Computer Communications xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Computer Communications

journal homepage: www.elsevier.com/locate/comcom

# Crowdsensing-based Wi-Fi radio map management using a lightweight site survey

Yungeun Kim, Hyojeong Shin, Yohan Chon, Hojung Cha *

*Department of Computer Science, Yonsei University, Seodaemun-gu, Shinchon-dong 134, Seoul 120-749, Republic of Korea*

ARTICLE INFO

*Article history:*
Received 1 April 2014
Received in revised form 24 November 2014
Accepted 24 November 2014
Available online xxxx

*Keywords:*
Wi-Fi fingerprinting
War-walking
Smartphone
Crowdsensing
Radio map management

ABSTRACT

Localization based on Wi-Fi fingerprinting (WF) necessitates training the radio signals of target areas. Manual training enables good accuracy but requires service providers to conduct thorough site surveys to collect the radio signals of target areas periodically. Several systems are capable of eliminating the training phase by collecting radio signals from users, but these schemes are unable to provide location-based services until enough data are collected from the participatory users. Moreover, the accuracy of such systems is generally worse than that of systems that conduct manual training. In this paper, we propose a radio map management scheme in which the two methods are combined to achieve high accuracy with reduced management costs. The proposed scheme entails only a lightweight site survey for the construction of the initial radio map and does not necessarily require coverage of the entire area of interest. The quality of the radio map is enhanced in terms of both coverage and accuracy through user collaboration. In our system, mobile users conduct automatic war-walking with smartphone-based pedestrian dead reckoning (PDR), and to match the war-walking path to the radio map accurately, we employ a particle filter using both WF and PDR. We also consider the received signal strength variance problem caused by the device type and environmental changes. The proposed scheme is elastic since the service provider can adjust the costs required for the initial site survey depending on the quality of the crowdsensing-based radio map, which would compensate for the lack of coverage and accuracy of the initial radio map. The experiment's result validates that our scheme achieves competitive accuracy and coverage in comparison with systems that conduct full site surveys.



## 1. Introduction

The diverse location-based services (LBS) on offer today, such as location tagging on photos [1], life-logging [2], and social networking, provide convenient technology for people to use in their daily lives. With the widespread use of smartphones, localization is now available for both indoor and outdoor spaces. Most outdoor spaces are covered by the global positioning system (GPS) [3], but indoor LBS systems employ different localization techniques because GPS signals cannot penetrate indoor spaces. Indoor LBS solutions use a variety of sensors, such as radio-frequency identification [4], ultra-wideband [5], the global system for mobile communications [6], and Wi-Fi [7]. In particular, Wi-Fi fingerprinting (WF) has been extensively applied because this technology estimates indoor locations with meter-level accuracy. Moreover, additional infrastructures are not required because the system uses existing access points (APs) and most smartphones are Wi-Fi enabled.

The WF training process generally incurs high costs because the surveyor has to create a radio map by collecting received signal strength (RSS) information from every location in a target area. Periodic re-training is also inevitable because the characteristics of radio signals change over time. Much of the recent research on WF has focused on overcoming the problems presented by user collaboration. Park et al. [8] proposed WF systems that construct radio maps by collecting RSS measurements and location information that are manually provided by users. These systems require the active participation of users who have knowledge of the layouts of buildings and their current locations. We previously developed a WF system that constructs radio maps by user collaboration in a non-intrusive way [9]. In the system, mobile users report RSS measurements with location information tracked via pedestrian dead reckoning (PDR). The accuracy of the system is lower than that of a manual training system because of the drift error from PDR. Rai et al. [13] proposed Zee, which improves the accuracy of PDR with information extracted from floor plans, such as the locations of walls, rooms, and obstacles. However, detailed floor plans

* Corresponding author.
*E-mail address:* hjcha@yonsei.ac.kr (H. Cha).

http://dx.doi.org/10.1016/j.comcom.2014.11.006
0140-3664/

are not always available in practice, and the accuracy of PDR is low in large, open spaces due to the absence of walls and obstacles.

To achieve high-accuracy WF at a reduced cost, in this paper, we propose an elastic radio map management scheme that constructs a partial radio map using a lightweight site survey. In this system, the radio map is updated and expanded with PDR-based non-intrusive user collaboration. The proposed scheme is elastic since service providers can readily adjust the coverage of the initial site survey. Nevertheless, the proposed system presents a number of challenges as users collect radio fingerprints with different types of devices at different times. The system should therefore ideally consider the RSS variance problem [10] when the radio map is updated with user data. Since the initial map may not cover the entire area, the system should be able to add radio fingerprints for places that have not been covered with the correction of location errors from the smartphone-based PDR.

We overcame these challenges by developing an algorithm that combines WF and PDR with a particle filter. To solve the RSS variance problem, we considered particles based on a series of RSS measurements rather than a single RSS measurement. To reduce PDR errors, we designed a mobility model that has the bias error of a gyroscope at each turn point to reflect the characteristics of smartphone-based PDR. We also reduced the localization error caused by RSS variances by updating the radio map with a median fingerprint that was highly similar to all the other fingerprints collected at the same location.

The contributions of our work are as follows:

- We designed an elastic radio map management scheme that allows the coverage of a site survey to be adjusted.
- We proposed a radio map update algorithm that minimizes the RSS variance problem and the error distance of smartphone-based PDR.
- We validated the feasibility of our system with real experiments in different environments.

The remainder of this paper is structured as follows: Section 2 presents a background on the technologies used and an overview of the proposed system. Section 3 describes the radio map update algorithm, and Section 4 provides an evaluation of the system, which was achieved through experiments. Section 5 presents related work and Section 6 concludes the paper.

## 2. Background

### 2.1. Wi-Fi fingerprinting

WF involves two phases: a training phase and a localization phase. In the training phase, a radio map is constructed by a surveyor who conducts site surveys to collect the RSS measurements of all the locations in the target area. Radio map $M$ is represented as follows:

$$M = \{F_1, F_2, \ldots, F_{n-1}, F_n\}$$

where $F_i$ denotes the radio fingerprint of location $i$, and $n$ is the number of locations from which the radio fingerprints are collected during the training phase. The radio fingerprint $F$ includes the location information $L$ and the RSS measurements $R$. The location $L$ is represented by the coordinates of the $x$-axis and the $y$-axis in the 2D plane. $R$ contains the RSS measurements from all the $m$ APs deployed in the target area. Hence, $F$ is represented as

$$F = (L, R)$$

$$L = (x, y), \quad R = (\mathrm{rss}^1, \mathrm{rss}^2, \ldots, \mathrm{rss}^m)$$

In the localization phase, mobile users report their current RSS measurements $R'$ to the server. Then, the server identifies the $F_i$ that has an RSS measurement most similar to $R$. Several kinds of similarity functions can be used for the two RSS measurements; these functions include the Euclidian distance [7] and the Tanimoto coefficient [12]. In this paper, we use the Tanimoto coefficient because it ranges from 0 to 1, while the Euclidian distance increases depending on the number of APs. Finally, the user location is estimated by using the location information on the selected fingerprint.

### 2.2. Received signal strength variance problem

Generally, WF suffers from an RSS variance problem because RSS values are monitored in different ways depending on the device type and environmental changes. According to our previous study [10], the RSS values measured in a single location can differ by more than 10 dBm under different conditions. Consequently, the accuracy of WF is significantly degraded due to RSS differences. Solving the RSS variance problem is important in user-collaboration scenarios because users report their RSS measurements at different times with different devices. In [11], the effect of device diversity was addressed via kernel density estimation with a wide kernel width, but the RSS variance caused by other conditions, such as user direction and environmental changes, was not explored in the work.

### 2.3. Smartphone-based pedestrian dead reckoning

PDR is a localization technique that tracks relative locations by estimating the moving distance and direction of pedestrians. We previously proposed a smartphone-based system that implements PDR by using the sensors that are built into commercial smartphones, such as an accelerometer and digital compass [9]. Moving distance is estimated by step counting with the accelerometer and direction is estimated by the digital compass. However, the compass reading cannot be aligned with the user's direction because of the smartphone's placement and magnetic distortion. According to [13], the difference between the compass reading and an actual user's direction might exceed 150°. We therefore did not use a digital compass for heading estimation but employed a gyroscope instead. Since the gyroscope provides only relative angular change, the PDR periodically reported relative mobility information (RMI) as follows:

$$D = (d, \theta_{\mathrm{gyro}})$$

where $d$ represents the moving distance, and $\theta_{\mathrm{gyro}}$ the relative directional change estimated with a gyroscope.

### 2.4. Limitations of map-based pedestrian dead reckoning

The main problem with PDR is that the start position is unknown and the positioning error is accumulated with the drift caused by sensor noises. To overcome this problem, researchers proposed a map-based PDR that exploits the layout information, such as the positions of pathways, doors, and obstacles. In particular, Zee [13] used a map-based PDR for radio map building based on crowdsensing. In the system, the moving path is estimated by finding a path that does not violate the physical constraints imposed by the layout, as shown in Fig. 1(a). However, map-based PDR fails to determine the moving path when there are multiple paths that do not violate the physical constraint, as illustrated in Fig. 1(b). The probability of failure increases when a user takes a short path or where few obstacles or walls exist that impose physical constraints.



Please cite this article in press as: Y. Kim et al., Crowdsensing-based Wi-Fi radio map management using a lightweight site survey, Comput. Commun. (2014), http://dx.doi.org/10.1016/j.comcom.2014.11.006

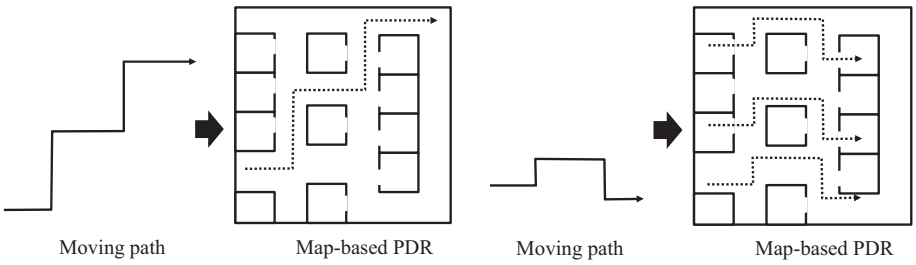



**Fig. 1.** Limitations of map-based PDR.

Moreover, it is not possible to assume that the floor plan of a building will always be available.

## 3. Radio map management scheme

### 3.1. Overview of the proposed scheme

The key feature of the proposed scheme is a hybrid mechanism that enables a combination of manual training and user collaboration to overcome the disadvantages of both approaches. Manual training is a costly process and periodic training is inevitable, while user collaboration may degrade tracking accuracy and cause initial service delays.

Fig. 2 illustrates the architecture of the proposed system, which comprises an LBS provider, a radio map server, and LBS users. The service provider first conducts a site survey to register the initial radio map on the server. In this process, the initial radio map may not cover the entire area and may consist only of the fingerprints of locations essential to the service. Such locations may include the entrances, main corridors, and hot locations in the target area. In this way, the proposed system enables the provision of LBS without having to wait for user contributions. The LBS users conduct non-intrusive war-walking while using the LBS offered by the server. They report RSS measurements with relative location information that is automatically tracked by smartphone-based PDR. When many users contribute war-walking data, the server updates the radio map to improve quality and coverage. The update process entails three phases: the path-matching phase, the accuracy-improving phase, and the coverage-expansion phase.

In the path-matching phase, the server matches each path on the radio map to a particle filter. The particle filter simulates the PDR noise model and estimates the optimal path that has the highest likelihood of RSS measurements on the radio map. After the optimal path has been estimated, the system determines if each fingerprint is at an existing location (EF) on the radio map.

In the accuracy-improving phase, the radio map is updated with the fingerprints of the existing location. The update algorithm chooses the representative fingerprint from among all the fingerprints accumulated at the same location. To overcome the RSS variance problem caused by device diversity and environmental changes, the fingerprint that has balanced similarity with all the other fingerprints is selected as the representative fingerprint.

In the coverage-expansion phase, the coverage of the radio map is expanded by using the fingerprints of new locations (NF). However, their location errors will be relatively larger compared to the fingerprints of the existing locations since there will be no similar fingerprints on the radio map. To solve this problem, the fingerprints of the new locations are clustered by comparing sequential RSS measurements. Then, the final new location is determined as the central location of the fingerprints in a group.

### 3.2. Path-matching phase

The war-walking data $W$ contributed by users include a series of RMI and RSS measurements. This is expressed as

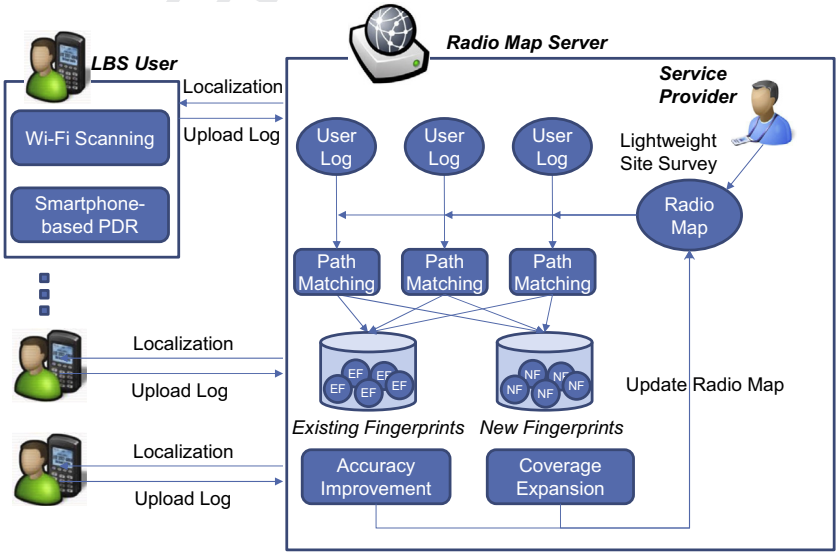



**Fig. 2.** Architecture of the proposed system.

Please cite this article in press as: Y. Kim et al., Crowdsensing-based Wi-Fi radio map management using a lightweight site survey, Comput. Commun. (2014), http://dx.doi.org/10.1016/j.comcom.2014.11.006

$$W = \{(D^{t_0}, R^{t_0}), (D^{t_1}, R^{t_1}), \ldots, (D^{t_e}, R^{t_e})\}$$

where $t_0$ and $t_e$ represent the start and end times of war-walking. $D^t$ and $R^t$ denote the RMI and RSS measurements at time $t$, respectively. Before applying the particle filter, we should decide the starting position on the path. We cannot start at time $t_0$ because uncertainty becomes high when the location at $t_0$ is not on the initial map. Hence, we divide the path $W$ into two sub-paths as

$$W_{\text{front}} = \{(D^{t_0}, R^{t_0}), (D^{t_1}, R^{t_1}), \ldots, (D^{t_s}, R^{t_s})\}$$

$$= \{(D^{t_s-1}, R^{t_s}), \ldots, (D^{t_1-1}, R^{t_1}), (D^{t_0-1}, R^{t_0})\}$$

$$W_{\text{tail}} = \{(D^{t_s}, R^{t_s}), (D^{t_{s+1}}, R^{t_{s+1}}), \ldots, (D^{t_e}, R^{t_e})\}$$

where $t_s$ is the time when the RSS measurement $R^{t_s}$ shows the highest similarity with the initial map. To start $W_{\text{front}}$ at time $t_s$, we reverse the order of $W_{\text{front}}$ and each item of RMI information. We then apply the update algorithm to each sub-path.

Updating the radio map with the war-walking path necessitates that the location of the RSS measurements in the data be accurately estimated. We estimate the accuracy of the locations by matching each sub-path on the radio map to a particle filter. The particle filter simulates the RMI noise models to identify an optimal path that shows the highest probability in the signal space. We conducted a preliminary experiment that showed the rationale for finding an optimal path based on the comparison of Wi-Fi fingerprints. We conducted war-walking four times in an office building and compared the RSS measurements between the war-walking and the initial radio map. As shown in Fig. 3(a), a higher fingerprint similarity was observed when the location error of PDR was smaller. Furthermore, the RSS variances between the paths were more stable than the RSS variances for each measurement, as is evident in Fig. 3(b). This means that the effect of RSS variance is reduced by using sequential Wi-Fi fingerprints on the trajectory. The results support the assertion that the optimal path close to the actual path exhibits high fingerprint similarity with the fingerprints on the initial map over the entire path. The implementation of the proposed particle filter is described as follows:

*Initialization.* To start, an initial location is roughly estimated by identifying the fingerprint $F^{\text{init}}$ that has the RSS measurement closest to $R^{t_0}$ in signal space. The actual location of $R^{t_0}$ could be far from the estimated location with $F^{\text{init}}$, because the radio map does not guarantee 100% coverage and the closest distance in signal space does not guarantee the closest distance in physical space because of the RSS variances. Hence, we calculate the maximum bound of the initial location from $L^{\text{init}}$ by using the log-distance path-loss model (LDPM) [31]. The LDPM is a radio propagation model that uses signal strength to predict the distance from an AP. The maximum distance between the estimated and actual locations is the sum of each distance from the AP. We calculate the maximum distance for each AP by the LDPM, and select the largest distance as the maximum bound from the estimated initial location. Then, we sample $K$ particles $\{p^k(t_0), k = 1 \ldots K\}$ according to the initial location and maximum bound. $p^k(t)$ represents the coordinates of the $k$-th particle at time $t$ as $p^k(t) = (x^k(t), y^k(t))$. Since we do not know the initial direction, the particle filter has to propagate particles in all directions. For each particle, $\theta^k_{\text{init}}$ is decided with the uniform distribution of $U(-180, 180)$.

*Propagation.* For each particle $p^k(t_j)$, a new particle $p^k(t_{j+1})$ is obtained on the basis of the RMI information at time $t_j$; thus:

$$p^k(t_{j+1}) = f(p^k(t_j), D^{t_j})$$

where $f$ represents the transition function, which is based on the RMI:

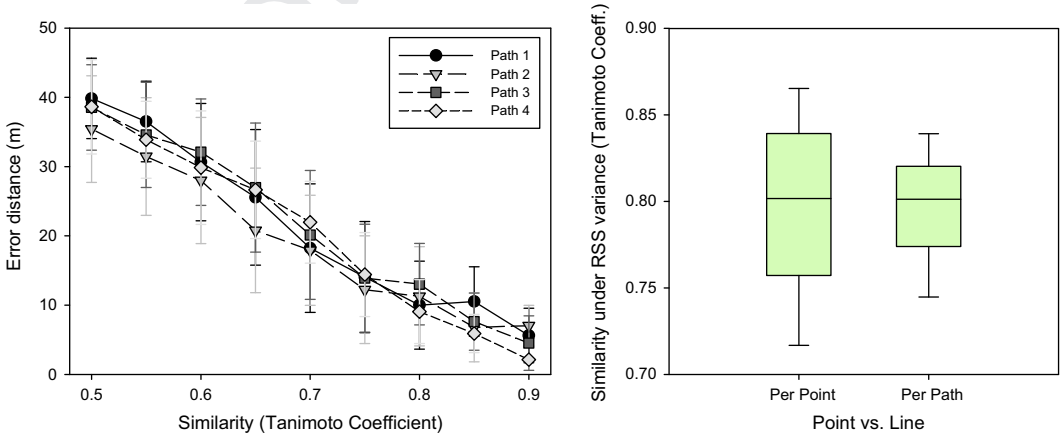
$$x^k(t_{j+1}) = x^k(t_j) + (d^{t_j} + \varepsilon_d^{t_j}) \cdot \cos\left(\theta^k_{\text{init}} + \theta^{t_j}_{\text{gyro}} + \varepsilon_\theta^{t_j}\right)$$

$$y^k(t_{j+1}) = y^k(t_j) + \left(d^{t_j} + \varepsilon_d^{t_j}\right) \cdot \sin\left(\theta^k_{\text{init}} + \theta^{t_j}_{\text{gyro}} + \varepsilon_\theta^{t_j}\right)$$

where $\varepsilon_d^{t_j}$ and $\varepsilon_\theta^{t_j}$ represent the PDR-estimated noise of the moving distance and direction, respectively. The performance of the particle filter is significantly affected by the modeling of such noise. To design the mobility model accurately, we investigated the noises of the smartphone-based PDR with real experiments. Fig. 4 illustrates the characteristics of the noises. We tracked a user walking along the dotted line two times. As shown in Fig. 4(b), the gyroscope rarely produced errors when the user walked forward in a straight line, but it produced large errors at each turn. This characteristic, also noted by Akiyama et al. [32], increases the number of particle filters to compensate for the increased heading error when pedestrians make a turn while walking. Considering this characteristic, we designed a RMI noise model as follows:

$$\varepsilon_\theta^{t_j} = \begin{cases} \varepsilon_\theta^{t_{j-1}} + U(-\alpha, \alpha), & \text{if } |\theta^{t_j} - \theta^{t_{j-1}}| > \theta_{\text{turn}} \\ \varepsilon_\theta^{t_{j-1}}, & \text{else} \end{cases}$$

$$\varepsilon_d^{t_j} = U(-\gamma \cdot d^{t_j}, \gamma \cdot d^{t_j})$$

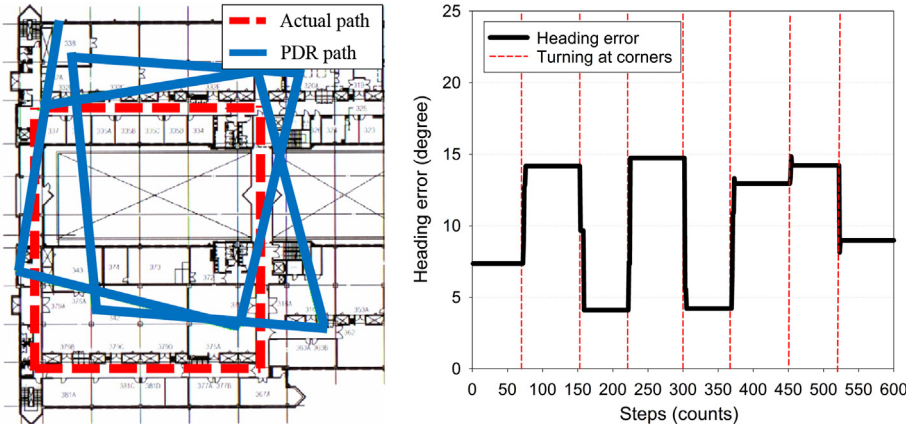



(a) PDR location error vs. similarity of Wi-Fi fingerprints

(b) RSS variances among the same points vs. the same paths

**Fig. 3.** Rationale for path matching based on Wi-Fi fingerprints.

Please cite this article in press as: Y. Kim et al., Crowdsensing-based Wi-Fi radio map management using a lightweight site survey, Comput. Commun. (2014), http://dx.doi.org/10.1016/j.comcom.2014.11.006




(a) Estimated path with PDR (b) Heading error of the gyroscope

**Fig. 4.** Characteristics of smartphone-based PDR using a gyroscope.

where $\gamma$ denotes the ratio of error distance to moving distance, and $\theta_{\text{turn}}$ stands for the threshold for detecting turn points. $\alpha$ represents the range of the heading error caused by turns.

*Importance sampling.* For each new particle $p^k(t_{j+1})$, we calculate weight $w^k(t_{j+1})$ on the basis of all the RSS measurements accumulated until time $t_{j+1}$; thus:

$$w^k(t_{j+1}) = \sum_{t=t_0}^{t_{j+1}} P(p^k(t), R^t)$$

where $P(p^k(t), R^t)$ is the probability that $R^t$ is observed at location $p^k(t)$. Since the radio map does not have the RSS measurements of all locations, we do not calculate $P(R^t|p^k(t))$ directly. Instead, we calculate $P(p^k(t), R^t)$ as follows:

$$P(p^k(t), R^t) = P\left(\left|\left|p^k(t) - L^{\text{best}}\right|\right| \left|\left|\text{Tanimoto}(R^t, R^{\text{best}})\right.\right)$$

where $L^{\text{best}}$ and $R^{\text{best}}$ represent the location and RSS information of the fingerprint that is closest to $R^t$ in signal space. $\text{Tanimoto}(R^t, R^{\text{best}})$ is the similarity between two RSS measurements calculated by the Tanimoto coefficient. That is, $P(p^k(t), R^t)$ is calculated based on the relationship of the physical distance and the RSS similarity between two RSS measurements.

*Normalization and resampling.* The weight of each particle is normalized. The particles are then duplicated on the basis of weight: the particles with low weight are deleted whereas those of high weight are duplicated.

*Finalization.* The previous steps are repeated until the end time of the war-walking $t_e$. The trajectory of the particle $k'$ that has the highest weight is selected as the optimal path:

$$k' = \underset{k}{\text{argmax}}\, w^k(t_e)$$

Then, the war-walking data $W$ are converted to $W'$, which includes the series of fingerprints:

$$W' = \left\{p^{k'}(t_0), \ldots, p^{k'}(t_e), R^{t_0}, \ldots, R^{t_e}\right\} = \left\{F^{t_0}, \ldots, F^{t_e}\right\},$$
$$F^t = \left(x^{k'}(t), R^t\right)$$

Finally, each fingerprint in $W'$ is classified as either an EF or an NF (Fig. 5). The fingerprints that overlap on the radio map are considered as EFs; otherwise, they are regarded as NFs.

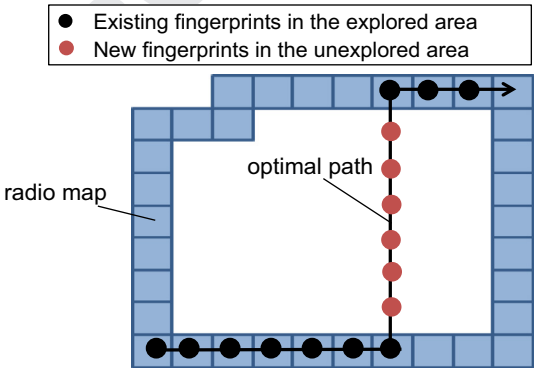



**Fig. 5.** Classification of EF and NF after the path-matching step.

### 3.3. Accuracy-improvement phase

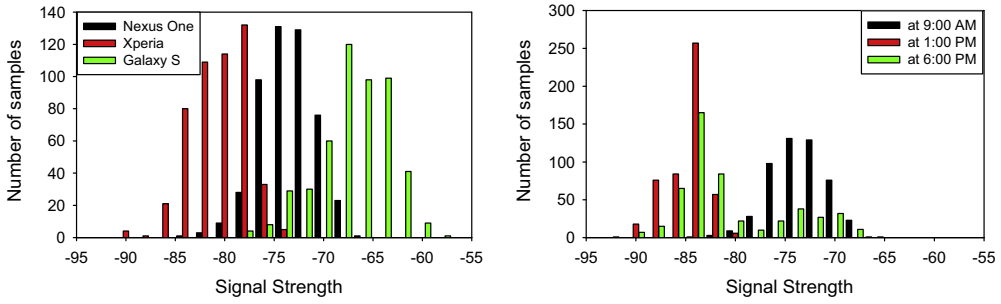
After the optimal path has been identified for all the war-walking data, multiple fingerprints are accumulated at the same location in the radio map as follows:

$$S_{\text{EF}}^i = \left\{F_i^1, \ldots, F_i^l\right\}$$

where $S_{\text{EF}}^i$ represents the set of $l$ fingerprints at the $i$-th location in the radio map. In the EF update, the new fingerprint should be produced using these fingerprints. The simple solution is to average all the RSS values for each AP. However, this approach is an ineffective solution for the proposed system because the RSS values collected by multiple users do not follow a Gaussian distribution. Fig. 6 illustrates the RSS values during 500 scans of the same location. RSS values vary depending on the device type and time differences. The extent of the RSS variance is expected to increase during user collaboration because users employ different devices and are located in different environments. Considering the RSS variance problem, we opt to select a representative fingerprint that makes up the RSS variances rather than to produce a new fingerprint by merging the fingerprints. The representative fingerprint $F_i'$ is selected as follows:

$$F_i' = (L_i, R_i')$$

$$R_i' = \underset{R_i^c}{\text{argmax}} \sum_{j=1}^{l} \text{Tanimoto}\left(R_i^c, R_i^j\right)$$




(a) RSS variances caused by device diversity (b) RSS variances caused by time differences

**Fig. 6.** RSS variances in the same location.

where $F'_i$ is the fingerprint that has the RSS measurement $R'_i$ showing the largest similarity with all other RSS measurements in location $L_i$.

### 3.4. Coverage-expansion phase

After the optimal paths have been identified for all the war-walking data, there may be NFs that have been collected at locations that the radio map does not cover. The system should then add these fingerprints to the radio map to increase the LBS coverage. However, the location error of the NFs will be relatively larger than that of the EFs because the initial radio map does not contain the fingerprints collected at the NFs' locations. Consequently, the NFs collected at the same location are spread out because of the error distance of the estimated location. To solve this problem, the NFs at the same location should be clustered by comparing their RSS measurements. This process is difficult because RSS measurements at the same location can differ significantly from one another due to the RSS variance problem. To reduce the effect of the RSS variance when clustering, we conduct clustering not by comparing each RSS measurement, but rather by comparing the set of fingerprints that were collected on the same trajectory. In this case, the effect of the RSS variance decreases because the average RSS variance over the entire trajectory is considerably more stable than for that of a single location. This is illustrated in Fig. 3(b). To cluster trajectories with similar RSS measurements, we convert the set of fingerprints in $W'$ to a set of lines that includes sequential fingerprints. The conversion is performed using the Ramer–Douglas–Peucker algorithm [14], a line-simplification algorithm. Then, the set of NFs for all the war-walking data $S_{NF}$ is represented as a set of $q$ line segments:

$$S_{NF} = \{V^1, \ldots, V^q\}, V = \{F_1, \ldots, F_v\}$$

where $V$ is the line segment that includes the set of sequential fingerprints. We then identify the same lines by $k$-means clustering [15], which minimizes the objective function $J$:

$$J = \sum_{i=1}^{k}\sum_{j=1}^{q}\left\|V_i^j - V_i^c\right\|$$

where $V_j^i$ and $V_c^i$ represent the $j$-th line segment and center line segment of the $i$-th cluster, respectively. $\left\|V_i^j - V_i^c\right\|$ stands for the distance between two line segments in the signal space as follows:

$$\|V^1 - V^2\| = \max_{0\le k\le v^2 - v^1} \frac{\sum_{j=1}^{v^1} \text{Tanimoto}\left(R_j^1, R_{j+k}^2\right)}{v^1}, v^1 < v^2$$

Here, $v^1$ and $v^2$ denote the number of fingerprints in each line. The distance between two lines in the signal space is estimated by averaging the similarity between each pair of RSS measurements. Line lengths can vary; thus, we determine distance by identifying the maximum average similarity, which is ascertained by moving the short line along the long line.

After identifying the same lines by clustering, a representative line is selected to determine the location of the NFs. In the proposed system, the line segment that has the largest weight in the path-matching step is selected as the representative line. Finally, we select the representative fingerprint of the NFs at the same location in the same way as in the EF update.

## 4. Evaluation

We first investigated the optimal parameters for the noise model of the proposed PDR. Then, we evaluated the proposed system in two different environments: an office building and a shopping mall. In the office building, we were unable to enter the offices themselves due to security issues. Hence, we evaluated the proposed system only via the paths along the corridors of the office building. We further evaluated the proposed system in a shopping mall where the users' paths covered the areas inside the shops as well as the corridors. We collected data for a single day in the office building, whereas data were collected for over a week in the shopping mall.

### 4.1. Noise model of smartphone-based pedestrian dead reckoning

To find the optimal parameters for the noise model, we conducted smartphone-based PDR 30 times with three users and investigated the noise characteristics. As shown in Table 1, step counting showed high accuracy regardless of the user. The error ratio of the moving distance was slightly larger than that of step counting due to stride-length error. In the case of heading estimation, errors in the straight sections were negligible. However, the gyroscope produced a large error of up to 28° when the users made turns. Based on the result, we fixed the parameters $\alpha$, $\gamma$ at 30° and 10%, respectively.

### 4.2. Experiment in the office building

*Experiment setup*. The test environment was an office building with dimensions of 100 m × 50 m and 58 APs. Fig. 7 shows the floor plan. We constructed an initial radio map by manual training, which was intended to collect RSS measurements at the locations marked with a solid line. We mobilized three participants to collect war-walking data at any time along any path on the red[1] line for a single day. In total, 94 sets of war-walking data were collected. Three devices were used for war-walking and the war-walking

[1] For interpretation of color in Fig. 7, the reader is referred to the web version of this article.

Please cite this article in press as: Y. Kim et al., Crowdsensing-based Wi-Fi radio map management using a lightweight site survey, Comput. Commun. (2014), http://dx.doi.org/10.1016/j.comcom.2014.11.006

**Table 1**
Characteristics of noise in smartphone-based PDR.

| User | Moving distance | | Direction | |
|---|---|---|---|---|
| | Step counting error ratio (%) | Distance error ratio (%) | Average error in a straight section (maxima) | Average error in a turning section (maxima) |
| User A | 4.2 | 8.7 | 1.5 (3.2) | 8.9 (21.2) |
| User B | 1.6 | 2.5 | 2.1 (4.4) | 10.375 (25.3) |
| User C | 2.5 | 3.0 | 1.8 (4.5) | 13.1 (28.4) |

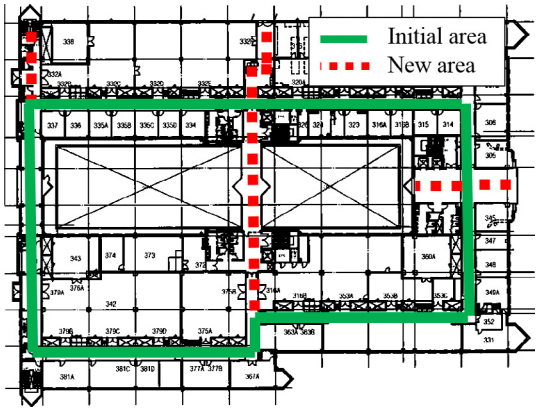


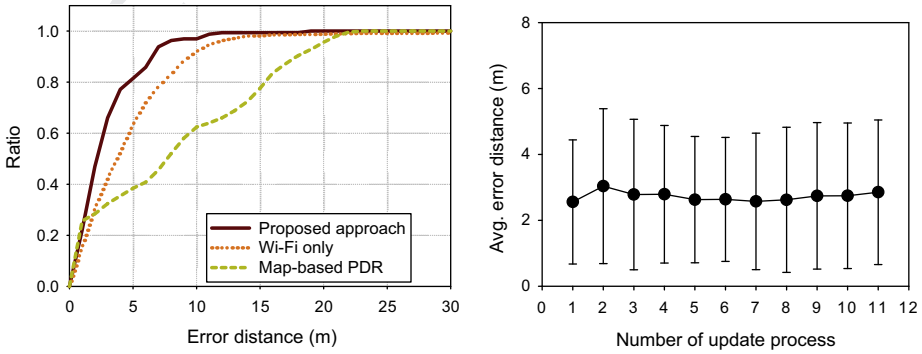

**Fig. 7.** Layout of the test environment on the third floor of an office building.

duration was evenly distributed from morning to night. Each war-walking data set contained not only consecutive RSS measurements and RMIs, but also the real trajectory of the war-walking. To record the real trajectory, the participants inputted tagging points on the planned path of the map before war-walking. The participants then pushed the tagging button whenever they passed each tagging point during war-walking. We used this information to evaluate the proposed system rather than to update the algorithm. The dotted lines represent the paths that were taken by the participants but not covered by the manual training. Due to security issues, we could not visit the offices themselves.

*Path-matching performance*. For each war-walking data set, we matched the paths on the radio map with a particle filter, and evaluated the distance between the estimated and real locations for each RSS measurement. To evaluate the accuracy of the path matching, we compared the proposed matching algorithm with two different algorithms: (1) the Wi-Fi-only method, which estimates the location of each RSS measurement by identifying the most similar fingerprints on the radio map; and (2) the map-based PDR used in Zee [13], which exploits layout information, such as the locations of walls and obstacles. Fig. 8 depicts the cumulative distribution function (CDF) of the error distance of each fingerprint after the path matching. Our approach outperformed the Wi-Fi-only method, indicating that the proposed particle filter decreases the WF error caused by RSS variances. The particle filter with map information showed poor performance because there were many possible paths due to the symmetric structure of the building. We further investigated if the error distance of the proposed path-matching method diverged due to the error of PDR when the update process was conducted repetitively. As shown in Fig. 8(b), the error distance did not diverge even when the update process was repeated multiple times. This occurred for two reasons. First, the error on path matching was further reduced by updating the map with only the representative fingerprint, which had a lower path-matching error than the other fingerprints; the path-matching precision of the representative fingerprints was 70% and their average error distance was only 1.7 m. Second, the cumulative errors with the repetitive update processes were canceled by particle filtering with sequential Wi-Fi fingerprints. The experiment's results validated that the proposed path-matching method is accurate and resilient to an erroneous PDR trajectory.

*Wi-Fi fingerprinting with the updated radio map*. To evaluate the accuracy of the updated radio map, we conducted a radio map update using 50% of the war-walking data then estimated the localization errors with the remaining 50%. We compared the accuracy of the initial and updated radio maps. Fig. 9(a) shows the CDF of the error distances when a user was at the locations of the EFs. The updated radio map yielded improved results: the ratio of errors lower than 3 m increased from 55% to 69% and the average error distance decreased by 20% from 3.2 to 2.6 m. These results indicate that the proposed update algorithm enhances the tolerance of radio maps against RSS variances.

We also evaluated the localization performance at the newly added locations. Fig. 9(b) illustrates the result of the WF-based localization at the newly added locations. The result derived with the initial radio map was characterized by a large error distance

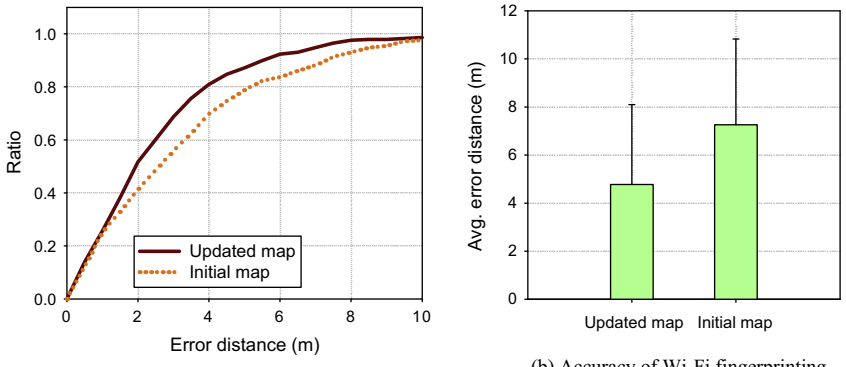



(a) Accuracy of the path matching according to the method

(b) Stable matching performance with repetitive update processes

**Fig. 8.** Accuracy of the path matching.

Please cite this article in press as: Y. Kim et al., Crowdsensing-based Wi-Fi radio map management using a lightweight site survey, Comput. Commun. (2014), http://dx.doi.org/10.1016/j.comcom.2014.11.006




(a) Accuracy of Wi-Fi fingerprinting at existing locations

(b) Accuracy of Wi-Fi fingerprinting at newly added locations

**Fig. 9.** Quality of the updated radio map.

of 7.2 m because the initial map did not contain the fingerprints collected at the new locations. With the updated radio map, users at the new locations had an average error distance of 4.5 m.

### 4.3. Experiment in the shopping mall

*Experiment setup.* We evaluated the proposed system in a shopping mall, which exhibited different characteristics compared to the office building, as shown in Fig. 10(b). An initial radio map was constructed by collecting fingerprints along the main corridor, which is marked with solid lines in Fig. 10(a). To evaluate the proposed system under long-term RSS variances, we conducted war-walking for a week. Compared to the office building, there were significantly more people in the shopping mall, and the density varied according to the time slot. We also increased the number of devices from three to six. Since we had free access to the stores, the war-walking paths covered not only the corridors, but also the areas inside the stores. In total, 100 sets of war-walking data were collected. We used a half of them for the radio map update and the remainder for the localization test.

*Effect of the received signal strength variance problem.* We first investigated the RSS variances caused by the different environments and device types during data collection. Fig. 11 shows the accuracy degradation due to RSS variances according to time and device type. "No diff." denotes the result when the RSS measurements were collected with the same device at the same time slot during training and localization. "Device diff." indicates the result when different devices were used and "Time diff." represents the localization error caused by time differences. "Device & Time diff." indicates that there were different device types and time slots for

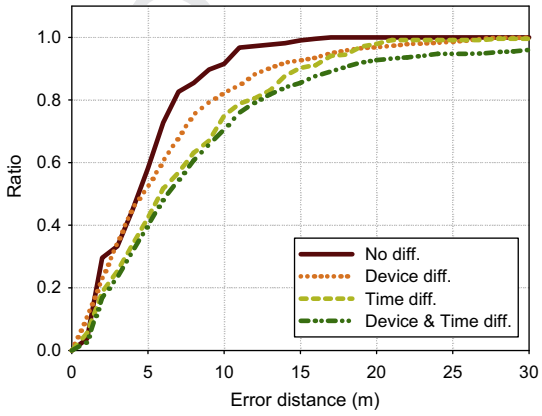
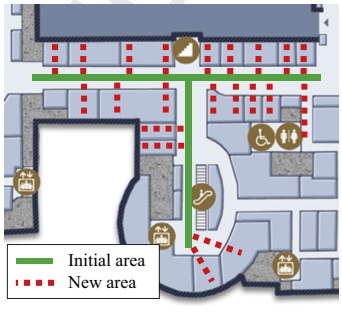



**Fig. 11.** Effect of the RSS variance problem in the shopping mall.

training and localization. For Time diff. and Device diff., the ratios of the errors larger than 15 m were 7% and 10%, respectively. In the case of Device & Time diff., the ratio was 15%. By contrast, the ratio of errors larger than 15 m for No diff. was 1%. The average error distances were 4.7 m, 6.1 m, 7.4 m, and 8.7 m, respectively. These results show that there was a significant RSS variance in the shopping mall, and this consequently increased the percentage of long-tail errors and threatened the reliability of the LBS.

*Path-matching performance.* Fig. 12(a) shows the average error distance of each fingerprint after the path matching. Similar to the office building result, our approach outperformed the Wi-Fi-only method and the particle filter with map information. The



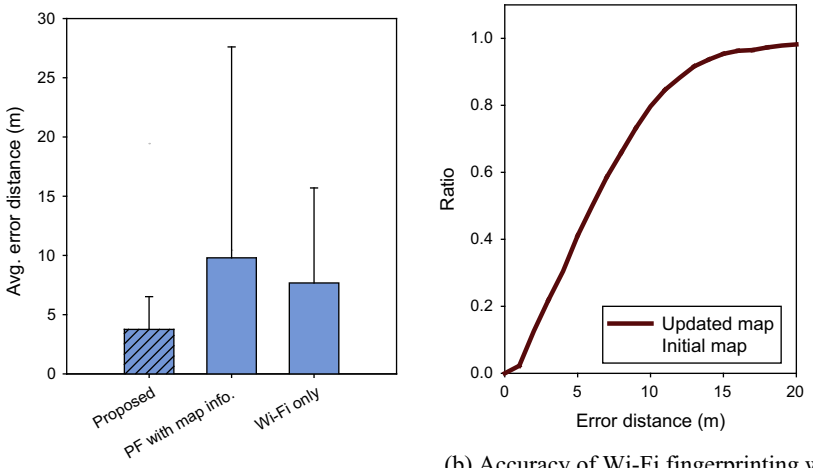

(a) Layout of the floor

| | *Office* | *Mall* |
|---|---|---|
| *Duration* | Single day | One week |
| *Population* | Dozens | Hundreds |
| *Devices* | 3 | 6 |
| *Coverage* | Corridors | Corridors and stores |

(b) Difference between the office building and the shopping mall

**Fig. 10.** Experimental environment in the shopping mall.

Please cite this article in press as: Y. Kim et al., Crowdsensing-based Wi-Fi radio map management using a lightweight site survey, Comput. Commun. (2014), http://dx.doi.org/10.1016/j.comcom.2014.11.006




(a) Accuracy of the path matching

(b) Accuracy of Wi-Fi fingerprinting with the updated radio map

**Fig. 12.** Performance of the proposed update algorithm.

average error distances were 3.75 m, 16.5 m, and 7.67 m, respectively. The proposed particle filter showed a 51% reduced error over the WF. The particle filter with map information also showed poor performance because the shopping mall had wide corridors, and there were several open spaces where walls and obstacles did not exist.

*Wi-Fi fingerprinting with the updated radio map*. Fig. 12(b) shows the result of WF-based localization with the initial map and the updated map. The localization with the updated map was more accurate than that with the initial map. Note that the ratio of error larger than 15 m was reduced from 17% to 7%. This means that the proposed update algorithm enhanced the reliability of the WF by reducing the occurrence ratio of the long-tail errors. We previously proposed a clustering-based radio map management scheme [33] that stores multiple fingerprints per location to overcome the RSS variance problem. Fig. 13 shows the Wi-Fi localization accuracy according to the updated map. The *Updated Map-n* represents the radio map that contains *n* representative fingerprints per location with the clustering-based radio map management. As more fingerprints were used for a single location, the accuracy and reliability of the WF increased. Since using multiple fingerprints per location increases the size of the radio map, the number of fingerprints should be carefully selected considering both the RSS variance problem and the scalability of the radio map database.

*Coverage and accuracy according to the volume of data used for the map update*. Fig. 14 shows the coverage and accuracy of the updated radio map according to the volume of data used for the update. There are several points in the figure showing accuracy degradation. This occurred when new data were added, but the amount of degradation was negligible. From a broad perspective, the average error distance of localization in the updated map decreased as the volume of used data increased. The effect of the RSS variance problem was reduced by selecting representative fingerprints from among the fingerprints collected in diverse situations. The coverage also increased as the volume of data increased. The results validate that crowdsensing-based radio map management is necessary to both maintain high accuracy and expand coverage.

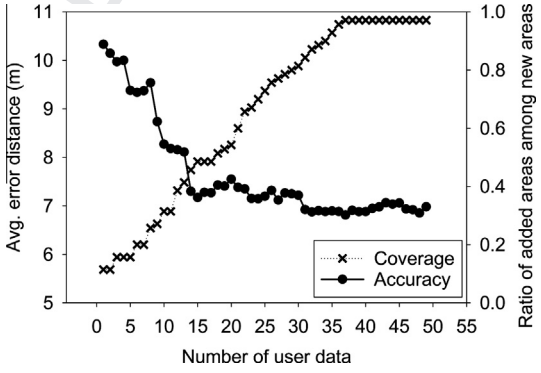



**Fig. 14.** Coverage and accuracy according to the data for the number of users.

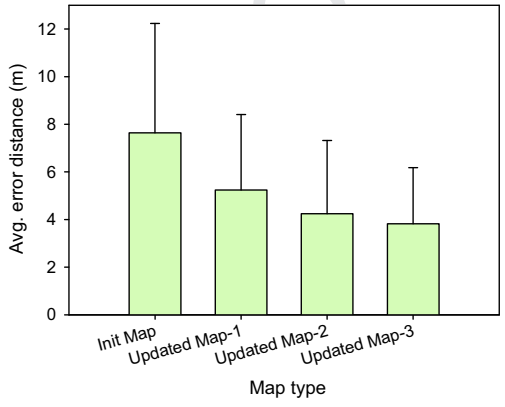
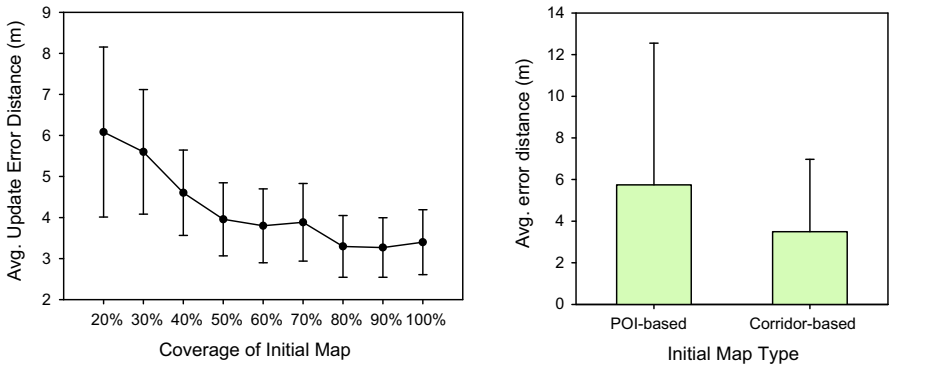



**Fig. 13.** Further enhancement with clustering-based radio map management.

*Condition of initial map vs. update performance*. To validate the elasticity of the proposed scheme, we investigated the path-matching accuracy according to the conditions of the initial site survey. We first investigated the relationship between the coverage of the initial map and the update performance. We constructed a radio map that covered both the initial area and the new area, as shown in Fig. 10. We then conducted the update process while changing the coverage of the radio map from 20% to 100%. Partial radio maps were constructed by choosing fingerprints randomly according to the given coverage. We repeated the experiment 100 times for each coverage. As shown in Fig. 15(a), the update performance was poor when the coverage of the initial map was lower than 30%. This was expected since there were few fingerprints with which the PDR error distance could be reduced. It is noteworthy that the update

Please cite this article in press as: Y. Kim et al., Crowdsensing-based Wi-Fi radio map management using a lightweight site survey, Comput. Commun. (2014), http://dx.doi.org/10.1016/j.comcom.2014.11.006




(a) Effect of the coverage of the initial map (b) Effect of the initial site-survey strategy

**Fig. 15.** Path-matching accuracy according to the conditions of the initial radio map.

accuracy improved significantly as the coverage increased. With only 50% coverage, the proposed update scheme achieved similar accuracy to 100% coverage. The result validated that the proposed scheme allows service providers to adjust the coverage of the site survey, allowing for due consideration of both accuracy and cost. We further investigated the effect of the initial map-construction strategy to the update performance. We constructed two types of initial maps: a point-of-interest (POI)-based map and a corridor-based initial map. The POI-based map contained Wi-Fi fingerprints at POIs, and the corridor-based map contained the fingerprints along the main corridors. As shown in Fig. 15(b), the accuracy of the path matching with the POI-based map was worse than that with the corridor-based map. This result was considered reasonable because most of the user movements during war-walking were along the corridors rather than at the POIs.

## 5. Related work

Active research on WF systems has been conducted recently. RADAR [7], which requires a training phase and a localization phase, was the first such system. In the training phase, a radio map is constructed by measuring RSSs from existing APs at all locations. In the localization phase, location is estimated by the *k*-nearest neighbor algorithm, which identifies the RSS vector that has the closest Euclidian distance to the currently observed RSS vector. Numerous WF systems have been proposed to eliminate the disadvantages of RADAR by improving localization algorithms, reducing the costs incurred in the training phase, and rectifying the RSS variance problem. We provide an overview of the key research on each improvement approach.

### 5.1. Schemes that improve localization algorithms

Brunato [16] introduced a location discovery technique based on a support vector machine. In this system, a regression engine tracks a mobile user, and a classification engine determines the room where the user is currently located. Nibble [17] is a probabilistic location service that uses Bayesian networks to infer the location of a device. Horus [18] reduces computational overheads in a localization algorithm by clustering locations that share a common set of APs. Kushki et al. [19] proposed a kernelized distance calculation algorithm to compare two RSS measurements. They also developed an AP filtering algorithm [20] that chooses APs whose signal strengths remain fixed over time but vary considerably over space.

### 5.2. Schemes that reduce training phase costs

Oliver et al. [21] proposed an automatic site-survey system that performs tracking with a wearable location system. The organic location system [8] constructs radio maps by user collaboration. In the system, users manually input their locations. The EZ localization system [22] was proposed to enable localization without manual training. In the system, numerous RSS measurements are collected without location information from anonymous users. The server then estimates the optimal location of each measurement by imposing constraints on the physics of wireless propagation. The SENIL localization system [30] builds a radio map automatically with indoor floor plans and numerous RSS samples collected by crowdsensing. SENIL matches the sequential RSS samples to the indoor floor plan using a graph-matching algorithm. Kim et al. [9] proposed a smartphone-based autonomous war-walking system in which an accelerometer and digital compass are used. However, the system has the limitation that an initial location and direction need to be given. Zee [13] and WILL [29] overcame this limitation by exploiting the information extracted from floor plans, such as the locations of walls, rooms, and obstacles. However, these systems have the potential to fail to infer user locations in large, open spaces where no obstacles or walls exist. Wang et al. [28] proposed an unsupervised localization system where users collect landmarks that each have a unique sensor signature at a certain location, e.g., a unique set of APs or magnetometer distortion. These landmarks are used to improve the smartphone-based PDR. To start the system, several landmarks are required, such as the locations of elevators, stairs, and entrances.

### 5.3. Schemes that solve the received signal strength variance problem

Haeberlen et al. [23] developed a manual calibration method that identifies the linear relationship between different devices. This approach is accurate but impractical because of scalability issues. Kjærgaard et al. [24] proposed an automated method that calibrates the RSS variance in different devices during the online learning phase. This method solves the scalability issue but exhibits less accurate performance than that achieved with manual calibration. The authors also proposed a localization algorithm [25] that uses the ratio of RSS values as a fingerprinting component instead of the absolute RSS values. Tsui et al. [26] put forward an unsupervised learning system that learns the relationship between two different Wi-Fi devices during localization. In the system, the current location is roughly estimated with the Pearson product-moment correlation coefficient, and the expectation maximization algorithm

is then applied to identify the optimal transformation function. Park et al. [11] reduced the RSS difference among devices by wide smoothing of the signal strength distribution function. Kim et al. [10] proposed a localization algorithm that uses the location of the RSS peak as a fingerprinting component. They considered device diversity as well as environmental changes, such as time, direction, and device placement.

## 6. Conclusion

In this paper, we proposed an elastic radio map management scheme in which radio maps are automatically updated and expanded by employing a combination of manual training and the user-collaboration approach. LBS users non-intrusively contribute their war-walking data with smartphone-based PDR. With the proposed scheme, service providers do not need to conduct a thorough site survey and instead just conduct a lightweight site survey that covers essential locations. We also proposed an accurate radio map update algorithm that overcomes the RSS variance problem. The effect of RSS variances is reduced by particle filtering with PDR and WF. The RSS variance problem is solved through the selection of representative fingerprints that are highly similar to other fingerprints at the same location. We evaluated the proposed algorithm by conducting thorough experiments in two different environments. The experiment's results validated the high accuracy and reduced costs presented by the system.

In future work, we intend to update radio maps by exploiting the RSS measurements reported by location-based social-networking service applications, such as LifeMap [2] and FourSquare [27]. The applications can provide accurate location information that is directly obtained by users. We anticipate that we will be able to use this information to construct more accurate radio maps and reduce the effect of RSS variances.

## Acknowledgements

Q3 This work was supported by a grant from the National Research
Q4 Foundation of Korea (NRF), funded by the Korean government, Ministry of Education, Science and Technology under Grant (No. 2013-027363).

## References

[1] K. Toyama, R. Logan, A. Roseway, Geographic location tags on digital images, in: Proc. 11th ACM Int. Conf. Multimedia, 2003, pp. 156–166.
[2] J. Chon, H. Cha, LifeMap: smartphone-based context provider for location-based services, IEEE Pervasive Comput. 10 (2) (2011) 58–67.
[3] E.D. Kaplan, C.J. Hegarty, Understanding GPS: Principles and Applications, Artech House, Boston, MA, 1996.
[4] M. Addlesee, R. Curwen, S. Hodges, J. Newman, P. Steggles, A. Ward, A. Hopper, Implementing a sentient computing system, Computer 34 (2001) 50–56.
[5] C. Zhang, M. Kuhn, B. Merkl, A.E. Fathy, M. Mahfouz, Accurate UWB indoor localization system utilizing time difference of arrival approach, in: IEEE Radio Wireless Symp., 2006, p. 515.
[6] V. Otsason, A. Varshavsky, A. LaMarca, E. de Lara, Accurate GSM indoor localization, UbiComp 2005, Lecture Notes Computer Science, vol. 3660, Springer-Verlag, 2005, p. 141.
[7] P. Bahl, V.N. Padmanabhan, RADAR: an in-building RF-based user location and tracking system, in: Proc. 19th IEEE Int. Comput. Commun. (InfoCom), vol. 2, 2000, pp. 775–784.
[8] J. Park, B. Charrow, D. Curtis, J. Battat, E. Minkov, J. Hicks, S. Teller, J. Ledlie, Growing an organic indoor location system, in: Proc. Int. Conf. Mobile Systems, Applications, and Services (MobiSys), San Francisco, CA, 2010, pp. 271–284.
[9] Y. Kim, Y. Chon, H. Cha, Smartphone-based collaborative and autonomous radio fingerprinting, IEEE Trans. Syst., Man., Cybern., Part C: Appl. Rev. 42 (1) (2012) 112–122.
[10] Y. Kim, H. Shin, H. Cha, Smartphone-based Wi-Fi pedestrian-tracking system tolerating the RSS variance problem, in: Proc. Int. Conf. on Pervasive Comput. and Commun. (PerCom), Lugano, Switzerland, 2012, pp. 11–19.
[11] J. Park, D. Curtis, S. Teller, J. Ledlie, Implications of device diversity for organic localization, in: Proc. IEEE Infocom, 2011, pp. 3182–3190.
[12] P. Jaccard, The distribution of the flora in the alpine zone, New Phytologist 11 (2) (1912) 37–50.
[13] A. Anshul Rai, K.K. Chintalapudi, P. Venkat, R. Sen, Zee: zero-effort crowdsourcing for indoor localization, in: Proc. 18th Annu. Int. Conf. Mobile Computing and Networking (MobiCom), 2012.
[14] D.H. Douglas, T.K. Peucker, Algorithms for the reduction of the number of points required to represent a digitized line or its caricature, Can. Cartogr. 10 (2) (1973) 112–122.
[15] J. MacQueen, Some methods of classification and analysis of multivariate observations, in: L.M. LeCam, J. Neyman (Eds.), Proc. 5th Berkeley Symp. Math., Stat., and Prob., U. California Press, Berkeley, CA, 1967, p. 281.
[16] M. Brunato, R. Battiti, Statistical learning theory for location fingerprinting in wireless LANs, Comput. Netw. 47 (6) (2005) 825–845.
[17] P. Castro, P. Chiu, T. Kremenek, R. Muntz, A probabilistic room location service for wireless networked environments, in: Proc. 3rd Int. Conf. Ubiquitous Computing, 2001, pp. 18–34.
[18] M. Youssef, A. Agrawala, The Horus location determination system, Wireless Netw. 14 (3) (2008) 357–374.
[19] A. Kushki, K.N. Plataniotis, A.N. Venetsanopoulos, Kernel-based positioning in wireless local area networks, IEEE Trans. Mob. Comput. 6 (6) (2007) 689–705.
[20] A. Kushki, K.N. Plataniotis, A.N. Venetsanopoulos, Intelligent dynamic radio tracking in indoor wireless local area networks, IEEE Trans. Mob. Comput. 9 (3) (2009) 405–419.
[21] O. Woodman, R. Harle, RF-based initialisation for inertial pedestrian tracking, in: Proc. 7th Int. Conf. Pervasive Comput. Commun., 2009, pp. 238–255.
[22] K. Chintalapudi, A.P. Iyer, V.N. Padmanabhan, Indoor localization without the pain, in: Proc. 16th Annu. Int. Conf. Mobile Comput. Netw., Chicago, IL, 2010, pp. 173–184.
[23] A. Haeberlen, E. Flannery, A.M. Ladd, A. Rudys, D.S. Wallach, L.E. Kavraki, Practical robust localization over large-scale 802.11 wireless networks, in: Proc. 10th Annu. Int. Conf. Mobile Comput. Netw., 2004, pp. 70–84.
[24] M.B. Kjærgaard, Automatic mitigation of sensor variations for signal strength based location systems, Location and Context Awareness, 2006, pp. 30–47.
[25] M.B. Kjærgaard, C.V. Munk, Hyperbolic location fingerprinting: a calibration-free solution for handling differences in signal strength, in: Proc. 6th Int. Conf. Pervasive Comput. Commun., 2008, pp. 110–116.
[26] A.W. Tsui, Y.H. Chuang, H.H. Chu, Unsupervised learning for solving RSS hardware variance problem in WiFi localization, Mob. Netw. Appl. 14 (5) (2009) 677–691.
[27] About Foursquare <https://foursquare.com/about>.
[28] H. Wang, S. Sen, A. Elgohary, M. Farid, M. Youssef, R.R. Choudhury, No need to war-drive: unsupervised indoor localization, in: Proc. 10th Int. Conf. Mobile Syst., Appl., Serv., 2012, pp. 197–210.
[29] C. Wu, Z. Yang, Y. Liu, W. Xi, Will: wireless indoor localization without site survey, in: Proc. IEEE Infocom, 2012, pp. 64–72.
[30] Z. Jiang, J. Zhao, J. Han, S. Tang, J. Zhao, W. Xi, Wi-Fi fingerprint based indoor localization without indoor space measurement, in: Proc. IEEE 10th Int. Conf. Mobile Ad-Hoc and Sensor Syst. (MASS), 2013, pp. 384–392.
[31] J. Koo, H. Cha, Unsupervised locating of WiFi access points using smartphones, IEEE Trans. Syst., Man., Cybern., Part C: Appl. Rev. 42 (6) (2012) 1341–1353.
[32] T. Akiyama, H. Ohashi, A. Sato, G. Nakahara, K. Yamasaki, Pedestrian dead reckoning using adaptive particle filter to human moving mode, in: Proc. Int. Conf. Indoor Positioning and Indoor Navigation (IPIN), 2013, pp. 1–7.
[33] Y. Kim, Y. Chon, M. Ji, S. Park, H. Cha, Scalable and consistent radio map management scheme for participatory sensing-based Wi-Fi fingerprinting, in: 3rd Int. Workshop on Mobile Sensing (WMS 2013), Philadelphia, USA, 2013.