFORMANCE EVALUATION

A. Experimental Setup

Testing area is located at the 8th floor of the Chow Yei Ching Building at the University of Hong Kong (HKU), with an area of $36 \times 18 m^2$. We divide the area into $1 m^2$ cells based on the layout, with each center representing the location of the cell. RSSIs were collected from all the calibration points where the pedestrians could pass. Performance evaluation of proposed S-T similarity model is at two levels. First, the interpolation of RSSI map accuracy is verified at the sampling percentage of 20%, 40%, 60%, 80%, and 100%, compared to four commonly used methods: linear interpolation, cubic interpolation, nearest neighbor interpolation, and compressive sensing. The sampled datasets at a certain percentage are randomly selected to imitate the case where volunteers collect and update RSSIs. Second, we evaluate the median and maximum localization errors based on the constructed RSSI map to verify the effectiveness of the S-T similarity model.

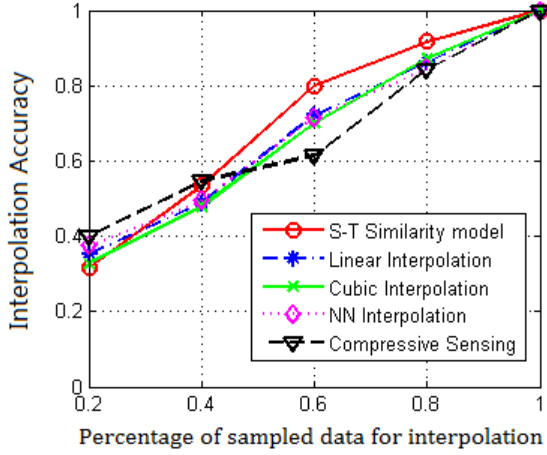


Fig. 7: Comparison of interpolation accuracy at different percentages of sampled data for interpolation.

The difference between our proposed model and the four baseline methods is, we interpolate the RSSI map based on S-T similarities and S-T metrics which are related to localization accuracy. While the other four interpolation methods focus on the linear or non-linear correlations among cells.

Here we give a brief introduction of four other RSSI map construction methods:

1) *Linear Interpolation*: This method fits the regular 2D grid using linear polynomials. The fitting process is to interpolate functions of two variables (eg, x and y), firstly in one direction, and then the other.

2) *Cubic Interpolation*: This method achieves smoother surface than corresponding surfaces obtained by using a third-degree polynomial specified in Hermite form [12].

3) *Nearest Neighbor (NN) Interpolation*: This method approximates the value of a non-given point according to the proximal point.

4) *Compressive Sensing*: This method [13] takes advantage of the sparsity of RSSI in DCT domain, thus reconstructing a fine-grained RSSI map with small set of sampled calibration points.

B. Interpolation Accuracy

Fig. 7 illustrates the interpolation accuracy at different sampling percentages. We randomly select 20%, 40%, 60%, 80%, and 100% of all the surveyed calibration points. Interpolation accuracy is measured by comparing the estimated RSSI values at unsampled points with the real RSSI values. Estimation accuracy for one unsampled point is counted if the estimated RSSI is within a 0.9 confidence level of real RSSI. We choose the 0.9 confidence level instead of the signal-noise-ratio (SNR) since the RSSIs are not stable and fluctuate over time.

It is revealed the interpolation accuracy of our proposed model is superior to four other RSSI map construction methods, when the sampling rate is over 40%. Compressive sensing model gains relatively high accuracy at a 20% sampling rate

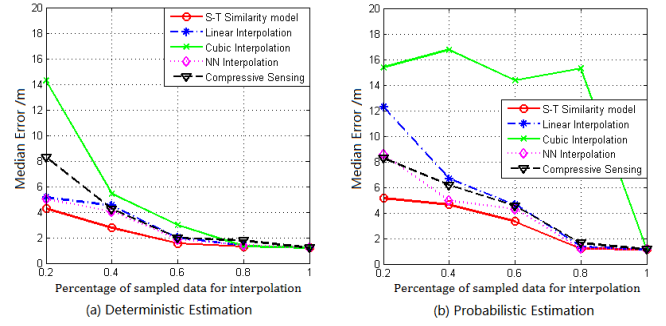


Fig. 8: Comparison of localization accuracy (median error) at different percentages of sampled data for interpolation.

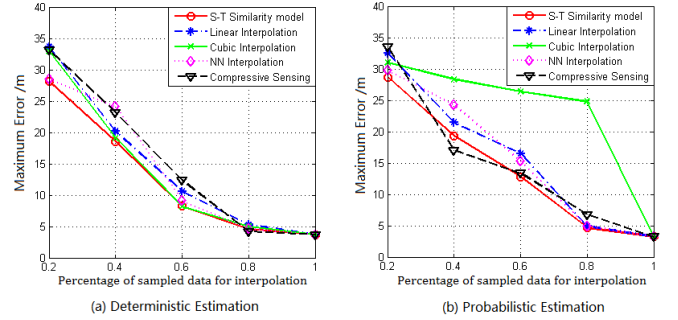


Fig. 9: Comparison of interpolation accuracy (maximum error) at different percentages of sampled data for interpolation.

yet is not quite productive when the sampling rate gets higher. This is because compressive sensing is efficient only with geologically sparse signals. S-T similarity model is more efficient for randomly selected sampling based on the S-T correlation. Statistics show the accuracy to be 79.89% at a 60% sampling. This is compared with a maximum accuracy of 72.25% based on four other methods, indicating an improvement of nearly 7% interpolation accuracy.

C. Impact of Constructed RSSI Map on Positioning Accuracy

To evaluate how the constructed RSSI map helps the localization, we compare positioning accuracy between the S-T similarity model and four other methods.

Fig. 8 shows the median localization error under the deterministic estimation and the probabilistic estimation. Results show a decend trend of median error when the sampling rate increases, with our proposed model corresponding to the lowest errors. For the deterministic estimation, we achieve a median error of 1.2-2.75m when the sampling rate is over 40%. This is about a 32% improvement comparing to a minimum median error of 1.2-4.02m achieved by other RSSI construction methods. The probabilistic estimation results show relatively higher median errors when the sampling rate is below 60%, for the original data is not enough to obtain an accurate pmf. However, when the sampling rate is higher than 80%, the median error tends to be smaller compared to deterministic estimation, reaching a minimum median error of 1.15m. We note that there is an exception by using the cubic interpolation, where the positioning accuracy is not accurate at all. This

is because the smoother function changes the pmf in the interpolation phase.

Fig. 9 shows the maximum localization error under the deterministic estimation and the probabilistic estimation, which acts as the upper bound of localization accuracy. Based on the S-T model, we can achieve a maximum error within 4.5m by deterministic estimation and 4.6m by probabilistic estimation, when the sampling rate is higher than 80%. The maximum error performance under the sampling rate of 60% is not quite good, since the amount of sampled datasets is not enough for generating pmf function, and thus any wrong interpolation will lead to a large maximum localization error.

VI. CONCLUSION AND FUTURE WORKS

In this paper we propose an S-T similarity model for WIFI-based RSSI fingerprinting. The model defines five S-T correlation metrics for construct a trustworthy RSSI map with the RSSI streams. We employ the model in the training phase of RSSI fingerprinting, to create a fine-grained and up-to-date RSSI map. The model is deployed in our WIFI positioning system, and both the interpolation accuracy and positioning accuracy are evaluated. We observe improvements in both the interpolation accuracy (up to 7%) and localization accuracy (up to 32%), compared to four commonly used RSSI map construction methods.

The future work will continue to focus on the real-time update of the RSSI map, enabled with streaming technologies, to handle vast number of data streams. MapReduce tasks, as well as complex event processing will be executed in the computing cluster to enhance the real-time processing capability.

REFERENCES

- [1] "Research report: Global advancements, market forecasts and analysis (2013 - 2018)," MarketsandMarkets.
- [2] M. Quan, E. Navarro, and B. Peuker, "Wi-fi localization using rssi fingerprinting," California Polytechnic State University 1 Grand Avenue, San Luis Obispo, CA, Tech. Rep.
- [3] 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 Communications Society Conference on*. IEEE, 2004, pp. 406–414.
- [4] K. Arai and H. Tolle, "Color radiomap interpolation for efficient fingerprint wifi-based indoor location estimation," *International Journal*, 2013.
- [5] R. Palaniappan, P. Mirowski, T. K. Ho, H. Steck, P. Whiting, and M. MacDonald, "Autonomous rf surveying robot for indoor localization and tracking," in *International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2011.
- [6] M. Gunawan, B. Li, T. Gallagher, A. G. Dempster, and G. Retscher, "A new method to generate and maintain a wifi fingerprinting database automatically by using rfid," in *Indoor Positioning and Indoor Navigation (IPIN), 2012 International Conference on*. IEEE, 2012, pp. 1–6.

- [7] D. Milioris, G. Tzagkarakis, A. Papakonstantinou, M. Papadopouli, and P. Tsakalides, "Low-dimensional signal-strength fingerprint-based positioning in wireless lans," *Ad Hoc Networks*, vol. 12, pp. 100–114, 2014.
- [8] J.-C. Chang, C.-C. Shen, A.-C. Chang, and Y.-C. Chung, "Indoor lbs based on svm and rssi method," in *Proceedings of BAI Conference*, 2011.
- [9] L. Chen, B. Li, K. Zhao, C. Rizos, and Z. Zheng, "An improved algorithm to generate a wi-fi fingerprint database for indoor positioning," *Sensors*, vol. 13, no. 8, pp. 11 085–11 096, 2013.
- [10] V. Honkavirta, T. Perala, S. Ali-Loytty, and R. Piché, "A comparative survey of wlan location fingerprinting methods," in *Positioning, Navigation and Communication, 2009. WPNC 2009. 6th Workshop on*. IEEE, 2009, pp. 243–251.
- [11] M. Klepal, D. Pesch *et al.*, "Influence of predicted and measured fingerprint on the accuracy of rssi-based indoor location systems," in *Positioning, Navigation and Communication, 2007. WPNC'07. 4th Workshop on*. IEEE, 2007, pp. 145–151.
- [12] C. Withers, "A simple expression for the multivariate hermite polynomials," *Statistics & probability letters*, vol. 47, no. 2, pp. 165–169, 2000.
- [13] R. Baraniuk, "Compressive sensing," *IEEE signal processing magazine*, vol. 24, no. 4, 2007.