Sectjunction: Wi-Fi Indoor Localization based on Junction of Signal Sectors

Suining He   S.-H. Gary Chan

Department of Computer Science and Engineering

The Hong Kong University of Science and Technology

Email: {sheaa, gchan}@cse.ust.hk

Abstract—In Wi-Fi fingerprint localization, a target sends its measured Received Signal Strength Indicator (RSSI) of access points (APs) to a server for its position estimation. Traditionally, the server estimates the target position by matching the RSSI with the fingerprints stored in database. Due to signal measurement uncertainty, this matching process often leads to a geographically dispersed set of reference points, resulting in unsatisfactory estimation accuracy.

We propose a novel, efficient and highly accurate localization scheme termed *Sectjunction* which does not lead to a dispersed set of neighbors. For each selected AP, Sectjunction sectorizes its coverage area according to discrete signal levels, hence achieving robustness against measurement uncertainty. Based on the received AP RSSI, the target can then be mapped to the sector where it is likely to be. To further enhance its computational efficiency, Sectjunction partitions the site into multiple area clusters to narrow the search space. Through convex optimization, the target is localized based on the cluster and the junction of the sectors it is within. We have implemented Sectjunction, and our extensive experiments show that it significantly outperforms recent schemes with much lower estimation error.

Keywords-Indoor localization; Wi-Fi fingerprint; clustering; sectoring; convex optimization.

I. INTRODUCTION

Indoor Location-Based Services (LBS) has attracted much attention in recent years due to its potential social and commercial values. The quality of such service largely depends on the localization accuracy of mobile devices.

Many sensor signals have been explored for indoor localization, e.g., Wi-Fi [1], FM signal [2], RFID [3], etc. Among all these, Wi-Fi fingerprinting emerges as a promising approach because it is easy to deploy and requires no extra sensor infrastructure beyond the existing Wi-Fi network.

In Wi-Fi fingerprinting, localization is usually conducted in two phases: an off-line "survey" phase followed by an on-line "query" phase. In the off-line phase, a site survey is conducted to collect the vectors of received signal strength indicator (*RSSI*) of the Wi-Fi access points (APs) at many given reference points (RPs) of known locations. The vectors of RSSI form the *fingerprints* of the site and are stored at a database for the on-line phase. In the on-line phase, a user (or a *target*) samples an RSSI vector at his own position and reports

it to the server. The server compares the received signal vector with the stored fingerprints in the database. By employing some matching algorithms according to some similarity metric in the signal space, the server then finds a set of RPs with fingerprints closely matched with the target's one (using, for examples, the nearest neighbor search or k-nearest neighbors search). These RPs are termed *neighbors* to the target. The server then estimates the target position based on them.

Due to statistical fluctuation of signal strengths, there is uncertainty in signal measurement. Therefore, the matching algorithm in the on-line phase may find a dispersed set of neighbors which are quite distant apart in the physical space. Because of that, localization error is inevitable. It has been observed that in many sites especially spacious indoor environment, Wi-Fi fingerprint-based localization can have error as much as 8 to 15 m [4].

In order to reduce estimation error, a close set of neighbors should be identified with high accuracy in the on-line phase. To achieve it, a better approach is to confine the target to a *sector* for each of its received AP RSSI. We can effectively identify the target location based on the overlap region of all these signal sectors. This is clearly far better than the traditional matching approach of comparing isolated fingerprints of reference points independently.

Based on the above observation, we propose Sectjunction, a novel, efficient and highly accurate indoor localization technique based on *junction* of received signal strength (RSS) sectors. Sectjunction does not lead to a dispersed set of nearest neighbors, and is highly robust to measurement uncertainty due to its use of sectors of discrete signal levels.

In Sectjunction, the site is sectorized according to the signal levels of APs. For an AP, its signal level generally decreases with distance. We can hence find the Wi-Fi *peak*, the reference point corresponding to the maximum signal of the AP. Given the measured RSS of the AP, we can then identify the sector with the Wi-Fi peak as the center where the target is likely to be. As each of the measured RSS values corresponds to a discrete signal sector, the location of a target can then be constrained at the *junction* of these sectors.

We illustrate Sectjunction in Figure 1 given three APs. Each of the sectors corresponds to a certain signal range, and is the region where the target is likely to be for that AP. By considering the junction of the sectors, Sectjunction tightens
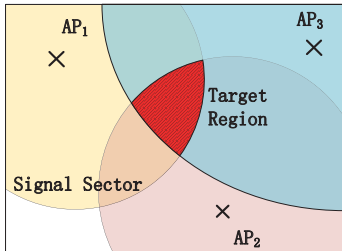
2611

Fig. 1. Illustration of signal sector inter-section. Crosses represent the Wi-Fi peaks. The overlapped (red) region denotes the constrained area.
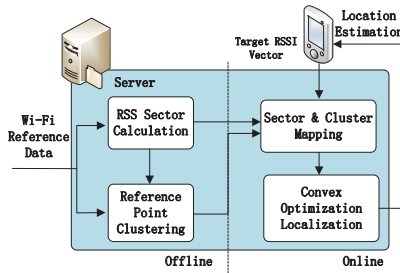
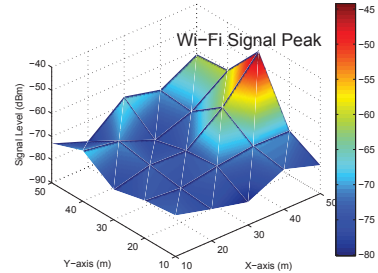Fig. 2. Overview of Sectjunction localization system.

Fig. 3. A Wi-Fi AP's site survey data in HKUST atrium. The peak of the signal map represents the reference point with the highest signal level.

the search space without leading to a dispersed set of nearest reference points.

Sectjunction achieves location efficiency and accuracy by a number of novel approaches:

- *RSS Sector Calculation:* For each AP, the area is sectorized based on the RSS range. This greatly improves the robustness against measurement uncertainty. In the sectorization process, clearly the APs with large signal dynamic range in the whole site are preferred. For computational efficiency, Sectjunction hence filters out the APs with narrow dynamic range making use of signal entropy. Using the filtered APs, the sectors are then calculated based on discrete signal levels with their Wi-Fi peaks as the centers.

- *Reference Point Clustering:* The search space in finger-print database can be large, consisting of large quantities of reference points (RPs). To narrow the search space for computational efficiency, Sectjunction further partitions the site into clusters consisting of similar signal vectors using the k-means algorithm. To efficiently identify the cluster where the target is, Sectjunction elects some exemplars in each cluster to represent each cluster. Given a target vector, Sectjunction then maps it to one of the clusters using these exemplars. The cluster serves as a constraint in the localization process.

- *Localization with Convex Optimization:* Given a target RSSI, the server first computes the sectors and the cluster where the target is in. Using a novel signal difference metric comparing signals with measurement uncertainty, Sectjunction then localizes the target by formulating a convex optimization problem to maximize fingerprint matching given the target's sectors and cluster.

Note that both the RSS sector calculation and RP clustering can be done off-line, while the mapping of target to sectors and cluster, and the convex optimization are done on-line. We have implemented Sectjunction as a real system and conducted large-scale experimental trials in our HKUST campus. Our experimental results further confirm the strengths of Sectjunction with substantial error reduction (typically by more than 30% in our results).

We briefly review related work here. Much of the previous work treats the RSSI signals together as a single vector and use some pattern matching technique on these vectors in localization. The work in [5] first stores in database the sequence of Wi-Fi RSS vectors while the user is travelling as patterns along different corridors in an indoor environment. Then the target compares its online measured RSS sequence with them and finds the most similar one matching its trajectory. In the work of [6], the order of different AP RSSIs at RPs is considered as matching pattern. The work in [7] first finds the unique RSSI vectors in the area as landmarks. The location of a target is then mapped to the RPs with landmarks most similar to the sampled RSSI. It is clear that, as pattern matching is applied on the vector of RSSI signals, these approaches are susceptible to measurement uncertainty and hence estimation inaccuracy. Sectjunction, on the other hand, treats the *individual* AP signal independently in forming a sector for each one. By considering the junction of these sectors, much higher accuracy can be achieved. In the work of [8] [9], Euclidean distance is used to compute the similarity between the RSSI values of observed target and reference points. The estimated location is then based on the most similar RP or a weighted sum of k nearest neighbors. We use a markedly different approach to compute signal difference which effectively captures the impact of measurement uncertainty.

The rest of this paper is organized as follows. After describing the system design of Sectjunction in Section II, we discuss in Section III how RSS sectors are calculated in order to constrain the target region for localization. In Section IV, we present how to cluster RPs to narrow the search space. We formulate the convex optimization for localization with a novel comparison metric of random signals in Section V. Illustrative results on our experimental trials are presented in Section VI. We conclude in Section VII.

II. SYSTEM FRAMEWORK

We show in Figure 2 the server setup of Sectjunction. In the offline phase, Wi-Fi reference data in terms of pairs <*reference point location*, *RSSI vector*> and version information is stored in a database (RSSI is in dBm). The module RSS Sector Calculation filters out the APs with narrow signal dynamic ranges and sectorizes the area after that. Then the server

conducts reference point clustering to partition the site into smaller regions and elects a number of exemplars for each cluster. After the two modules, the system is prepared for online estimation.

In the online phase, a target's RSSI vector is first mapped to a reference point cluster and a set of RSS sectors. Using a novel comparison metric for random signals as its objective, Sectjunction then estimates the target location using convex optimization with the cluster and sectors as constraints.

III. RSS Sector Calculation

In RSS sector calculation, for each AP the area is sectorized corresponding to different discrete signal levels. Clearly, the AP of wider signal range in the site is preferred because this leads to many sectors and hence higher localization accuracy. To achieve this, Sectjunction first filters away those APs of small signal range using entropy (Section III-A). Using the reduced set of APs, sectors are then formed by first finding the Wi-Fi peak for the AP in the area (Section III-B). With the Wi-Fi peak as the center, Sectjunction then calculates the radius of the sector for online mapping (Section III-C).

A. AP Filtering Based on Entropy

In the survey stage, we measure different Wi-Fi signal levels at different reference points. For all the measured signals in the site, some APs may have large signal range and various measurement values. These APs can generate more sectors given the same signal strength interval. These are the APs needed in localization estimation. To find them, Sectjunction measures each AP's entropy and keeps those with high entropy [10] for signal sectoring. The basic idea is as follows.

We first discretize the signal range of each AP into fixed intervals. Given an AP l, denote $RSSI_l$ as the set of measurement values in the whole survey site. Let $\min_l(RSSI_l)$ and $\max_l(RSSI_l)$ be the minimum and maximum signal value of AP l, respectively. The width of an interval, denoted as w_b, is defined as $w_b = (\max_l(RSSI_l) - \min_l(RSSI_l))/n_b$, where n_b is the number of intervals used. For each sector i of AP l, we use a fixed interval, i.e., $[\min_l(RSSI_l) + (i-1)w_b, \min_l(RSSI_l) + iw_b], i \in \{1, \ldots, n_b\}$. Let ψ_{jl} be AP l's signal value at RP j. Interval i corresponds to the reference points whose measurement value ψ_{jl} from AP l satisfies $\min_l(RSSI_l) + (i-1)w_b \le \psi_{jl} < \min_l(RSSI_l) + iw_b$.

We then count the number of RPs within each interval and calculate the entropy of each AP as follows. Let N^l be the number of reference points that can detect AP l. Denote the number of RPs whose signal measurement is within the interval i as N_i^l. Then the probability that an RP measures AP $l's$ signal value within interval i is given by $P_i = N_i^l/N^l$. After the signal range discretization of AP l, the entropy of AP l is given by

$$Entropy(l) = -\sum_{i=1}^{n_b} (P_i \cdot \log_2(P_i))$$
$$= -\sum_{i=1}^{n_b} \left(\frac{N_i^l}{N^l} \log_2 \left(\frac{N_i^l}{N^l} \right) \right). \qquad (1)$$

Using this definition, the AP with high entropy has large signal range and there are similar number of RPs within each interval. Therefore, Sectjunction filters out other APs with low entropy and generates a set of APs for sectors calculation. We denote the resultant set of APs as Λ.

B. Wi-Fi Peak as Sector Center

Sectjunction keeps the APs with high entropy. For each of them, there is a center for all its sectors (which correspond to different signal levels). The center is the RP with the highest signal level in the area, and is called the *Wi-Fi peak* of the AP.

For each selected AP $l \in \Lambda$, its Wi-Fi peak, denoted as WP_l, is represented by the following three tuples:

$$WP_l \triangleq \{MAC_l, \mathbf{x}_l, RSSI_l\}, l \in \Lambda, \qquad (2)$$

where MAC_l is its Wi-Fi MAC address, $\mathbf{x}_l$ is the peak location and $RSSI_l$ represents all its measurement values. Wi-Fi AP MAC address uniquely identifies an AP. Sometimes the RPs with maximum signal level may not be unique. In such a case, we use the centroid of these points as $\mathbf{x}_l$. In Figure 3 we show the Wi-Fi peak of a selected AP as measured in the HKUST atrium. Its high peak and large signal range can help differentiate the RPs and discretize more sectors for Sectjunction.

C. Sector Radius Calculation and Mapping

Based on the selected APs, we can generate the sector radius, i.e., the maximum distance between reference points and the Wi-Fi peak for a certain signal level. Let R_i be the set of RPs whose measurements are in interval i. Let δ_{il} be the radius corresponding to signal level of interval i for AP l. Mathematically, the radius is defined as the maximum distance between the RPs $\mathbf{r}_j \in R_i$ and the Wi-Fi peak of AP_l, i.e.,

$$\delta_{il} = \max\{d_j | d_j = \|\mathbf{r}_j - \mathbf{x}_l\|_2, \mathbf{r}_j \in R_i,$$
$$\min_l(RSSI_l) + (i-1)w_b \le \psi_{jl} < \min_l(RSSI_l) + iw_b\}.$$

Given the above, a target signal is mapped to a sector in a straightforward manner. If a target's measurement ϕ_l from AP l is within the interval i, i.e., $\min_l(RSSI_l) + (i-1)w_b \le \phi_l < \min_l(RSSI_l) + iw_b$, the corresponding sector constraint d_l equals to δ_{il}.

Using maximum distance constrains the target within a sector, and hence ensures junction of multiple RSS sectors. For an AP with high entropy, the larger set of radii results in tighter area sector and hence higher localization accuracy.

IV. Reference Point Clustering

The search space in fingerprint database usually consists of a large quantity of reference points, while the target location is estimated based on only a few of them. To filter out most of the RPs efficiently, we implement k-means clustering method to partition the site into small areas (Section IV-A). Then we elect some exemplars in each cluster. Given a target vector, we map it first to the cluster which has the most similar exemplars (Section IV-B) to narrow the search scope to a smaller region.

A. K-Means Clustering

After Wi-Fi fingerprinting, we implement k-means clustering method on RSSI vectors to partition the whole survey site into smaller regions. Physically close RPs usually have similar vectors and therefore are clustered into the same group.

The clustering process is as follows. After the filtering, we have an overall L selected APs, i.e., $L = |\Lambda|$. Let the average RSSI of multiple samples from AP l at RP i be $\overline{\psi}_{il}$. Then the Wi-Fi RSSI vector at RP i is $\mathbf{q}_i = [\overline{\psi}_{i1}, \overline{\psi}_{i2}, \ldots, \overline{\psi}_{iL}], i \in \{1, 2, \ldots, n_r\}$.

We evaluate similarity between two signal vectors by cosine metric [11]. The cosine value of the angle between two RSSI vectors $\mathbf{q}_i$ and $\mathbf{q}_j$ is defined as follows:

$$
\begin{aligned}
sim(\mathbf{q}_i, \mathbf{q}_j) &= \frac{\mathbf{q}_i \cdot \mathbf{q}_j}{\|\mathbf{q}_i\| \cdot \|\mathbf{q}_j\|} \\
&= \frac{\sum_{l=1}^{L} \overline{\psi}_{il} \cdot \overline{\psi}_{jl}}{\sqrt{\sum_{l=1}^{L} \overline{\psi}_{il}^2} \cdot \sqrt{\sum_{l=1}^{L} \overline{\psi}_{jl}^2}}.
\end{aligned}
\tag{3}
$$

Denote the cluster m as C_m. We implement the k-means clustering in order to minimize the within-cluster sum of dissimilarity of K clusters, i.e.,

$$
\min_{C_m} \sum_{m=1}^{K} \sum_{\mathbf{q}_j \in C_m} (1 - sim(\mathbf{q}_j, \boldsymbol{\mu}_m)),
\tag{4}
$$

where $\boldsymbol{\mu}_m$ represents the mean of the vectors within the cluster m. All RPs are then partitioned into several smaller regions.

B. Election of Exemplars and Cluster Mapping

Based on the reference points clusters, we describe below how to map the target to a cluster given an RSSI measurement.

One naive way of mapping is to add target vector into the set of RSSI vectors and do the clustering again. However, this is not computationally efficient. To improve the mapping efficiency, we first elect the exemplars within each RP cluster, i.e., the RPs which have the highest mean similarity with all other members in the same cluster. We define the exemplar RSSI vector $\mathbf{q}_j$ of the cluster m as follows:

$$
\arg\max_{\mathbf{q}_j} \frac{1}{|C_m| - 1} \sum_{s \in C_m, s \neq j} sim(\mathbf{q}_j, \mathbf{q}_s).
\tag{5}
$$

To improve mapping accuracy, we use several exemplars in one cluster. And all clusters have equal number of exemplars for comparison. Denote the set of exemplars of cluster m as Q_m. We compare the target signal vector with each exemplar signal vector with cosine similarity in Equation (3). By summing up the similarity within each set of reference points' exemplars, we map the target with RSSI vector $\mathbf{p}$ to the cluster whose exemplars are the most similar, i.e.,

$$
\arg\max_{C_m} \sum_{j \in Q_m} sim(\mathbf{p}, \mathbf{q}_j).
\tag{6}
$$

In this way we can filter out other clusters efficiently and focus on a single one (small region) for localization.

V. CONVEX OPTIMIZATION FOR INDOOR LOCALIZATION

In this section, we present a convex optimization formulation for indoor localization, by *jointly* considering the sectors and cluster the target is at. We first present an objective function taking into account the impact of signal measurement uncertainty, followed by the convex optimization formulation.

A. An Objective Function to Compare Random Signals

Much of previous works have not sufficiently considered the measurement uncertainty for random signals. In order to properly capture such uncertainty for accurate indoor localization, we introduce a comparison metric which takes into account the signal fluctuation (or variance).

In the survey stage, multiple samples of Wi-Fi RSS are collected at each reference point. Denote the RSS at RP i from AP l at time τ as $\{\psi_{il}(\tau), \tau = 1, \ldots, S, S > 1\}$, with S being the total number of samples collected at the RP. Then the average RSS readings over time domain from AP l at RP i is $\overline{\psi}_{il} = \left(\sum_{\tau=1}^{S} \psi_{il}(\tau)\right)/S$. The unbiased estimate of variance in RSS time samples for AP l at RP i is $\sigma_{il}^2 = \left(\sum_{\tau=1}^{S} \left(\psi_{il}(\tau) - \overline{\psi}_{il}\right)^2\right)/(S-1)$.

We consider that the received signal strength ψ_{il} of AP l at RP i follows a certain (unknown) distribution with mean $\overline{\psi}_{il}$ and standard deviation σ_{il}. The Wi-Fi RSSI vector at RP i is $\mathbf{q}_i = [\overline{\psi}_{i1}, \overline{\psi}_{i2}, \ldots, \overline{\psi}_{iL}]$. Denote the RSSI of AP l at the target as ϕ_l and the corresponding vector as $\mathbf{p} = [\phi_1, \phi_2, \ldots, \phi_L]$. Denote the set of APs shared by the target and the reference point i's data as J_i ($0 < |J_i| \leq L$). Given a target's Wi-Fi measurement ϕ_l (constant) from AP $l \in J_i$, the expected signal difference between RP i and a given target is defined as:

$$
\begin{aligned}
\Delta_l(\mathbf{p}, \mathbf{q}_i) &= E\left((\phi_l - \psi_{il})^2\right) \\
&= E\left(\phi_l^2 - 2\phi_l\psi_{il} + \psi_{il}^2\right) \\
&= \phi_l^2 - 2\phi_l E(\psi_{il}) + E(\psi_{il}^2) \\
&= \phi_l^2 - 2\phi_l E(\psi_{il}) + E^2(\psi_{il}) + \sigma_{il}^2 \\
&= \left(\phi_l - \overline{\psi}_{il}\right)^2 + \sigma_{il}^2.
\end{aligned}
\tag{7}
$$

Thus the overall expected signal difference between the reference point i and the target is given by

$$
\begin{aligned}
\Delta(\mathbf{p}, \mathbf{q}_i) &= \frac{1}{|J_i|} \sum_{l=1}^{|J_i|} \Delta_l(\mathbf{p}, \mathbf{q}_i) \\
&= \frac{1}{|J_i|} \sum_{l=1}^{|J_i|} \left(\left(\phi_l - \overline{\psi}_{il}\right)^2 + \sigma_{il}^2\right).
\end{aligned}
\tag{8}
$$

If there is no shared APs between the target and an RP, this RP will be excluded in the final localization.

B. Convex Optimization Formulation

Let $\mathbf{r}_i$ be the position of corresponding RP i. After cluster mapping, the search space of the target is narrowed to cluster C_m with n_m reference points, i.e., $C_m = \{\mathbf{r}_i | i = 1, \ldots, n_m\}$.

We assign normalized weight to each RP in the cluster to represent the RP's contribution in final estimation. Then we formulate the estimation problem of the target into finding the weighted combination of RPs in C_m. Let ω_i be the weight assigned to RP i when locating the target, and the estimated position can be represented as

$$\hat{\mathbf{x}} = \sum_{i=1}^{n_m} \omega_i \cdot \mathbf{r}_i, \mathbf{r}_i \in C_m, \qquad (9)$$

where $\sum_{i=1}^{n_m} \omega_i = 1, 0 \le \omega_i \le 1$. We illustrate as follows how we calculate the weights.

First, the weights are constrained by the sector constraints, which are obtained in Section III-C. Let d_l be the predicted sector radius constraint between target estimation position $\hat{\mathbf{x}}$ and AP l's signal peak $\mathbf{x}_l$. We can relax their distance constraints as $\|\hat{\mathbf{x}} - \mathbf{x}_l\|_2 - d_l \le 0$. Such a relaxation, which corresponds to a signal sector, can help covert the signal sector constraint into a convex one. Therefore we can solve the problem by formulating a convex optimization problem to conduct Sectjunction localization.

Then we obtain optimal weights by solving a convex optimization. We construct the weighted sum of expected signal difference as our objective function. By jointly considering the Wi-Fi fingerprint map and the Wi-Fi sector distance constraints, Sectjunction localization is formulated into a convex optimization as follows:

$$\begin{aligned}
\underset{\omega_i}{\text{minimize}} \quad & \sum_{i=1}^{n_m} \Delta \left(\mathbf{p}, \mathbf{q}_i \right) \cdot \omega_i \\
\text{subject to} \quad & \hat{\mathbf{x}} = \sum_{i=1}^{n_m} \omega_i \cdot \mathbf{r}_i, \mathbf{r}_i \in C_m, \\
& \|\hat{\mathbf{x}} - \mathbf{x}_l\|_2 \le d_l, l \in \Lambda, \\
& \sum_{i=1}^{n_m} \omega_i = 1, 0 \le \omega_i \le 1.
\end{aligned} \qquad (10)$$

The output of the formulation above, ω_i, can be used to estimate the corresponding position by Equation (9).

A target compares its signal vector with the Wi-Fi fingerprint and obtains a series of expected signal difference with the whole signal map. Then the solution to this optimization problem ensures that the weight assigned to each reference point in estimation can jointly minimize the above difference in signal map and satisfy the sector constraints. In other words, the RPs that have small signal difference and lie in junction of sectors will obtain higher weights in Equation (9) and contribute more in target estimation. Hence in the minimization problem, the influence from the wrong nearest neighbors (corresponding RPs) in signal space is reduced.

VI. EXPERIMENTAL RESULTS

A. Experiment Setting and Comparison Schemes

To study Sectjunction performance, we have conducted experiments in the atrium in HKUST of $2,000\ m^2$. For each reference point, 15 samples of RSSI vector are collected each from 4 different directions (north, west, south and east). Overall there are 6 clusters and Figure 4 shows the clustered reference points in the site.

Unless otherwise stated, we use the values below as our baseline parameters: number of selected APs is 40; the grid size of reference points in site survey is 5 meters; total number of reference points is 183; each target collects 3 samples and uses the mean of their estimation results for localization; 4 exemplars are elected at each cluster; and the width of interval in signal range discretization is 3 dB.

We compare our scheme with two algorithms: the state-of-the-art probability-based localization method [12] and the traditional k-NN (Nearest Neighbors) algorithm [13] based on Wi-Fi RSSI vector *cosine* similarity (k = 10 in our baseline).

We mainly use the localization error as our performance metric, which is the error between the target u's estimated location $\hat{\mathbf{x}}_u$ and its true location $\mathbf{x}_u$, i.e., $e_u = \|\hat{\mathbf{x}}_u - \mathbf{x}_u\|_2$. Denote all the sampled targets as V. We evaluate the overall performance by mean error (ME) of all targets $ME = (\sum_{u=1}^{|V|} e_u)/(|V|)$.

B. Experimental Results

In Figure 5, we plot the performance of the three algorithms against the number of APs used. As the number of APs used increases, all three algorithms' performance improves. Then all the mean errors converge. Sectjunction is less susceptible than the other two algorithms under sparse APs coverage. It is mainly because the constraint from reference point clustering and sector junction ensures that the target is in a small region and prevent a dispersed set of nearest neighbors in location estimation.

In Figure 6, we plot the mean error versus site survey grid size for the three schemes. As the grid size increases, the performance decreases. However, Sectjunction achieves better performance under large grid size because it reduces the error by junction of RSS sectors. Even though grid size increases, we can still match the target to a physically close RP.

In Figure 7, we show the cumulative probability of location error for the three schemes. By considering intersection of multiple Wi-Fi sectors, Sectjunction achieves much higher accuracy than the other two algorithms. K-NN algorithm is prone to signal measurement uncertainty and thus generates a disperse set of nearest neighbors. Probability-based algorithm assumes a certain distribution of signal level at each reference point and therefore cannot represent real signal distribution under limited sampling.

In Figure 8, we show the cumulative error probability with and without cluster mapping. We see that applying cluster mapping can improve the localization accuracy. Through cluster mapping, the target is constrained in a small area. Therefore, we achieve a close set of nearest neighbors and hence more accurate estimation.

In Figure 9, we compare the cumulative error probability for different signal comparison metrics in the objective function of *Sectjunction*. Denote the target RSSI vector as $\mathbf{p}$ and fingerprint data at RP j as $\mathbf{q}_j$. Both Euclidean distance
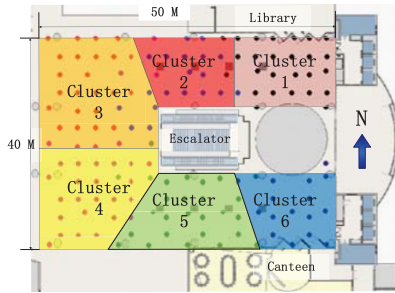
Fig. 4. Reference points clustering in the HKUST atrium. Different colors represent different clusters.
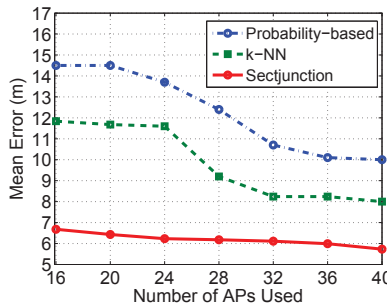

Fig. 5. Impact of APs number used over different algorithms in HKUST atrium.
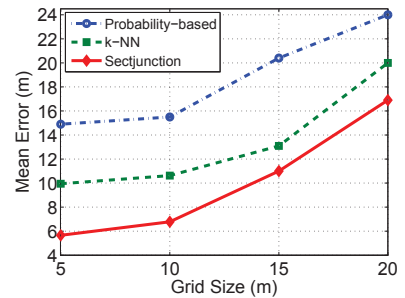

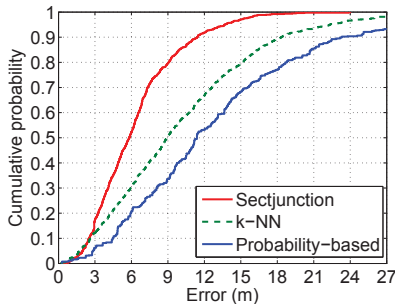Fig. 6. Impact of grid size over different algorithms in HKUST atrium.

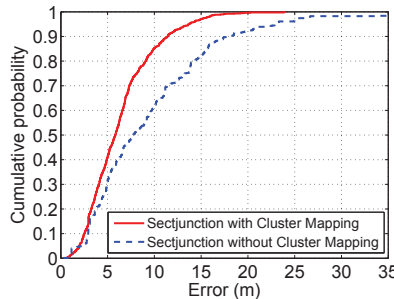Fig. 7. Performance comparison of the algorithms in HKUST atrium.

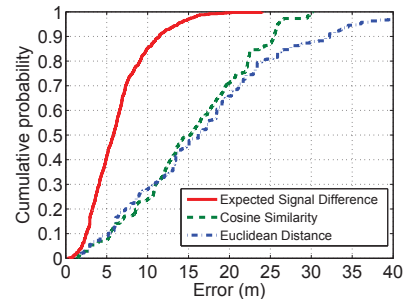Fig. 8. Performance of applying cluster mapping in Sectjunction.

Fig. 9. Performance of different signal comparison metrics in Sectjunction.

$(Eud(j) = \|\mathbf{p} - \mathbf{q}_j\|_2)$ and Cosine similarity (Equation (3)) do not consider random signal fluctuation in the measurement. In contrast, expected signal difference as proposed in our scheme discriminates the fingerprints much better by penalizing the RPs that have high signal randomness. It effectively mitigates the measurement uncertainty and achieves much better performance than the other two similarity metrics.

VII. CONCLUSION

Traditional Wi-Fi indoor fingerprint-based localization is often based on comparing signal vector at each reference point with the target RSSI vector. Due to measurement uncertainty, this often leads to a dispersed set of nearest reference points and hence unsatisfactory estimation error.

In this work, we address the problem by proposing a novel, efficient and highly accurate algorithm called Sectjunction. Sectjunction maps each of the AP signals measured at a target to a sector in the site. It also maps the target RSSI vector to a cluster area in the site. Using convex optimization based on an objective to compare random signals, the target location can be estimated by the junction of the sectors and cluster.

We have implemented Sectjunction and conducted extensive experimental study in our campus. Sectjunction is shown to effectively mitigate measurement uncertainty, achieving substantially lower estimation error as compared with state-of-the art and traditional schemes.

REFERENCES

[1] C. Wu, Z. Yang, Y. Liu, and W. Xi, "Will: Wireless indoor localization without site survey," *IEEE Trans. Parallel and Distributed Systems*, vol. 24, no. 4, pp. 839–848, 2013.

[2] Y. Chen, D. Lymberopoulos, J. Liu, and B. Priyantha, "Indoor localization using fm signals," *IEEE Trans. Mobile Computing*, vol. 12, no. 8, pp. 1502–1517, 2013.

[3] J. Wang and D. Katabi, "Dude, where's my card?: Rfid positioning that works with multipath and non-line of sight," in *Proc. ACM SigComm*, 2013, pp. 51–62.

[4] P. Mirowski, H. Steck, P. Whiting, R. Palaniappan, M. MacDonald, and T. K. Ho, "Kl-divergence kernel regression for non-gaussian fingerprint based localization," in *Proc. IPIN*, 2011, pp. 1–10.

[5] G. Shen, Z. Chen, P. Zhang, T. Moscibroda, and Y. Zhang, "Walkie-markie: indoor pathway mapping made easy," in *Proc. USENIX NSDI*, 2013, pp. 85–98.

[6] Y. Jiang, Y. Xiang, X. Pan, K. Li, Q. Lv, R. P. Dick, L. Shang, and M. Hannigan, "Hallway based automatic indoor floorplan construction using room fingerprints," in *Proc. UbiComp*, 2013, pp. 315–324.

[7] H. Wang, S. Sen, A. Elgohary, M. Farid, M. Youssef, and R. R. Choudhury, "No need to war-drive: unsupervised indoor localization," in *Proc. ACM MobiSys*, 2012, pp. 197–210.

[8] W. Sun, J. Liu, C. Wu, Z. Yang, X. Zhang, and Y. Liu, "Moloc: On distinguishing fingerprint twins," in *Proc. IEEE ICDCS*, Jul. 2013.

[9] M. Atia, A. Noureldin, and M. Korenberg, "Dynamic online-calibrated radio maps for indoor positioning in wireless local area networks," *IEEE Trans. Mobile Computing*, vol. 12, no. 9, pp. 1774–1787, 2013.

[10] J. Han, M. Kamber, and J. Pei, *Data mining: concepts and techniques*. Morgan kaufmann, 2006.

[11] G. Salton and M. J. McGill, *Introduction to Modern Information Retrieval*. New York, NY, USA: McGraw-Hill, Inc., 1986.

[12] M. Youssef and A. Agrawala, "The horus wlan location determination system," in *Proc. ACM MobiSys*, 2005, pp. 205–218.

[13] H. Liu, H. Darabi, P. Banerjee, and J. Liu, "Survey of wireless indoor positioning techniques and systems," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 37, no. 6, pp. 1067–1080, 2007.