Hindawi Publishing Corporation
Mobile Information Systems
Volume 2015, Article ID 416197, 14 pages
http://dx.doi.org/10.1155/2015/416197

Research Article

Crowd-Sourced Mobility Mapping for Location Tracking Using Unlabeled Wi-Fi Simultaneous Localization and Mapping

Mu Zhou,[1] Kunjie Xu,[2] Zengshan Tian,[1] Haibo Wu,[3] and Ruikang Shi[1]

[1]*Chongqing Key Lab of Mobile Communications Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China*
[2]*Ericsson, San Jose, CA 95134, USA*
[3]*China Internet Research Lab, Computer Network Information Center, Chinese Academy of Sciences, Beijing 100190, China*

Correspondence should be addressed to Mu Zhou; zhoumu@cqupt.edu.cn

Received 1 March 2015; Revised 23 April 2015; Accepted 29 April 2015

Academic Editor: Laurence T. Yang

Due to the increasing requirements of the seamless and round-the-clock Location-based services (LBSs), a growing interest in Wi-Fi network aided location tracking is witnessed in the past decade. One of the significant problems of the conventional Wi-Fi location tracking approaches based on received signal strength (RSS) fingerprinting is the time-consuming and labor intensive work involved in location fingerprint calibration. To solve this problem, a novel unlabeled Wi-Fi simultaneous localization and mapping (SLAM) approach is developed to avoid the location fingerprinting and additional inertial or vision sensors. In this approach, an unlabeled mobility map of the coverage area is first constructed by using the crowd-sourcing from a batch of sporadically recorded Wi-Fi RSS sequences based on the spectral cluster assembling. Then, the sequence alignment algorithm is applied to conduct location tracking and mobility map updating. Finally, the effectiveness of this approach is verified by the extensive experiments carried out in a campus-wide area.

1. Introduction

Location tracking in wireless mobile environments plays an important role in recent development of Location-based services (LBSs), such as visitor navigation, elderly health care, facility management, transportation, and emergency rescue [1–3]. With the significant growing interests in Wi-Fi technology and the ubiquitous deployment of Wi-Fi infrastructures, intense attention has been paid to various approaches of Wi-Fi network aided location tracking [4]. Location tracking using Wi-Fi received signal strength (RSS) fingerprints presents many challenges due to less predictable variations of RSSs with respect to the distances from the access points (APs) to the receiver. The Wi-Fi network aided location tracking approaches can be generally classified into two categories. In the first one, only Wi-Fi RSS data are used to estimate the parameters in propagation models [5, 6] or construct a radio map corresponding to the coverage area [7, 8]. Then, the target locations are estimated based on the propagation model aided triangulation algorithm or radio map aided fingerprint matching. The major limitations of such category are that (i) the RSS propagation modeling requires the prior knowledge of AP locations and easily leads to the low accuracy when the relations between the RSSs and their corresponding distances are not captured appropriately and (ii) the construction of radio map usually involves the site survey over a large number of reference points (RPs) and their associated RSSs, which deteriorates the adaptability for the large coverage area, and meanwhile the radio map can easily become outdated.

In the second one, the Wi-Fi RSSs and the acquired measurements from several different types of inertial and vision sensors are integrated to achieve simultaneous localization and mapping (SLAM) [9]. The SLAM approach can be used to construct the virtual floor plan corresponding to the coverage area, meanwhile conducting location tracking. Although the RSS propagation modeling and radio map construction are not required, the combined problem of target localization and mobility mapping needs to be significantly concerned in order to make the SLAM approach robust to

the radio noise. Specifically, the mobility mapping built up from the environment helps to keep tracking of the target locations, while the results of location tracking feed the process of the mobility mapping. The aim of this paper is to use solely Wi-Fi network to construct a mobility map without location fingerprinting and inertial or vision sensing (or called unlabeled mobility map) for the coverage area. Moreover, the proposed approach is expected to achieve a high-enough probability of locating the new RSS data into the corridor where these data are actually recorded, namely, the corridor-level location accuracy. In summary, the following two problems are resolved in this paper.

(i) How to construct an unlabeled mobility map by using crowd-sourcing based on the sporadically recorded RSS data without explicit information of their physical coordinates?

(ii) How to simultaneously conduct location tracking and mobility map updating?

The solution to the construction of the unlabeled mobility map is by spectral cluster assembling, which consists of the intrasequence spectral clustering step and the intersequence cluster assembling step. The solution to the target location tracking and mobility map updating is called sequence alignment algorithm, which matches the new RSS data against the constructed mobility map and selects the location point (LP) with the highest alignment similarity as the tracked location for each location query.

Solution 1 (spectral cluster assembling). Mobility map construction by using crowd-sourcing consists of two main steps: intrasequence clustering and intersequence assembling. In the intrasequence clustering step, each string of time-stamped consecutive RSS samples recorded by the person is first represented as an RSS sequence, where the RSS samples in each sequence are sequenced in chronological order. Then, for any recorded RSS sequence, spectral clustering is performed on this sequence to classify the RSS samples into different clusters. The spectral clustering used in this paper preserves the locality of RSS samples in RSSs and time-stamps. Another significant advantage of spectral clustering is that the RSS samples which are high-dimensional can be mapped into a low-dimensional space, which is beneficial to the information retrieval and data mining [10]. Specifically, after an RSS sequence is recorded, the similarity of every two RSS samples in this sequence is first calculated, and then a low-dimensional mapped space is constructed by using Laplacian embedding. Finally, the K-means clustering is performed to classify the RSS samples into different clusters.

In the intersequence assembling step, the concept of winning path used in Smith-Waterman algorithm for protein sequencing [11] is first applied to identify the clusters which have high similarities between each other. The similarity of any two clusters is measured by their cumulative matching score, which is calculated based on the Kullback-Leibler (KL) divergence of their RSS distributions and the cumulative matching scores of several previous pairs of clusters (or called pairs of clusters with time-stamps before the current pair of clusters). After the matrix of cumulative matching scores between the clusters, namely, the scoring space, for each pair of RSS sequences is obtained, the pairs of RSS sequences having one or more pairs of clusters with the cumulative matching scores higher than a given threshold are required to be assembled, while the corresponding pairs of clusters are merged into a new cluster. This process is repeated until all the remaining pairs of clusters are with the cumulative matching scores lower than the threshold. At this point, each remaining cluster is recognized as an LP in mobility map.

Solution 2 (sequence alignment algorithm). For the purpose of simultaneously conducting location tracking and mobility map updating, the three RSS sequences containing the largest number of RSS samples, namely, the three longest RSS sequences, in each LP is first selected as the virtual fingerprints. The concept of sequence alignment in protein sequencing is then used to find the strings of tracking LPs which have the highest alignment similarities to the new RSS sequences. Finally, based on the time-stamped transition relations of LPs, the target locations can be tracked by using the modified strings of tracking LPs with the improved accuracy performance. A preliminary discussion on the time-stamped transition relations of LPs in mobility map can be found in the previous work [4]. In this paper, to validate the construction of mobility map by using crowd-sourcing and the unlabeled Wi-Fi SLAM, extensive experiments are conducted in a campus-wide area.

The remainder of this paper is organized as follows. Section 2 reviews some related work on SLAM based location tracking. In Section 3, the steps of the proposed crowd-sourced mobility mapping for location tracking without the site survey on location fingerprinting and inertial or vision sensing are described in detail. The experimental results and the related discussions are shown in Section 4. Finally, this paper is concluded in Section 5.

2. Related Work

Tracking the target locations cost-efficiently by using the conventional RSS propagation modeling and construction of radio map is a challenging work since both of them involve the time-consuming and labor intensive site survey on the relations between RSSs and physical locations [12, 13]. To achieve the cost-efficiency purpose, the recent works began to focus on the SLAM based location tracking approaches, which have good adaptability to the environment with low site survey effort and high robustness to the environment noise.

Wang and Thorpe [14] integrated the SLAM and detection and tracking of moving objects (DTMO) approaches and verified the efficiency of the integrated SLAM and DTMO at high speed in a large crowded city environment. A simulation system is designed to be implemented to analyze and validate the SLAM based on the well-known extended Kalman filter (EKF) [15]. There are generally three major approaches studied for SLAM, optimal control approach, local submap approach, and frontier based approach, to achieve the localization and mapping purposes in an active and intelligent way. Similar work on Kalman filter (KF) based SLAM has been addressed extensively in the literature [16, 17].

Chatterjee and Matsuno [16] introduced a new approach of using the neurofuzzy assisted EKF to enhance the performance of SLAM. There are two limitations of this approach. First of all, to suitably use the neurofuzzy supervision, the free parameters of neurofuzzy system are required to be learned carefully. Second, this approach could be unavailable when the process and sensor noise covariance matrices are inaccurate. To improve the accuracy and fast convergence of state estimation involved in KF, the pseudolinear model based Kalman filter (PLKF) based SLAM is proposed [17]. The PLKF based SLAM outperforms the conventional EKF based SLAM since the pseudolinear model preserves the nonlinearity in the system, motion, and observation models.

Wang et al. [18] proved that the particle filtering (PF) can also be used to improve the performance of SLAM. The PF can not only reduce the complexity of the data but also enhance the real-time capacity of SLAM. The weakness of the PF based SLAM is that many types of sensors and the related data fusion process are required to guarantee the effectiveness of the feature extraction in the unknown and highly complex environments. The Rao-Blackwellised particle filter (IRBPF) based SLAM is used to achieve the accurate localization and generate a consistent map of the environment [19]. The IRBPF based SLAM first uses PF to estimate the posterior probability distributions of the target. The adaptive resampling approach is then applied to reduce the risk of sample depletion. Finally, the motion and observation models are constructed based on the data from a ranging sensor and an odometer. However, this approach is not suitable for the highly dynamic environment.

Luo et al. [20] presented a new concept of vision based SLAM (V-SLAM), in which the visual feature point buffer and human body elimination are used to decrease the estimative errors to the system. Specifically, the V-SLAM implements the visual feature point buffer to filter the temporary feature points which are extracted from features from accelerated segment test (FAST) corner detector and conducts the human body elimination to help V-SLAM to be more accurate. As an application example, the V-SLAM is used to aid inertial navigation by compensating for inertial navigation divergence [21]. Sazdovski and Silson [21] proved that such an integrated inertial navigation system with V-SLAM requires the coordination between the guidance and control measurements to achieve the high navigation accuracy.

The aforementioned SLAM is based on the single target and single model. The decentralized platform developed by Saeedi et al. [22] is known as the first application of neural network for SLAM in multiple targets condition. Their proposed approach consists of five modules of the high-level map segmentation for map preprocessing, application of self-organizing maps for preprocessed map clustering, inclusion of map uncertainty in learning phase, estimation of the relative transformation matrix of every two maps, and use of surface norms for relative transformation determination. Yingmin and Ding [23] investigated the nonlinear interacting multiple model (IMM) based SLAM to solve the problem of the statistical property mutation of SLAM. There are five steps involved in the nonlinear IMM based SLAM, the model condition reinitialization, model condition filtering and data association, model probability updating and estimate fusion, and state augmentation and map building.

By considering the indoor Wi-Fi network, the Wi-Fi SLAM is invented to gather the location and mapping information simultaneously [24]. Specifically, the Wi-Fi SLAM first uses the location fingerprinting to get an idea of what the construction of a particular building is going to do to Wi-Fi RSS distributions. The initial trajectories are then constructed based on the measurements from multiple sensors on a smartphone including the accelerometer, gyroscope, and magnetometer. Finally, the constructed trajectories are mated with the results of Wi-Fi RSS trilateration to serve fine-grain localization and create accurate indoor maps. Since the measurements are gathered by different sensors, the Wi-Fi SLAM uses pattern recognition and machine learning to draw the correlations between these measurements for data fusion purpose. The SmartSLAM is recognized as another representative Wi-Fi network based SLAM by using the measurements from inertial sensors and Wi-Fi network [9]. One of the significant problems of SmartSLAM is that the high energy cost involved in continuously scanning multiple sensors and Wi-Fi channels seriously limits the practical use.

In this paper, a better solution will be provided to the simultaneous mobility mapping and location tracking by using crowd-sourcing from the sporadically recorded Wi-Fi RSS data without location fingerprinting and inertial or vision sensing. In summary, the three major contributions of this paper are as follows.

(i) An unlabeled Wi-Fi SLAM approach is developed to avoid the location fingerprinting and additional inertial or vision sensors.

(ii) The mobility map is constructed by using crowd-sourcing based on the sporadically recorded Wi-Fi RSS data without explicit information of their physical coordinates.

(iii) The concept of alignment similarities between the newly recorded RSS data and prestored virtual fingerprints is utilized to achieve the location tracking.

The major notations and parameters used in this paper are summarized in Notations and Parameters section.

3. System Description

3.1. Intrasequence Clustering

3.1.1. Problem Statement. In the system, each person holding a Wi-Fi RSS receiver walks around the coverage area and collects Wi-Fi RSS sequences. It is assumed that N such sequences are collected where $R^{\ell} = \{\boldsymbol{\mu}_1^{\ell}, \ldots, \boldsymbol{\mu}_{N^{\ell}}^{\ell}\}$ ($\ell = 1, \ldots, N$) is the ℓth sequence and $\boldsymbol{\mu}_i^{\ell} = (\mu_{i,1}^{\ell}, \ldots, \mu_{i,M}^{\ell})$ ($i = 1, \ldots, N^{\ell}$) is the ith sample (of dimensions M) in R^{ℓ} which contains N^{ℓ} samples. Each sample is a vector containing the RSS values from M APs. The jth ($j = 1, \ldots, M$) element in $\boldsymbol{\mu}_i^{\ell}$, $\mu_{i,j}^{\ell}$ is the RSS value from the jth AP. The difference in RSSs and time-stamps between any two RSS samples in the same sequence is calculated by $\mathrm{diff}_R(\boldsymbol{\mu}_i^{\ell}, \boldsymbol{\mu}_{i'}^{\ell}) = \|\boldsymbol{\mu}_i^{\ell} - \boldsymbol{\mu}_{i'}^{\ell}\|_2$ and

$\text{diff}_T(\boldsymbol{\mu}_i^\ell, \boldsymbol{\mu}_{i'}^\ell) = |T_i^\ell - T_{i'}^\ell|$ where $T_i^\ell = (i-1)\delta$ is the time-stamp of $\boldsymbol{\mu}_i^\ell$ in R^ℓ and δ is the sampling interval.

Here, the intrasequence clustering problem is as follows. Given a set of N^ℓ RSS samples $\{\boldsymbol{\mu}_1^\ell, \ldots, \boldsymbol{\mu}_{N^\ell}^\ell\}$ belonging to a K^ℓ-dimensional mapped space $\mathbf{R}^{K^\ell}$, which is embedded in the raw M-dimensional RSS space $\mathbf{R}^M$ ($M \geq K^\ell$), a set of N^ℓ mapped vectors $\{\widehat{\boldsymbol{\mu}}_1^\ell, \ldots, \widehat{\boldsymbol{\mu}}_{N^\ell}^\ell\}$ in $\mathbf{R}^{K^\ell}$ is first found such that $\widehat{\boldsymbol{\mu}}_i^\ell$ is mapped from $\boldsymbol{\mu}_i^\ell$, and then the N^ℓ mapped vectors $\widehat{\boldsymbol{\mu}}_1^\ell, \ldots, \widehat{\boldsymbol{\mu}}_{N^\ell}^\ell$ are classified into Φ^ℓ RSS clusters. Finally, the RSS samples which have the corresponding mapped vectors in the same cluster are also classified into the same RSS cluster.

The first advantage of intrasequence clustering is that the similarities of RSS samples in RSSs and time-stamps between any two clusters are minimized. Second, the graph Laplacian which is selected for dimensionality reduction can avoid the situation that the isolated RSS samples accidentally form some outlier RSS clusters by minimizing the values of Ncut over all the RSS clusters [10, 25].

3.1.2. Steps. The steps of intrasequence clustering are provided as follows.

Step 1. The similarity of $\boldsymbol{\mu}_i^\ell$ and $\boldsymbol{\mu}_{i'}^\ell$ is computed as $W_{i,i'} = \exp(-\alpha_R \text{diff}'_R(\boldsymbol{\mu}_i^\ell, \boldsymbol{\mu}_{i'}^\ell) - \alpha_T \text{diff}'_T(\boldsymbol{\mu}_i^\ell, \boldsymbol{\mu}_{i'}^\ell))$, where α_R and α_T ($\alpha_R + \alpha_T = 1$; $0 \leq \alpha_R, \alpha_T \leq 1$) are two tunable weighting factors and $\text{diff}'_R(\boldsymbol{\mu}_i^\ell, \boldsymbol{\mu}_{i'}^\ell)$ and $\text{diff}'_T(\boldsymbol{\mu}_i^\ell, \boldsymbol{\mu}_{i'}^\ell)$ are the normalized values of $\text{diff}_R(\boldsymbol{\mu}_i^\ell, \boldsymbol{\mu}_{i'}^\ell)$ and $\text{diff}_T(\boldsymbol{\mu}_i^\ell, \boldsymbol{\mu}_{i'}^\ell)$, respectively.

Step 2. Consider the problem of mapping the RSS samples in each sequence onto a line such that the RSS samples with large similarities in RSSs and time-stamps are corresponding to the mapped points which can stay as close together as possible on the line. Let $\widehat{R}_1^\ell = \{r_{1,1}^\ell, \ldots, r_{1,N^\ell}^\ell\}$ be the set of the N^ℓ mapped points. The optimal objective function to this problem can be described as $\min_{\widehat{R}_1^\ell}\{\sum_{i'',i'''=1}^{N^\ell}(r_{1,i''}^\ell - r_{1,i'''}^\ell)^2 W_{i'',i'''}\}$.

Step 3. Compute the Laplacian matrix $\mathbf{L}^\ell = \mathbf{D}^\ell - \mathbf{W}^\ell = [D_{i,i'}^\ell - W_{i,i'}^\ell]_{i,i'=1}^{N^\ell}$ for each RSS sequence. If $i = i'$, $D_{i,i}^\ell$ is set as $\sum_{i''=1}^{N^\ell} W_{i,i''}^\ell$. Otherwise, $D_{i,i'}^\ell$ is set as 0. By using the Lagrange multiplier method [4], the problem described in Step 2 is converted to $\min_{\widehat{R}_1^\ell}\{\lambda\}$, where λ's are the eigenvalues of $\mathbf{L}^\ell$.

Step 4. Consider the general problem of mapping each RSS sequence (of dimensions M) into the K^ℓ-dimensional space, where $K^\ell \leq M$. The set of mapped vectors, namely, the mapped sequence, is given by a $N^\ell \times K^\ell$ matrix $\boldsymbol{\Psi}^\ell = [\widehat{\boldsymbol{\mu}}_1^\ell \cdots \widehat{\boldsymbol{\mu}}_{N^\ell}^\ell]^T$, where $\widehat{\boldsymbol{\mu}}_i^\ell = (r_{1,i}^\ell, \ldots, r_{K^\ell,i}^\ell)$ is the mapped vector of $\boldsymbol{\mu}_i^\ell$ and the superscript "T" represents the transpose operation. Thus, the optimal objective function to this general problem equals $\min_{\widehat{R}^\ell}\{\text{tr}((\boldsymbol{\Psi}^\ell)^T \mathbf{L}^\ell \boldsymbol{\Psi}^\ell)\}$. Based on the previous work [4], the solution to this optimal objective function can be provided by the K^ℓ eigenvectors which are corresponding to K^ℓ smallest eigenvalues of the generalized eigenvalue problem $\mathbf{L}^\ell \widehat{\boldsymbol{\mu}}_i^\ell = \lambda \mathbf{D}^\ell \widehat{\boldsymbol{\mu}}_i^\ell$.

Step 5. As discussed in the literature [10, 26], the process of the previous dimensionality reduction which preserves the locality of RSS samples in RSSs and time-stamps can also yield a similar solution to the RSS clustering. In concrete terms, after the dimensionality reduction from M to K^ℓ, K-means clustering is conducted on the mapped vectors to obtain Φ^ℓ RSS clusters $C_1^\ell, \ldots, C_{\Phi^\ell}^\ell$ for each RSS sequence, where C_k^ℓ is the kth ($k = 1, \ldots, \Phi^\ell$) RSS cluster in R^ℓ.

3.2. Intersequence Assembling

3.2.1. Problem Statement. The problem to be solved by intersequence assembling is to assemble the RSS clusters obtained in intrasequence clustering step into a mobility map. Since each RSS sequence can be represented by a string of consecutive time-stamped RSS clusters, $\widehat{R}^\ell = \{C_1^\ell, \ldots, C_{\Phi^\ell}^\ell\}$, the concept of winning path in scoring space used by Smith-Waterman alignment is applied to identify the RSS clusters required to be merged. By considering the process of clusters combination graphically, each string of consecutive time-stamped RSS clusters is first viewed as a string of consecutive vertices in a graph, where any two adjacent vertices are connected by an edge. Second, the string assembling is conducted by merging the specific vertices in different strings. Finally, after all the strings of consecutive vertices have been assembled, the mobility map $G = (V, E)$ is constructed in which V and E represent the sets of vertices and edges in graph, respectively.

3.2.2. Steps. The four steps of intersequence assembling are as follows.

Step 1. For every two strings of consecutive vertices $\widehat{R}^\ell$ and $\widehat{R}^{\ell'}$, their initial scoring space is set as a zero matrix $[\mathbf{0}]_{\Phi^\ell \times \Phi^{\ell'}}$. The entry at position (s,t) ($s = 1, \ldots, \Phi^\ell$; $t = 1, \ldots, \Phi^{\ell'}$) in scoring space is the cumulative matching score $\Upsilon(C_s^\ell, C_t^{\ell'})$ between C_s^ℓ and $C_t^{\ell'}$.

Step 2. The entries in scoring space are calculated starting at position (1, 1) and proceeding from the first to the last rows in scoring space. The entry at position (s,t), $\Upsilon(C_s^\ell, C_t^{\ell'})$, is calculated that (i) when $1 \leq t \leq \Phi^{\ell'}$, $\Upsilon(C_1^\ell, C_t^{\ell'}) = 1/\Gamma(C_1^\ell, C_t^{\ell'})$; (ii) when $1 \leq s \leq \Phi^\ell$, $\Upsilon(C_s^\ell, C_1^{\ell'}) = 1/\Gamma(C_s^\ell, C_1^{\ell'})$; (iii) when $2 \leq s \leq \Phi^\ell$, $2 \leq t \leq \Phi^{\ell'}$, and $\Gamma(C_s^\ell, C_t^{\ell'}) \leq \varepsilon_D$, $\Upsilon(C_s^\ell, C_t^{\ell'}) = \Upsilon(C_{s-1}^\ell, C_{t-1}^{\ell'}) + 1/\Gamma(C_s^\ell, C_t^{\ell'})$; and (iv) otherwise, one has

$$\Upsilon\left(C_s^\ell, C_t^{\ell'}\right) = \max\left\{\begin{array}{l} \max_{s'=1}^{s-1}\left\{(1-F_M)^{s-s'}\,\Upsilon\left(C_{s'}^\ell, C_t^{\ell'}\right) + \dfrac{(1-F_D)^{s-s'}}{\Gamma\left(C_{s'}^\ell, C_t^{\ell'}\right)}\right\}, \\ \max_{t'=1}^{t-1}\left\{(1-F_M)^{t-t'}\,\Upsilon\left(C_s^\ell, C_{t'}^{\ell'}\right) + \dfrac{(1-F_D)^{t-t'}}{\Gamma\left(C_s^\ell, C_{t'}^{\ell'}\right)}\right\} \end{array}\right\}, \quad (1)$$

where F_M and F_D ($0 \leq F_M, F_D \leq 1$) are the missing factor and the damping factor, respectively; ε_D is a threshold; and

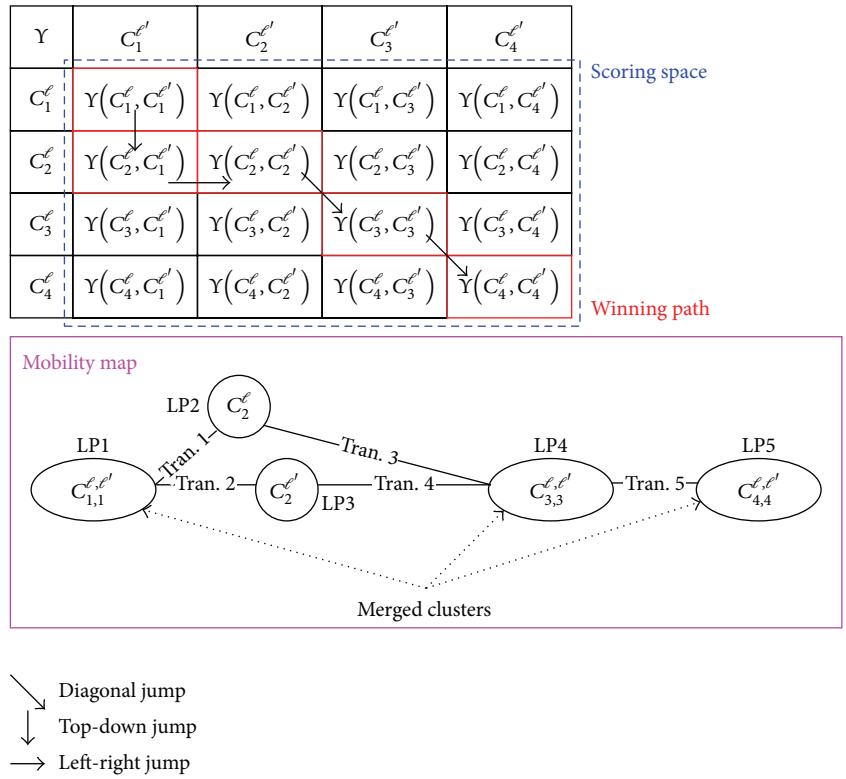


FIGURE 1: An example of intersequence assembling.

$\Gamma(C_s^{\ell}, C_t^{\ell'}) = \sum_{j=1}^{M} \sum_{\mu} p_{s,j}^{\ell}(\mu) \log_2[p_{s,j}^{\ell}(\mu)/p_{t,j}^{\ell'}(\mu)]/M$. $p_{s,j}^{\ell}(\mu)$ and $p_{t,j}^{\ell'}(\mu)$ are the probabilities of RSS value μ under the RSS distributions in C_s^{ℓ} and $C_t^{\ell'}$, respectively.

Step 3. The winning path creation is started at the position (s_1, t_1) with the largest entry, namely, the first position on the winning path, in scoring space, such that

$$(s_1, t_1) = \arg\max_{s=1,\ldots,\Phi^{\ell};t=1,\ldots,\Phi^{\ell'}} \left\{\Upsilon\left(C_s^{\ell}, C_t^{\ell'}\right)\right\},$$
$$\Upsilon\left(C_{s^1}^{\ell}, C_{t^1}^{\ell'}\right) \geq \varepsilon_{\mathrm{WC}}, \tag{2}$$

where $\varepsilon_{\mathrm{WC}}$ is a threshold.

After the first position on the winning path is obtained, it is required to go backwards to one of positions, $(s_1 - 1, t_1)$, $(s_1, t_1 - 1)$, and $(s_1 - 1, t_1 - 1)$, and compare the entries of these three positions, $\Upsilon(C_{s_1-1}^{\ell}, C_{t_1}^{\ell'})$, $\Upsilon(C_{s_1}^{\ell}, C_{t_1-1}^{\ell'})$, and $\Upsilon(C_{s_1-1}^{\ell}, C_{t_1-1}^{\ell'})$. A connection from (s_2, t_2) to (s_1, t_1), notated as $(s_2, t_2) \rightarrow (s_1, t_1)$, is constructed as the first jump on the winning path, where

$$(s_2, t_2) = \arg\max_{(s,t)\in\{(s_1-1,t_1),(s_1,t_1-1),(s_1-1,t_1-1)\}} \left\{\Upsilon\left(C_s^{\ell}, C_t^{\ell'}\right)\right\}. \tag{3}$$

This process is continued until the winning path reaches the first row or the first column in scoring space or reaches a position with the entry not larger than the threshold $\varepsilon_{\mathrm{WT}}$.

Step 4. After the winning path is created, the specific RSS clusters in $\widehat{R}^{\ell}$ and $\widehat{R}^{\ell'}$, respectively, are merged based on the concept of jumps used in Smith-Waterman alignment [27]. It is required to start from the last jump and go forward to the first jump on the winning path to identify the specific RSS clusters to be merged. There are three types of jumps involved [27]: (i) diagonal jump $(s_m, t_m) \rightarrow (s_{m-1}, t_{m-1}) = (s_m + 1, t_m + 1)$ implying that $C_{s_m+1}^{\ell}$ and $C_{t_m+1}^{\ell'}$ are required to be merged into a LP; (ii) top-down jump $(s_m, t_m) \rightarrow (s_{m-1}, t_{m-1}) = (s_m + 1, t_m)$ implying that $C_{s_m+1}^{\ell}$ is an isolated LP; and (iii) left-right jump $(s_m, t_m) \rightarrow (s_{m-1}, t_{m-1}) = (s_m, t_m + 1)$ implying $C_{t_m+1}^{\ell'}$ is an isolated LP. Figure 1 gives an example of the winning path creation, while the two RSS sequences $\widehat{R}^{\ell} = \{C_1^{\ell}, \ldots, C_4^{\ell}\}$ and $\widehat{R}^{\ell'} = \{C_1^{\ell'}, \ldots, C_4^{\ell'}\}$ happen to be both with four RSS clusters in length. Based on the created winning path in Figure 1, we can obtain three pairs of RSS clusters to be merged into $C_{1,1}^{\ell,\ell'}$, $C_{3,3}^{\ell,\ell'}$, and $C_{4,4}^{\ell,\ell'}$, respectively, and eventually construct a mobility map containing 5 LPs, notated as LP $1, \ldots, 5$, and 5 transitions, notated as Tran. $1, \ldots, 5$, between the LPs.

3.3. Sequence Alignment Algorithm. So far, the mobility map has been constructed based on the intrasequence clustering and intersequence assembling steps. In this subsection, the algorithm used for location tracking will be investigated based on the constructed mobility map. The target locations are tracked as follows.

Step 1. In each LP C_k^{ℓ}, the K RSS sequences with the largest number of RSS samples are selected as the K virtual

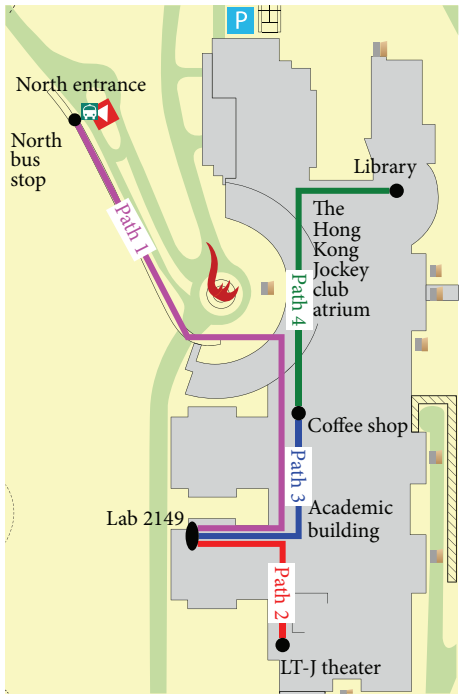


Figure 2: Layout of four paths.

fingerprints $R_k^{\ell,n} = \{\boldsymbol{\mu}_1^{\ell,n}, \ldots, \boldsymbol{\mu}_{N_k^{\ell,n}}^{\ell,n}\}$ $(n = 1, \ldots, K)$, where $N_k^{\ell,n}$ is the number of RSS samples in $R_k^{\ell,n}$, namely, length of $R_k^{\ell,n}$.

Step 2. By assuming that the interval of location query is I_S, the segment of RSS samples used for the τth $(\tau \geq 1)$ location query is $R_\tau^{\text{New}} = \{\boldsymbol{\mu}_{(\tau-1)\varepsilon_S+1}^{\text{New}}, \ldots, \boldsymbol{\mu}_{\tau\varepsilon_S}^{\text{New}}\}$, where $\varepsilon_S = [I_S/\delta]$ and the notation "$[\cdot]$" represents the rounding operation. If

$$C_{\hat{k}}^{\hat{\ell}}(\tau) = \arg\max_{k=1,\ldots,\Phi^\ell;\ell=1,\ldots,N} \left\{ \max_{n=1}^{K} \left\{ \Omega'\left(R_\tau^{\text{New}}, R_k^{\ell,n}\right) \right\} \right\},$$
$$\Omega'\left(R_\tau^{\text{New}}, R_k^{\ell,n}\right) = \max_{u=(\tau-1)\varepsilon_S+1}^{\tau\varepsilon_S} \max_{v=1}^{N_k^{\ell,n}} \left\{ \Omega\left(\boldsymbol{\mu}_u^{\text{New}}, \boldsymbol{\mu}_v^{\ell,n}\right) \right\}, \quad (4)$$

$C_{\hat{k}}^{\hat{\ell}}(\tau)$ is selected as the tracked location for the τth location query, where $\max_{n=1}^{K}\{\Omega'(R_\tau^{\text{New}}, R_k^{\ell,n})\}$ is the alignment similarity between R_τ^{New} and C_k^ℓ. In (4), $\Omega(\boldsymbol{\mu}_u^{\text{New}}, \boldsymbol{\mu}_v^{\ell,n})$ is calculated that (i) when $u = 1$ or $v = 1$, $\Omega(\boldsymbol{\mu}_u^{\text{New}}, \boldsymbol{\mu}_v^{\ell,n}) = 1/\|\boldsymbol{\mu}_u^{\text{New}}, \boldsymbol{\mu}_v^{\ell,n}\|_2$; (ii) when $u \neq 1, v \neq 1$, and $\|\boldsymbol{\mu}_u^{\text{New}}, \boldsymbol{\mu}_v^{\ell,n}\|_2 \leq \varepsilon_N$, $\Omega(\boldsymbol{\mu}_u^{\text{New}}, \boldsymbol{\mu}_v^{\ell,n}) = \Omega(\boldsymbol{\mu}_{u-1}^{\text{New}}, \boldsymbol{\mu}_{v-1}^{\ell,n}) + 1/\|\boldsymbol{\mu}_u^{\text{New}}, \boldsymbol{\mu}_v^{\ell,n}\|_2$; and (iii) otherwise, one has

$$\Omega\left(\boldsymbol{\mu}_u^{\text{New}}, \boldsymbol{\mu}_v^{\ell,n}\right) = \max \left\{ \begin{array}{l} \max_{u'=1}^{u-1} \left\{ (1-F_M)^{u-u'} \Omega\left(\boldsymbol{\mu}_{u'}^{\text{New}}, \boldsymbol{\mu}_v^{\ell,n}\right) + \dfrac{(1-F_D)^{u-u'}}{\left\|\boldsymbol{\mu}_{u'}^{\text{New}}, \boldsymbol{\mu}_v^{\ell,n}\right\|_2} \right\}, \\ \max_{v'=1}^{v-1} \left\{ (1-F_M)^{v-v'} \Omega\left(\boldsymbol{\mu}_u^{\text{New}}, \boldsymbol{\mu}_{v'}^{\ell,n}\right) + \dfrac{(1-F_D)^{v-v'}}{\left\|\boldsymbol{\mu}_u^{\text{New}}, \boldsymbol{\mu}_{v'}^{\ell,n}\right\|_2} \right\} \end{array} \right\}, \quad (5)$$

where ε_N is a threshold.

Step 3. After the tracked locations are obtained, the tracked motion path will be constructed by the following steps.

First of all, it is defined that if the difference of average time-stamps of the RSS samples in $C_{\hat{k}}^{\hat{\ell}}(\tau)$ and $C_{\hat{k}}^{\hat{\ell}}(\tau')$, respectively, is smaller than ε_S, then $C_{\hat{k}}^{\hat{\ell}}(\tau)$ is in the ε_S-neighborhood of $C_{\hat{k}}^{\hat{\ell}}(\tau')$. On this basis, it is noted that the ε_S-neighborhood relationship is naturally symmetric.

Second, by setting $\mathscr{A}_{\hat{k}}^{\hat{\ell},\tau}(\varepsilon_S)$ as the set of tracked locations which are in the ε_S-neighborhood of $C_{\hat{k}}^{\hat{\ell}}(\tau)$, one has that (i) if $C_{\hat{k}}^{\hat{\ell}}(\tau+1) \in \mathscr{A}_{\hat{k}}^{\hat{\ell},\tau}(\varepsilon_S)$, $C_{\hat{k}}^{\hat{\ell}}(\tau+1)$ is left in the set of tracked locations, and it is continued to examine whether $C_{\hat{k}}^{\hat{\ell}}(\tau+2)$ is in the ε_S-neighborhood of $C_{\hat{k}}^{\hat{\ell}}(\tau+1)$; and (ii) otherwise, it is continued to examine whether $C_{\hat{k}}^{\hat{\ell}}(\tau+2)$ is in the ε_S-neighborhood of $C_{\hat{k}}^{\hat{\ell}}(\tau)$. When $C_{\hat{k}}^{\hat{\ell}}(\tau+2) \in \mathscr{A}_{\hat{k}}^{\hat{\ell},\tau}(\varepsilon_S + \max_{\boldsymbol{\mu}_i^\ell \in C_{\hat{k}}^{\hat{\ell}}(\tau); \boldsymbol{\mu}_{i'}^{\ell'} \in C_{\hat{k}}^{\hat{\ell}}(\tau+1)} \{\text{diff}_T(\boldsymbol{\mu}_i^\ell, \boldsymbol{\mu}_{i'}^{\ell'})\})$, both $C_{\hat{k}}^{\hat{\ell}}(\tau+1)$ and $C_{\hat{k}}^{\hat{\ell}}(\tau+2)$ are left in the set of tracked locations, and it is continued to examine whether $C_{\hat{k}}^{\hat{\ell}}(\tau+3)$ is in the ε_S-neighborhood of $C_{\hat{k}}^{\hat{\ell}}(\tau+2)$. Otherwise, $C_{\hat{k}}^{\hat{\ell}}(\tau+1)$ is removed from the set of tracked locations, and it is continued to examine whether $C_{\hat{k}}^{\hat{\ell}}(\tau+3)$ is in the $(\varepsilon_S + \max_{\boldsymbol{\mu}_i^\ell \in C_{\hat{k}}^{\hat{\ell}}(\tau); \boldsymbol{\mu}_{i'}^{\ell'} \in C_{\hat{k}}^{\hat{\ell}}(\tau+2)} \{\text{diff}_T(\boldsymbol{\mu}_i^\ell, \boldsymbol{\mu}_{i'}^{\ell'})\})$-neighborhood of $C_{\hat{k}}^{\hat{\ell}}(\tau)$. This process is repeated until the most recent tracked location is reached or one obtains

$$C_{\hat{k}}^{\hat{\ell}}(\tau+\chi) \in \mathscr{A}_{\hat{k}}^{\hat{\ell},\tau}\left(\varepsilon_S + \max_{\boldsymbol{\mu}_i^\ell \in C_{\hat{k}}^{\hat{\ell}}(\tau); \boldsymbol{\mu}_{i'}^{\ell'} \in C_{\hat{k}}^{\hat{\ell}}(\tau+\chi-1)} \left\{\text{diff}_T\left(\boldsymbol{\mu}_i^\ell, \boldsymbol{\mu}_{i'}^{\ell'}\right)\right\}\right), \quad \chi \geq 3. \quad (6)$$

In the former case, all the tracked locations with the time-stamps after $C_{\hat{k}}^{\hat{\ell}}(\tau)$ are removed, whereas, in the latter case, $C_{\hat{k}}^{\hat{\ell}}(\tau+\chi-1)$ is left in the set of tracked locations and it is continued to examine whether $C_{\hat{k}}^{\hat{\ell}}(\tau+\chi+1)$ is in the ε_S-neighborhood of $C_{\hat{k}}^{\hat{\ell}}(\tau+\chi)$.

Finally, the tracked locations that remain are connected in chronological order along the path passing by the smallest number of LPs in mobility map as the tracked motion path. To conduct mobility map updating, the segments of RSS samples are merged into their corresponding tracked locations.

4. Experimental Results

4.1. Experimental Setup. In this section, we will evaluate the performance of our proposed approach for mobility map construction and location tracking on four representative paths in HKUST campus, as shown in Figure 2. They are path 1 between Lab 2149 and North Entrance; path 2 between Lab 2149 and LT-J Theater; path 3 between Lab 2149 and Coffee Shop; and path 4 between Coffee Shop and Library. A summary of these four paths is given in Table 1.

Table 1: Experimental setups of four testing paths.

Paths	Description	Average number of RSS samples
Path 1 between Lab 2149 and North Entrance	The path between the indoor and outdoor environments	190
Path 2 between Lab 2149 and LT-J Theater	The path on the 2nd floor in the building	80
Path 3 between Lab 2149 and Coffee Shop	The path between the 1st and 2nd floors in the building	85
Path 4 between Coffee Shop and Library	The path between the ground and 1st floor in the building	110

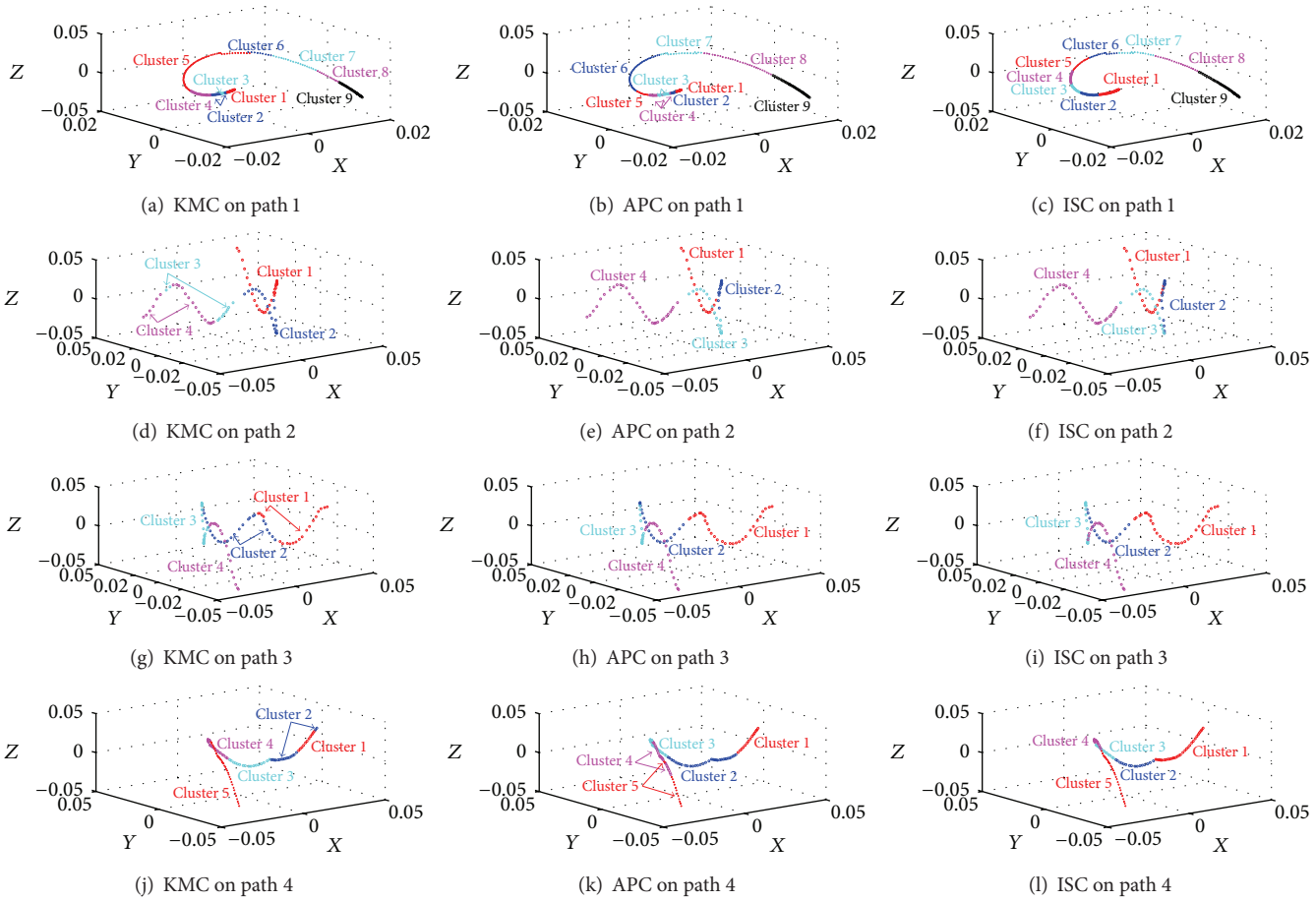


Figure 3: Intrasequence clustering on each path.

Two planned volunteers hold tablet computers which have the self-developed Wi-Fi RSS scanner software installed and walk along the four paths. On each path, each volunteer has recorded 10 RSS sequences with 5 of them in forward direction, whereas the rest in backward direction. The RSS sequences on each path are with the same sampling rate of 1 Hz (i.e., $\delta = 1$ s) and the same dimensions of 650 (i.e., $M = 650$). During the testing, the data recorded by volunteer 1 are first used to show the result of the constructed mobility map, and then the data recorded by volunteer 2 are used to examine the performance of location tracking.

4.2. Test of Mobility Map Construction

4.2.1. Results of Intrasequence Clustering. For comparison, the other two widely used approaches for intrasequence clustering are conducted: K-means clustering, a baseline approach designed for data clustering [28], and affinity propagation clustering, a state-of-the-art approach that is based on the concept of exchanging the responsible and availability messages among the data [29].

The parameters involved in intrasequence clustering are not specifically fine-tuned. Instead, they are determined at a coarse level. It is set that $K^{\ell} = 3$, $\Phi^{\ell} = [N^{\ell}/20]$, $\alpha_R = \alpha_T = 0.5$, $\varepsilon_T = 30$ s, and ε_R equals the average of distances of every two RSS samples with 30 s time-stamp difference.

The results, shown in Figure 3, are produced by K-means clustering (KMC), affinity propagation clustering (APC), and intrasequence clustering (ISC) which are conducted on each path. Different clusters are indicated by different symbols.

In Figure 3, it is observed that the proposed intrasequence clustering not only preserves the locality of RSS samples in

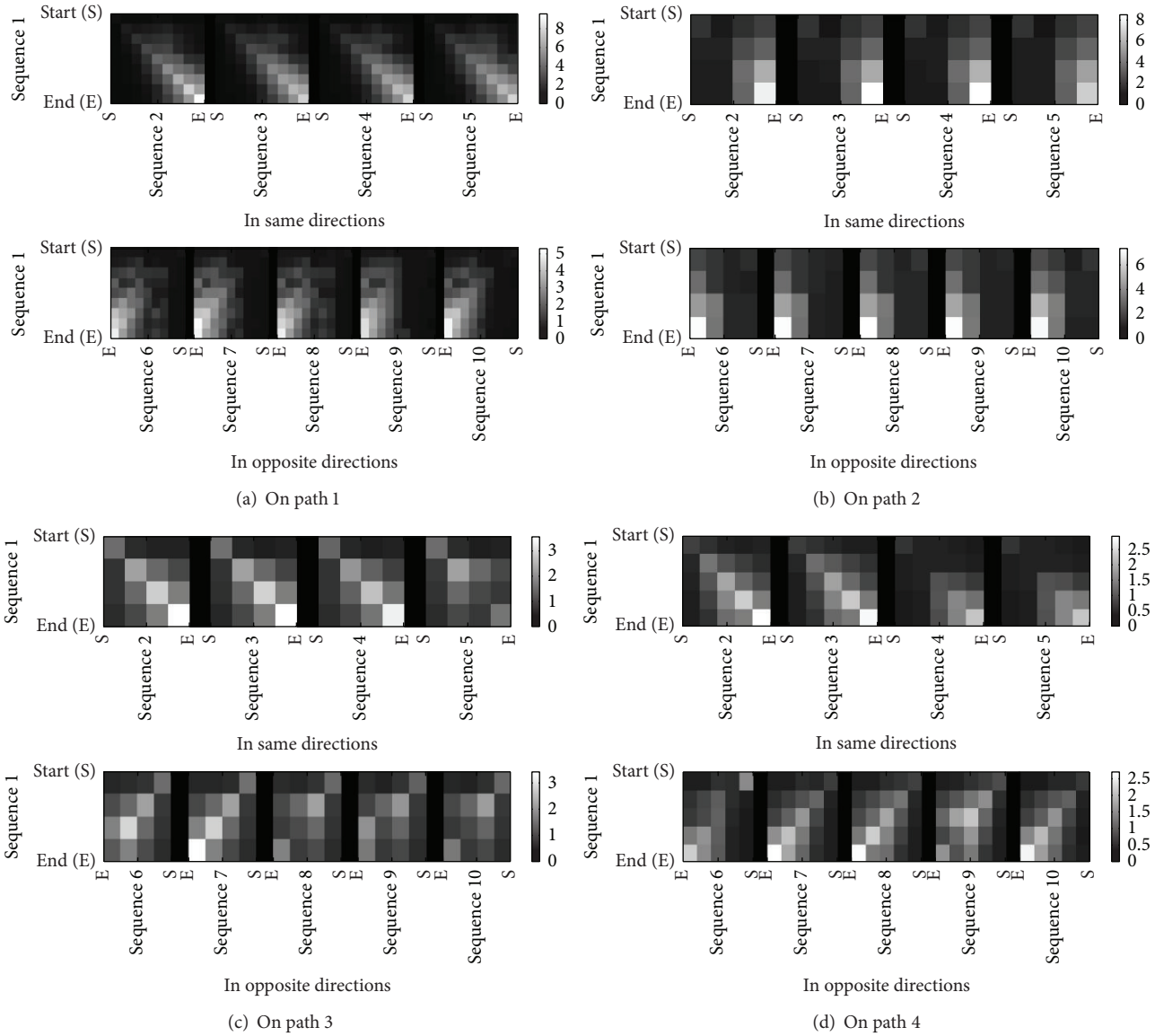
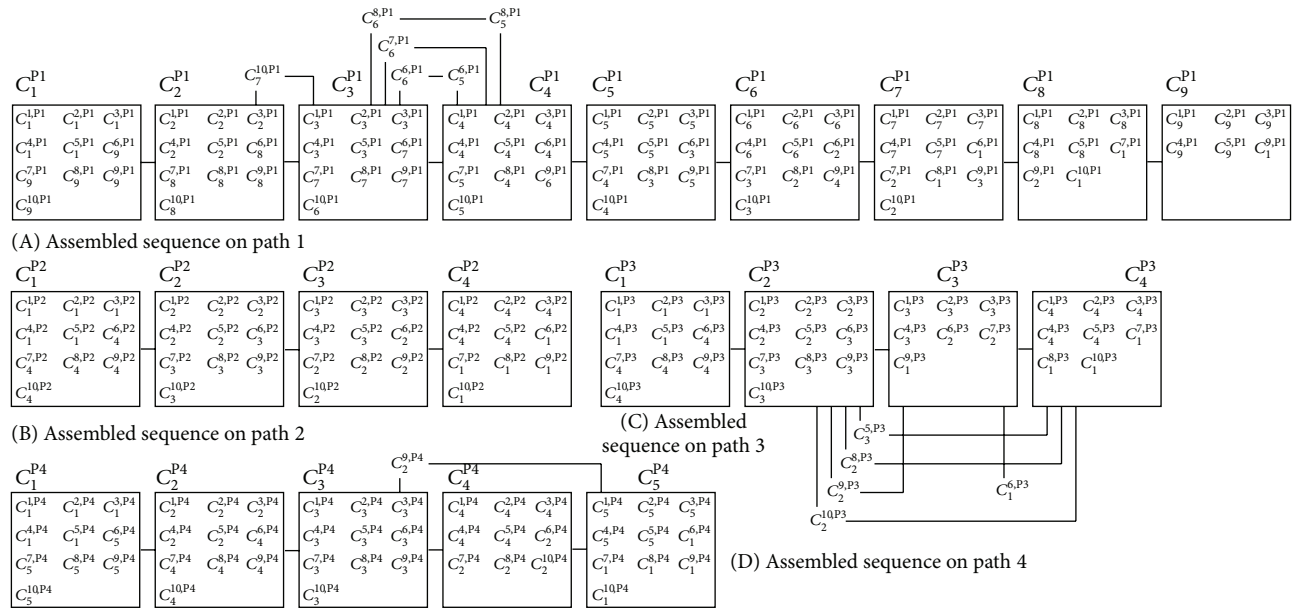


Figure 4: Scoring spaces on each path.

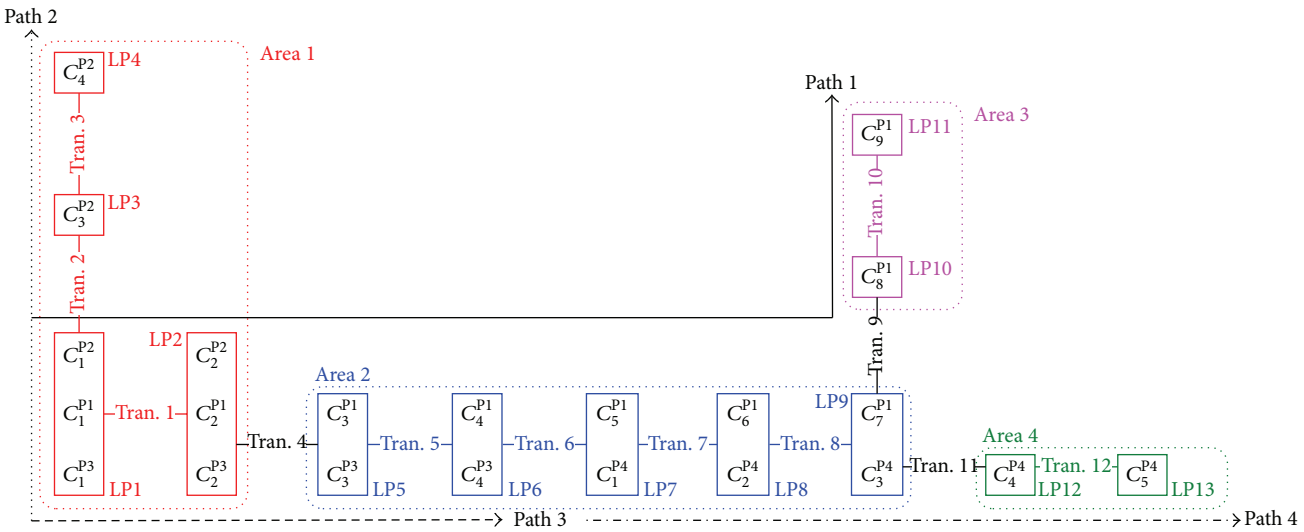
each sequence but also exhibits better adaptation to different path lengths compared to the K-means clustering and affinity propagation clustering. The results on path 1 are taken for instance. By K-means clustering, Cluster 3 which contains very few RSS samples tends to lead to poor clustering performance, while Cluster 6 which is generated by affinity propagation clustering contains too many RSS samples results in the degradation of position resolution of RSS clusters.

4.2.2. Results of Intersequence Assembling. In this subsection, the result of mobility map construction by intersequence assembling will be shown. For simplicity, the intersequence assembling is first conducted on the RSS sequences which are recorded on each path, and then they are assembled into the complete mobility map. Figure 4 shows the scoring spaces between the first recorded RSS sequence and the other 9 RSS sequences on each path. It is set that $F_M = F_D = 0.5$, $\varepsilon_{\mathrm{WT}} = 0$, and $\varepsilon_{\mathrm{WC}}$ and ε_D equal the largest entry in scoring space and the average of KL divergence of RSS distributions of every two RSS clusters, respectively.

As can be seen from Figure 4, it is found that (i) similar variation patterns of entries in scoring space can be found for any two pairs of RSS sequences with the same directions on each path and (ii) the entries in scoring space corresponding to the pair of sequences in the same direction are generally larger than the ones in scoring space corresponding to the pair of sequences in opposite directions. After the scoring spaces are obtained, it is continued to identify the specific RSS clusters required to be merged based on the concept of jumps used in Smith-Waterman alignment. Figure 5(a) shows the results of intersequence assembling on each path. The kth RSS cluster in sequence ℓ on paths 1, 2, 3, and 4 is



(a) Assembling on each path



(b) Assembling among different paths

Figure 5: Results of intersequence assembling.

notated as $C_k^{\ell,\mathrm{P1}}$, $C_k^{\ell,\mathrm{P2}}$, $C_k^{\ell,\mathrm{P3}}$, and $C_k^{\ell,\mathrm{P4}}$, respectively, and the kth merged RSS cluster on paths 1, 2, 3, and 4 is notated as C_k^{P1}, C_k^{P2}, C_k^{P3}, and C_k^{P4}, respectively. Furthermore, for simplicity, the RSS clusters that are not merged (e.g., $C_7^{10,\mathrm{P1}}$) will not be considered during the intersequence assembling among different paths since these isolated RSS clusters which contain very few RSS samples generally represent the locations where the target rarely visited.

After the assembled RSS sequence on each path is obtained, it is continued to assemble the RSS sequences among different paths into the complete mobility map. To do this, the cumulative matching scores between the merged RSS clusters in different assembled RSS sequences are first calculated. Figure 6 shows the cumulative matching scores between the merged RSS clusters on path 1 and the ones on paths 2, 3, and 4. Based on the created winning paths which are indicated by red box symbols, the mobility map consisting of 13 LPs and 12 transitions is eventually obtained, as shown in Figure 5(b).

4.3. Test of Location Tracking. The aim of this subsection is to verify that, by employing the proposed sequence alignment algorithm, the target locations can be tracked accurately in the mobile scenario. To examine whether sequence alignment algorithm can achieve corridor-level accuracy, the corridor-level segmentation is conducted on the constructed mobility map. The segmented corridors in mobility map are Corridor 1 "LP1, LP2," Corridor 2 "LP3, LP4," Corridor 3 "LP5, LP6,"

Scoring space with respect to the assembled RSS sequences on paths 1 and 2

Clusters on path 2	Clusters on path 1								
	C_1^{P1}	C_2^{P1}	C_3^{P1}	C_4^{P1}	C_5^{P1}	C_6^{P1}	C_7^{P1}	C_8^{P1}	C_9^{P1}
C_1^{P2}	1.427317	0.629255	0.646382	0.783221	0.670462	0.511344	0.504384	0.433722	0.400282
C_2^{P2}	0.723711	2.123117	1.211372	1.361272	1.112919	0.511344	0.504384	0.433722	0.374900
C_3^{P2}	1.291901	2.720135	1.968532	1.940996	1.214408	0.503487	0.484375	0.419083	0.411679
C_4^{P2}	2.085204	2.793861	2.775443	2.062959	1.221253	0.517083	0.503487	0.433106	0.416678

Scoring space with respect to the assembled RSS sequences on paths 1 and 3

Clusters on path 3	Clusters on path 1								
	C_1^{P1}	C_2^{P1}	C_3^{P1}	C_4^{P1}	C_5^{P1}	C_6^{P1}	C_7^{P1}	C_8^{P1}	C_9^{P1}
C_1^{P3}	0.717110	0.536025	0.847633	0.778365	0.670170	0.689631	0.530320	0.452790	0.416459
C_2^{P3}	0.536025	1.200853	0.974249	1.307996	0.884179	1.097977	0.530320	0.452790	0.305147
C_3^{P3}	0.449630	0.979501	1.667016	1.464400	1.733060	1.316990	0.445463	0.389454	0.308752
C_4^{P3}	0.408233	0.711488	1.434775	1.898906	1.166705	0.874900	0.404809	0.360943	0.299606

Scoring space with respect to the assembled RSS sequences on paths 1 and 4

Clusters on path 4	Clusters on path 1								
	C_1^{P1}	C_2^{P1}	C_3^{P1}	C_4^{P1}	C_5^{P1}	C_6^{P1}	C_7^{P1}	C_8^{P1}	C_9^{P1}
C_1^{P4}	0.357116	0.388600	0.431591	0.435638	0.546780	0.532444	0.572965	0.512987	0.524185
C_2^{P4}	0.336051	0.388600	0.431591	0.435638	0.546780	1.074075	0.763216	0.610079	0.524185
C_3^{P4}	0.382433	0.382433	0.413346	0.413658	0.932314	1.036606	1.512311	0.990606	0.709634
C_4^{P4}	0.470937	0.902460	0.892833	0.895578	1.022400	1.508909	1.572075	1.025370	0.724987
C_5^{P4}	0.631410	0.979061	1.384182	0.932952	1.459725	1.562408	1.998719	1.242563	0.835923

Figure 6: Cumulative matching scores between the RSS clusters.

Corridor 4 "LP7, LP8, and LP9," Corridor 5 "LP10, LP11," and Corridor 6 "LP12, LP13." It is set that $K = 3$, $I_S = 20$ s, and ε_N equals the average of distances of the newly recorded RSS samples and the ones in virtual fingerprints. Figure 7 shows the alignment similarities of the segments of RSS samples and the LPs. The real location of the target for each location query is indicated by a red box symbol. The A-1, A-5, and A-10 stand for the situations that only the first sequence, the 5 sequences in the same direction, and all the 10 sequences are used, respectively, for the testing.

From Figure 7, it is found that (i) similar patterns of alignment similarities for A-1, A-5, and A-10 situations can be found on each path and (ii) compared to A-1 situation, higher alignment similarities are obtained in A-5 and A-10 situations, whereas the ones in A-5 and A-10 situations are similar. To illustrate this result clearer, Figure 8 shows the alignment similarities achieved by the tracked locations, namely, highest alignment similarities, in A-1, A-5, and A-10 situations on each path.

Finally, Figure 9 illustrates the variations of probabilities of locating the target locations in correct corridors and LPs as the interval of location query increases from 5 s to 20 s. In Figure 9, It is found that (i) the largest mean of probabilities in correct corridors and LPs, 91.0% and 69.4%, are obtained in the condition of $I_S = 10$ s; and (ii) the probabilities in correct corridors and LPs with respect to $I_S = 5$ s, 88.1% and 65.6%, are slightly lower than the ones achieved by $I_S = 10$ s, whereas the real-time capacity is significantly enhanced.

5. Conclusions

In this paper, the crowd-sourced mobility mapping approach is proposed to solve the problem of location tracking in Wi-Fi area by using a new concept of unlabeled Wi-Fi SLAM. In the system, spectral cluster assembling is first applied to construct an unlabeled mobility map corresponding to the Wi-Fi coverage area, and then sequence alignment algorithm is used to simultaneously track the target locations and conduct mobility map updating. Extensive experiments in a campus-wide area prove that high accuracy of location tracking can be achieved without the site survey on location fingerprinting and motion sensing which are adopted by many existing Wi-Fi location systems. Furthermore, the proposed approach can also construct mobility map by using the crowd-sourcing

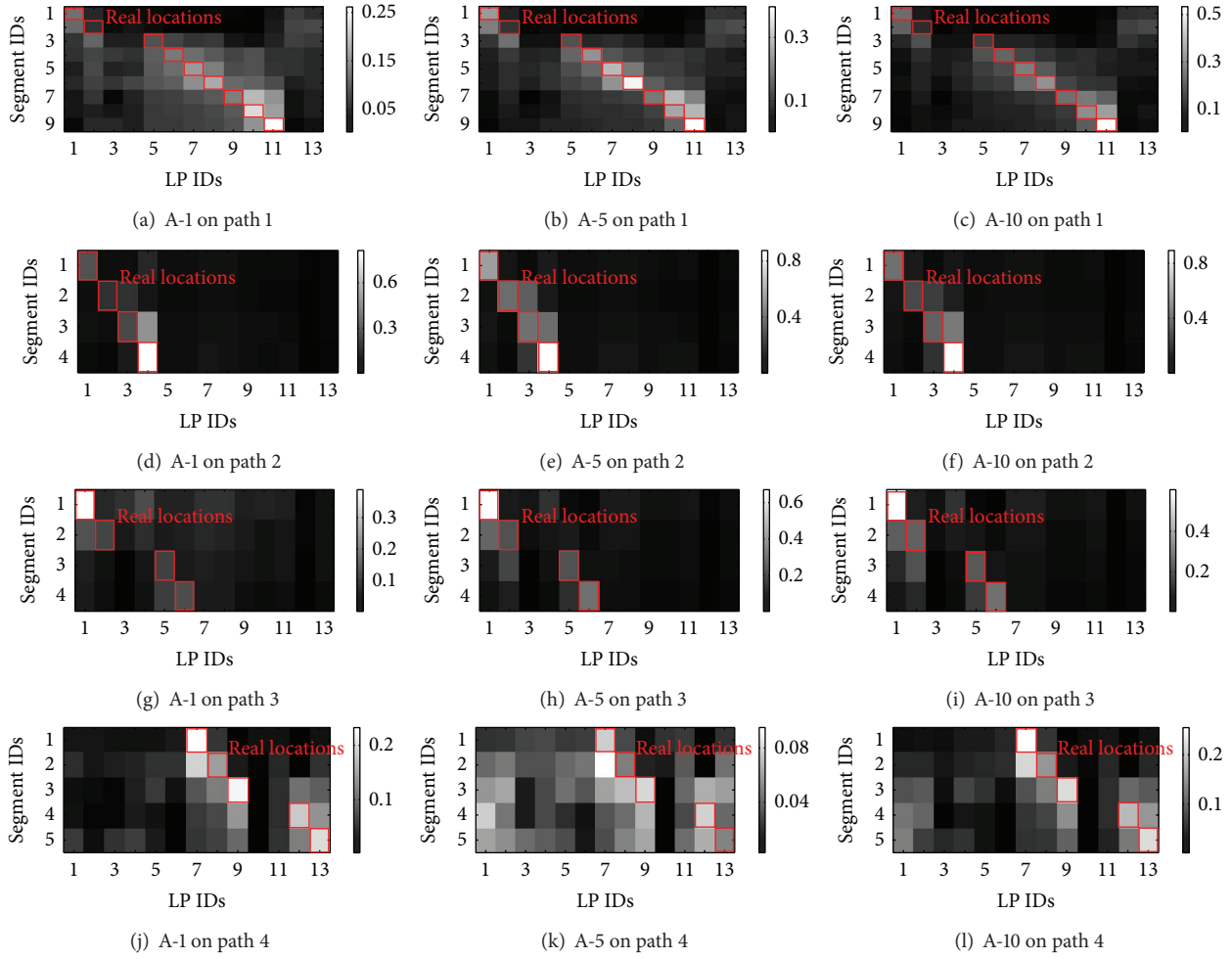


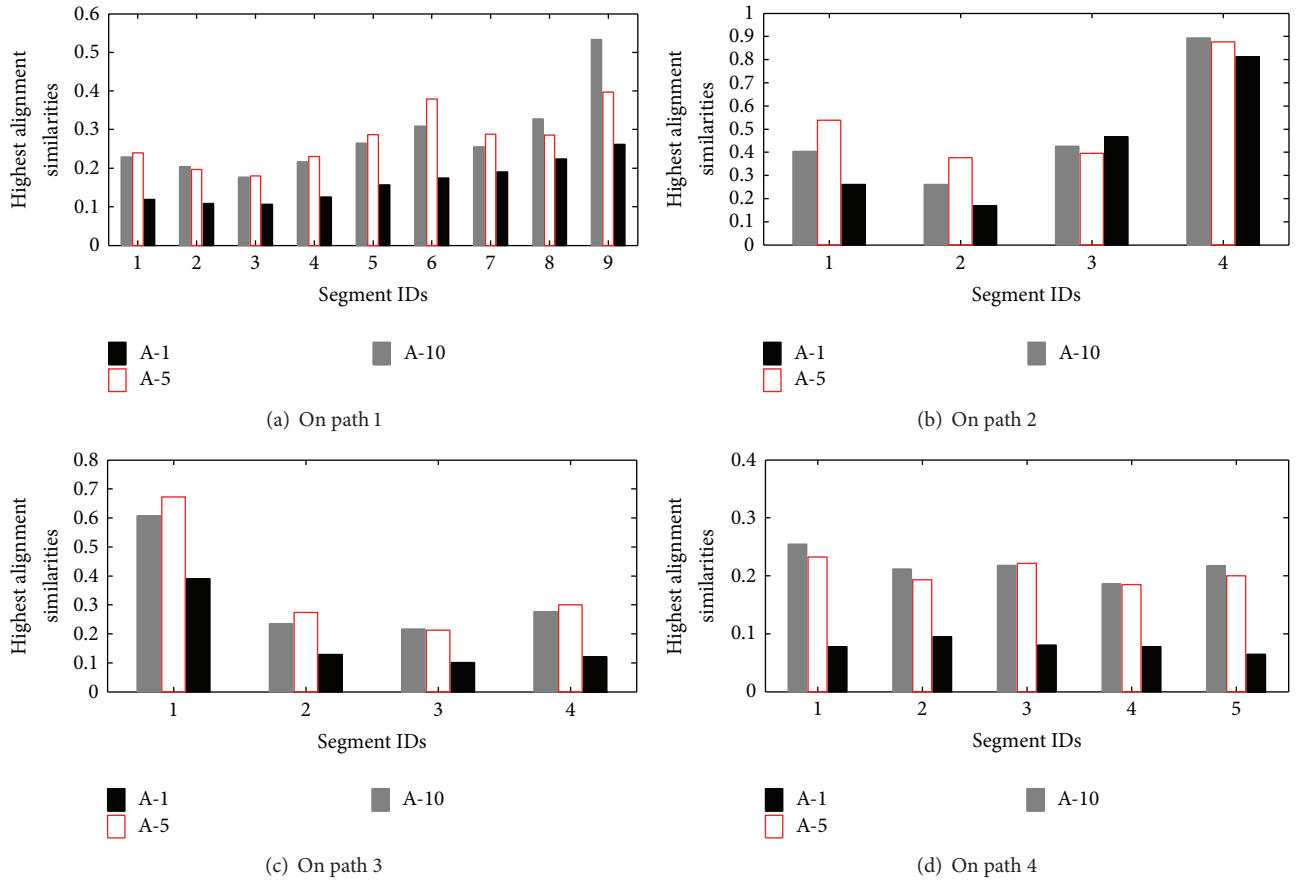
Figure 7: Alignment similarities on each path.

from massive and high-dimensional Wi-Fi RSS data which are sporadically recorded in large-scale area.

The significance of this paper is that both the RSS and time-stamp relations of the sporadically recorded Wi-Fi RSS data are used to conduct the crowd-sourced mobility mapping, and meanwhile alignment similarities between the newly recorded segments of RSS samples and prestored virtual fingerprints are considered to track the target locations. In future, the performance of integrating the time-stamped Wi-Fi RSS data with the data from inertial measurement unit (IMU) will be investigated to try to construct a more precise mobility map. For instance, the transitions among different LPs in mobility map can be better described by the moving speed and orientation of the target compared to the difference of time-stamps of RSS samples, and meanwhile more parameters may also be varied, such as the thresholds for scoring space construction and winning path creation in the intersequence assembling step, and also the threshold for alignment similarity calculation in sequence alignment algorithm.

Notations and Parameters

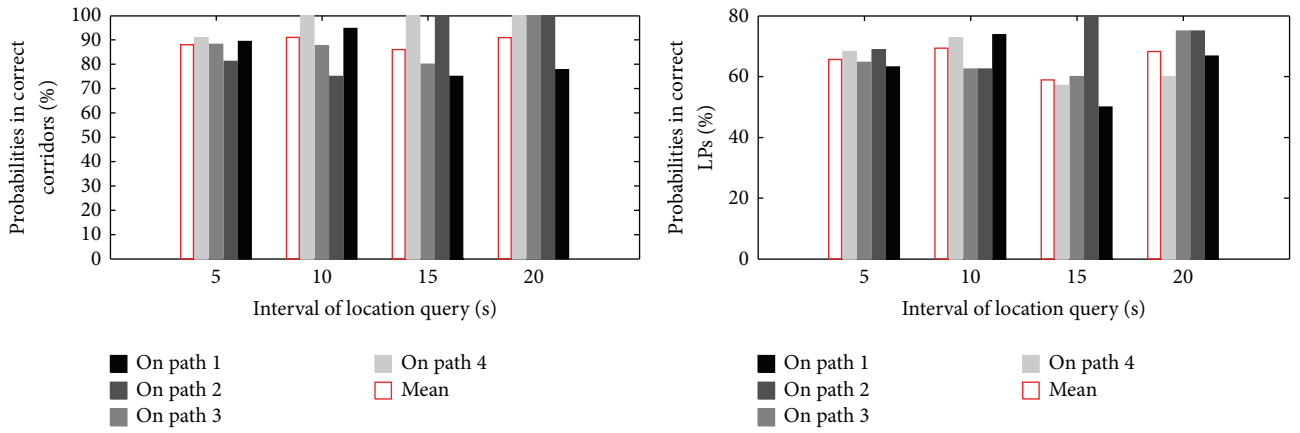
N:	Number of RSS sequences
N^ℓ and $N_k^{\ell,n}$:	Number of samples in R^ℓ and $R_k^{\ell,n}$
M:	Number of Aps
K:	Number of virtual fingerprints in each RSS cluster
R^ℓ and $\widehat{R}^\ell$:	ℓth RSS sequence of samples and clusters
Φ^ℓ:	Number of RSS clusters in R^ℓ
$\boldsymbol{\mu}_i^\ell$:	ith sample in R^ℓ
$\widehat{\boldsymbol{\mu}}_i^\ell$:	Sample mapped from $\boldsymbol{\mu}_i^\ell$
C_k^ℓ:	kth RSS cluster in R^ℓ
$C_{\widehat{k}}^\ell(\tau)$:	Tracked location for the τth location query
T_i^ℓ:	Time-stamp of $\boldsymbol{\mu}_i^\ell$ in R^ℓ
$\mu_{i,j}^\ell$:	RSS value from the jth AP in $\boldsymbol{\mu}_i^\ell$



(a) On path 1 (b) On path 2

(c) On path 3 (d) On path 4

Figure 8: Alignment similarities of the tracked locations on each path.



(a) Corridor-level accuracy (b) LP-level accuracy

Figure 9: Results of localization accuracy.

$W_{i,i'}$: Similarity between $\boldsymbol{\mu}_i^\ell$ and $\boldsymbol{\mu}_{i'}^\ell$

$R_k^{\ell,n}$: nth virtual fingerprint in C_k^ℓ

R_τ^{New}: Segment of samples used for the τth location query

$\Upsilon(C_s^\ell, C_t^{\ell'})$ and $\Gamma(C_s^\ell, C_t^{\ell'})$: Cumulative matching score and KL divergence between C_s^ℓ and $C_t^{\ell'}$

$\mathscr{A}_{\hat{k}}^{\hat{\ell},\tau}(\varepsilon_S)$: Set of tracked locations in ε_S-neighborhood of $C_{\hat{k}}^{\hat{\ell}}(\tau)$

δ: Sampling interval

α_R and α_T: Weighting factors for time-stamps and RSSs

F_M and F_D: Missing and damping factors

$\mathbf{R}^{K^\ell}$ and $\mathbf{R}^M$: K^ℓ and M-dimensional spaces.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors wish to thank the editor and reviewers for the careful review and valuable suggestions. This work was supported in part by the Program for Changjiang Scholars and Innovative Research Team in University (IRT1299), National Natural Science Foundation of China (61301126, 61471077), Special Fund of Chongqing Key Laboratory (CSTC), Fundamental and Frontier Research Project of Chongqing (cstc2013jcyjA40041, cstc2013jcyjA40032), Scientific and Technological Research Program of Chongqing Municipal Education Commission (KJ130528), Startup Foundation for Doctors of CQUPT (A2012-33), Science Foundation for Young Scientists of CQUPT (A2012-77), and Student Research Training Program of CQUPT (A2013-64).

References

[1] Y. Kim, S. Lee, and H. Cha, "A GPS sensing strategy for accurate and energy-efficient outdoor-to-indoor handover in seamless localization systems," *Mobile Information Systems*, vol. 8, no. 4, pp. 315–332, 2012.

[2] T. Cura, "A parallel local search approach to solving the uncapacitated warehouse location problem," *Computers and Industrial Engineering*, vol. 59, no. 4, pp. 1000–1009, 2010.

[3] Y. S. Lee, J. W. Park, and L. Barolli, "A localization algorithm based on AOA for ad-hoc sensor networks," *Mobile Information Systems*, vol. 8, no. 1, pp. 61–72, 2012.

[4] M. Zhou, A. K.-S. Wong, Z. Tian, V. Y. Zhang, X. Yu, and X. Luo, "Adaptive mobility mapping for people tracking using unlabelled Wi-Fi shotgun reads," *IEEE Communications Letters*, vol. 17, no. 1, pp. 87–90, 2013.

[5] C.-L. Wang, Y.-S. Chiou, and S.-C. Yeh, "A location algorithm based on radio propagation modeling for indoor wireless local area networks," in *Proceedings of the 61st IEEE Vehicular Technology Conference (VTC '05)*, vol. 5, pp. 2830–2834, IEEE, Stockholm, Sweden, May-June 2005.

[6] M. Zhou, Z. Tian, K. Xu, X. Yu, and H. Wu, "Error analysis for RADAR neighbor matching localization in linear logarithmic strength varying Wi-Fi environment," *The Scientific World Journal*, vol. 2014, Article ID 647370, 15 pages, 2014.

[7] P. Bahl and V. N. Padmanabhan, "RADAR: an in-building RF-based user location and tracking system," in *Proceedings of the 19th IEEE Conference on Computer Communications (INFOCOM '00)*, vol. 2, pp. 775–784, March 2000.

[8] M. Zhou, Z. Tian, K. Xu, X. Yu, and H. Wu, "Theoretical entropy assessment of fingerprint-based Wi-Fi localization accuracy," *Expert Systems with Applications*, vol. 40, no. 15, pp. 6136–6149, 2013.

[9] H. Shin, Y. Chon, and H. Cha, "Unsupervised construction of an indoor floor plan using a smartphone," *IEEE Transactions on Systems, Man and Cybernetics, Part C: Applications and Reviews*, vol. 42, no. 6, pp. 889–898, 2012.

[10] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Computation*, vol. 15, no. 6, pp. 1373–1396, 2003.

[11] T. F. Smith and M. S. Waterman, "Identification of common molecular subsequences," *Journal of Molecular Biology*, vol. 147, no. 1, pp. 195–197, 1981.

[12] M. Zhou, Z. Tian, K. Xu, X. Yu, X. Hong, and H. Wu, "SCaNME: location tracking system in large-scale campus Wi-Fi environment using unlabeled mobility map," *Expert Systems with Applications*, vol. 41, no. 7, pp. 3429–3443, 2014.

[13] M. Zhou, Z. Tian, K. Xu, H. Wu, Q. Pu, and X. Yu, "Construction of time-stamped mobility map for path tracking via smith-waterman measurement matching," *Mathematical Problems in Engineering*, vol. 2014, Article ID 673159, 17 pages, 2014.

[14] C.-C. Wang and C. Thorpe, "Simultaneous localization and mapping with detection and tracking of moving objects," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2918–2924, Washington, DC, USA, May 2002.

[15] B. Xi, R. Guo, F. Sun, and Y. Huang, "Simulation research for active simultaneous localization and mapping based on extended Kalman filter," in *Proceedings of the IEEE International Conference on Automation and Logistics (ICAL '08)*, pp. 2443–2448, September 2008.

[16] A. Chatterjee and F. Matsuno, "A neuro-fuzzy assisted extended kalman filter-based approach for simultaneous localization and mapping (SLAM) problems," *IEEE Transactions on Fuzzy Systems*, vol. 15, no. 5, pp. 984–997, 2007.

[17] C. D. Pathiranage, K. Watanabe, B. Jayasekara, and K. Izumi, "Simultaneous localization and mapping: a pseudolinear kalman filter (plkf) approach," in *Proceedings of the 4th International Conference on Information and Automation for Sustainability (ICIAFS '08)*, pp. 61–66, December 2008.

[18] K. Wang, L. Su, S. Wang, and Y. Yu, "Simultaneous localization and map building based on improved particle filter in grid map," in *Proceedings of the International Conference on Electronic and Mechanical Engineering and Information Technology (EMEIT '11)*, pp. 963–966, August 2011.

[19] Z. M. Wang, D. H. Miao, and Z. J. Du, "Simultaneous localization and mapping for mobile robot based on an improved particle filter algorithm," in *Proceedings of the IEEE International Conference on Mechatronics and Automation (ICMA '09)*, pp. 1106–1110, IEEE, Changchun, China, August 2009.

[20] R. C. Luo, K. Y. Chen, and M. Hsiao, "Visual simultaneous localization and mapping using stereo vision with human body elimination for service robotics," in *Proceedings of the IEEE/ASME Advanced Intelligent Mechatronics Conference*, pp. 244–249, July 2012.

[21] V. Sazdovski and P. M. G. Silson, "Inertial navigation aided by vision-based simultaneous localization and mapping," *IEEE Sensors Journal*, vol. 11, no. 8, pp. 1646–1656, 2011.

[22] S. Saeedi, L. Paull, M. Trentini, and H. Li, "Neural network-based multiple robot simultaneous localization and mapping," *IEEE Transactions on Neural Networks*, vol. 22, no. 12, pp. 2376–2387, 2011.

[23] Y. Yingmin and L. Ding, "Robot simultaneous localization and mapping based on non-linear interacting multiple model," in *Proceedings of the International Workshop on Intelligent Systems and Applications (ISA '09)*, pp. 1–6, May 2009.

[24] S. Yang, S. X. Yang, and L. Yang, "Method of improving WiFi SLAM based on spatial and temporal coherence," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '14)*, pp. 1991–1996, Hong Kong, China, May 2014.

[25] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, 2000.

[26] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.

[27] K. Rezahanjani, E. Dervishi, and J. Manzoor, *Smith-Waterman Sequence Alignment with MPI*, Group B6, WordPress, 2010.

[28] X. Wang, A. K. S. Wong, and Y. Kong, "Mobility tracking using GPS, Wi-Fi and cell ID," in *Proceedings of the 26th International Conference on Information Networking (ICOIN '12)*, pp. 171–176, February 2012.

[29] C. Feng, W. S. A. Au, S. Valaee, and Z. Tan, "Compressive sensing based positioning using RSS of WLAN access points," in *Proceedings of IEEE INFOCOM*, pp. 1–9, San Diego, Calif, USA, March 2010.