Smartphone-based Wi-Fi Pedestrian-Tracking System Tolerating the RSS Variance Problem

Yungeun Kim, Hyojeong Shin, and Hojung Cha
Department of Computer Science
Yonsei University
Seoul, Korea
{ygkim, hjshin, hjcha}@cs.yonsei.ac.kr

***Abstract*— The Wi-Fi fingerprinting (WF) technique normally suffers from the RSS (Received Signal Strength) variance problem caused by environmental changes that are inherent in both the training and localization phases. Several calibration algorithms have been proposed but they only focus on the hardware variance problem. Moreover, smartphones were not evaluated and these are now widely used in WF systems. In this paper, we analyze various aspect of the RSS variance problem when using smartphones for WF: device type, device placement, user direction, and environmental changes over time. To overcome the RSS variance problem, we also propose a smartphone-based, indoor pedestrian-tracking system. The scheme uses the location where the maximum RSS is observed, which is preserved even though RSS varies significantly. We experimentally validate that the proposed system is robust to the RSS variance problem.**

Keywords-Wi-Fi fingerprinting; RSS variance problem; smartphone

I. INTRODUCTION

Localization techniques are essential for location-based services in pervasive computing [1, 2]. In recent years, the Wi-Fi fingerprinting (WF) technique [3] has been actively studied, since Wi-Fi is chiefly available in indoor environments where GPS signals cannot penetrate. Hence, the technique is becoming increasingly important as a key component for mobile applications.

The WF-based localization generally requires two phases of operation: an off-line training phase and an on-line localization phase. In the training phase, vectors of Received Signal Strength (RSS) samples from the Access Points (APs) are collected at all grid points that cover the area of interest. The collected RSS vectors are stored with the location coordinates in the fingerprint server. During the localization phase, an RSS vector denoting the user's current position is relayed to the server in real-time. The server then finds the most similar RSS vector from the database that was constructed during the training phase. Finally, the current position is estimated from the coordinates of the located RSS vector.

Since the process is divided into two phases, WF-based localization suffers from the RSS variance problem: that is, accuracy is degraded when the RSS vectors observed in the localization phase are different from the ones collected during the training phase. The RSS variance problem is caused by differences in device type, user direction, and environmental changes between the two phases. The problem becomes serious in a pervasive environment, where users may have diverse types of smartphones and carry their smartphones in different places such as in a pocket or bag, or in their hand, and so on.

Several solutions have been proposed for the RSS variance problem. To reduce the error caused by user direction, RADAR [3] collected four RSS vectors with different directions at each grid point during the training phase. Haeberlen et al. [4] proposed a manual-calibration model to solve the device-diversity problem. However, the system cannot solve the scalability issue, since new devices are continuously appearing on the market. Kjærgaard et al. [5] proposed a Hyperbolic Location Fingerprinting (HLF) method, which tried to overcome the device-diversity issue without manual calibration. In this system, fingerprints are recorded as the RSS ratio between pairs of APs rather than as absolute RSS values. Tsui et al. [6] proposed an unsupervised learning system that automatically learns the linear-transformation function between different devices. However, these solutions have limitations, in that they only focus on user direction or device type, whereas many other factors exist in real environments. Moreover, the algorithms were evaluated with old-fashioned laptops or PDAs, yet WF systems are widely used with smartphones nowadays.

In order to incorporate both smartphones and additional environmental factors into our system, we first conduct experiments on the magnitude of RSS variances caused by differences in device type, device placement, user direction, and environmental changes over time. We also analyze performance degradation due to the RSS variance problem in two types of location-based services: Point Of Interest (POI) detection, and indoor pedestrian tracking. We also propose a smartphone-based pedestrian-tracking system that tolerates the RSS variance problem without an additional calibration phase. The proposed system exploits the locations where the maximum RSS of a particular AP is observed. Wi-Fi localization is performed when a signal peak is detected and the location is determined by the location of the maximum RSS recorded during the training phase. The location of maximum RSS is preserved regardless of the RSS variance problem, since RSS mapping from the training phase to localization phase is modeled as a linear function [4]. This approach has the disadvantage of the location being estimated only when a signal peak is detected. Hence, we combine the method with Pedestrian Dead Reckoning (PDR) [7, 8], which provides relative location information by using THE accelerometer and digital compass that are built into smartphones. Our system is implemented in off-the-shelf smartphones. The experiments in real environments demonstrate that the proposed system improves positional accuracy by overcoming the RSS variance problem.

The main contributions of this paper are as follows:

- We investigate the RSS variance problem by considering differences between the training phase and the localization phase. We not only consider the device-diversity issue but also other factors that are involved when using smartphones.
- We propose an indoor pedestrian-tracking system that overcomes the RSS variance problem by using the location of the signal peak.
- The proposed system is implemented in off-the-shelf smartphones and evaluated through real-world experiments that take place in an office building and a shopping mall.

The remainder of this paper is structured as follows. Section II presents the experimental analysis of the RSS variance problem. Section III describes the proposed indoor pedestrian-tracking system and is followed by the evaluation of the real-world experiments in Section IV. Section V presents related work, and we conclude the paper in Section VI.

II. Analysis of the RSS Variance Problem

In the WF system, location is estimated by finding the RSS vector most similar to a currently observed RSS vector in the fingerprint server. However, the system cannot correctly estimate the location due to the RSS variance problem: the characteristics of radio signals in the localization phase are different from those in the training phase. In this section, we first identify the magnitude of the RSS differences between the two phases. Then, the performance decreases in POI detection and indoor pedestrian tracking are analyzed, respectively.

A. *RSS Variance Problem in Smartphone-based Wi-Fi Fingerprinting Systems*

In the smartphone-based WF system, many differences exist between the training phase and the localization phase. We classify them into four categories: device type, device placement, user direction, and environmental changes over time. Differences in device type mean that the device in the localization phase is different from the one in the training phase. This is a common case, since people use diverse kinds of smartphones that have different features. Differences in device placement mean that the devices are carried in positions that are different from the ones used in the training phase. People carry their smartphones in various places, such as in a bag or a trouser pocket [9], where the signal is attenuated differently. Since many location-based services run in the background, we cannot assume that the carrying position is in stable. Differences in user direction are also important because the RSS is significantly influenced by the human body. Time difference yields many environmental changes: the humidity and the temperature can change, doors can be open or closed, and the number of people in the building can increase or decrease.

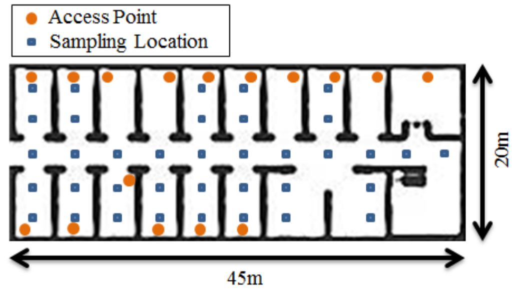


Figure 1. Test environment for the preliminary experiments

We conducted a series of experiments to observe the RSS variance problem caused by these four differences. The test environment was an office building, which was 45 m by 20 m and had sixteen APs marked with circles, as shown in Fig. 1. We collected the RSS samples with nine different configurations at the locations marked with blue rectangles. Table I shows the detailed setup of each configuration. Data 0 and 1 were sampled facing east with an HTC Nexus One in the messaging position at daytime during the first week of the month. Data 0 is used for the reference data. Data 2–4 used the Samsung Galaxy S, Motorola CLIQ, and Sony Xperia Arc, which have different characteristics to the Nexus One, as shown in Table II. Data 5–7 were collected facing west, north, and south. In data sets 8 and 9, the device placement was changed to pocket and bag, respectively, from the messaging position. Data 10 and 11 were collected to check short-term and long-term variation of the RSS. Data 10 was collected at night during the first week and data 11 was collected at daytime during the second week. Data 12 was constructed by incorporating all the possible changes over the four categories. Each data set contains 740 RSS vectors from 37 different locations with 4 m spacing and 20 RSS vectors were sampled at each location based on the time–space sampling analysis of the radio signal [10]. We then calculated the differences in the radio signals between the reference data and the other data. We were not able to enter several rooms due to security issues.

Fig. 2 shows the RSS variance denoted by box plots. The changed configurations (data 2–12) cause a larger RSS difference than the unchanged configuration. The average difference ranges from 3 to 11 dB and the maximum difference reaches about 18 dB according to the configurations. The changes in device cause RSS variances from 1 to 10 dB. Arc showed the largest RSS difference from Nexus One. The RSS differences caused by direction change were from 1 to 15 dB. Facing in the opposite direction to the reference data showed a slightly larger RSS difference than for other directions. Pocket placement showed small RSS differences of less than 10 dB, whereas the difference is bigger than 10 dB in the case of the bag. Changing time from daytime to night (Time-S) caused significant RSS differences of up to 16 dB and long-term

TABLE I. DETAILED SETUP FOR RSS SAMPLING DATA

Data	Name	Configurations			
		Device	*Placement*	*Dir.*	*Time*
0	Reference	**Nexus One**	**Messaging**	**East**	**Daytime 1^{st} week**
1	No diff.	Nexus One	Messaging	East	Daytime 1^{st} week
2	Device-S	**Galaxy S**	Messaging	East	Daytime 1^{st} week
3	Device-C	**CLIQ**	Messaging	East	Daytime 1^{st} week
4	Device-A	**Xperia Arc**	Messaging	East	Daytime 1^{st} week
5	Dir-W	Nexus One	Messaging	**West**	Daytime 1^{st} week
6	Dir-N	Nexus One	Messaging	**North**	Daytime 1^{st} week
7	Dir-S	Nexus One	Messaging	**South**	Daytime 1^{st} week
8	Placement-P	Nexus One	**Pocket**	East	Daytime 1^{st} week
9	Placement-B	Nexus One	**Bag**	East	Daytime 1^{st} week
10	Time-S	Nexus One	Messaging	East	**Night** 1^{st} week
11	Time-L	Nexus One	Messaging	East	Daytime **2^{nd} week**
12	Mixed	**Xperia Arc**	**Bag**	**West**	**Night 2^{nd} week**

TABLE II. CHARACTERISTICS OF DIFFERENT SMARTPHONES

Device	Characteristics		
	OS	*Wi-Fi chipset*	*Antenna*
HTC Nexus One	Android 2.3	BCM 4329	At the top
Samsung Galaxy S	Android 2.3	SWB-B23	At the bottom
Motorla CLIQ	Android 1.5	BCM 4329	At the bottom
Sony Xperia Arc	Android 2.3	BCM 4329	At the top

variation (Time-L) also showed relatively large RSS differences from 1 to 11 dB. Mixed data that incorporated the changes over the four categories showed the largest RSS difference of up to 18 dB. The result clarifies that we should not only consider the device–diversity problem, but also the differences in placement, user direction, and environmental changes over time in smartphone-based WF systems.

B. *Performance Degradation due to the RSS Variance Problem*

We showed that the RSS difference between the training phase and localization phase is up to 18 dB, and this is large enough to cause significant performance degradation in the WF system. In this section, we analyze the effect of the RSS variance in two representative types of location-based services: the POI detection service [8] and indoor pedestrian–tracking service [11, 12]. The POI detection services should decide if a user is at a particular POI. Since most POIs are set in different rooms, these types of services generally require room-level accuracy. On the other hand, indoor pedestrian–tracking systems should provide high positional accuracy while a user is moving.

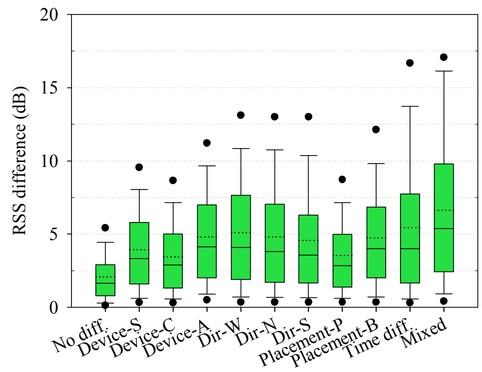


Figure 2. RSS differences with the refrence data

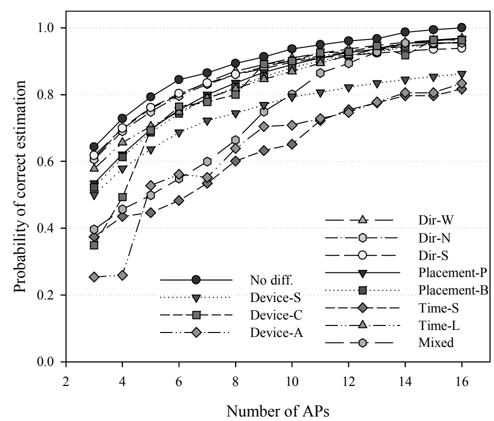


Figure 3. Performance of POI detection

1) Localization: For the localization experiment, we reused the data collected in the previous section; the reference data is used as a radio map, and we performed location estimation with the K-Nearest Neighbors (KNN) algorithm for 740 RSS vectors in each test data set. We used the Tanimoto coefficient [13], which is the similarity function that is known to be efficient for comparing Wi-Fi fingerprints. Since the number of APs influences the performance of the WF system, we conducted the experiment by changing the number of APs used in localization. Additionally, we also differentiated the number of samples used for location estimation: five samples for POI detection and two samples for indoor pedestrian tracking. This is reasonable because only two or three Wi-Fi scans are available in each location while the user is moving.

2) Effect on POI Detection Services: To see the performance decrease in POI detection, we conducted the location estimation inside rooms. We then calculated the probability of the estimated location and real location being in the same room. Fig. 3 shows this probability according to the different situations. The performance improved with an increased number of APs. The probability is higher than 0.6 with eight APs and 0.8 with twelve APs. The lowest

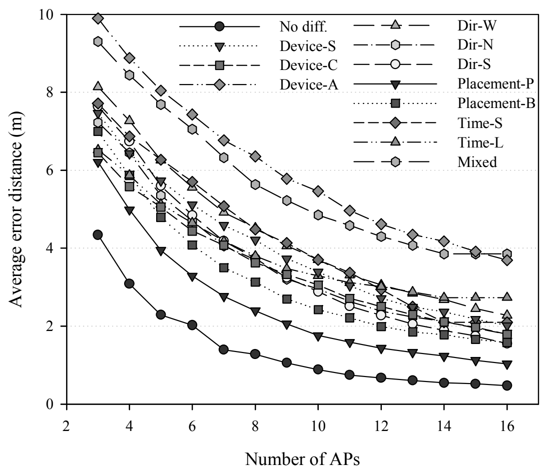


Figure 4. Performance of indoor pedestrian tracking

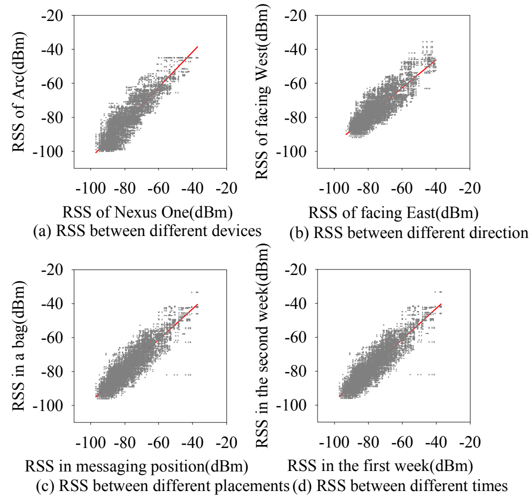


Figure 5. RSS comparisions of four different configurations

TABLE III. RATIO OF LONG–TAIL ERRORS

	Ratio of (error > 4 m)	*Ratio of (error > 8 m)*
No diff.	11 %	5 %
Others	33 %	10 %

probability of 0.81 with sixteen APs was shown when changes were made over the four categories.

3) Effect on Indoor Pedestrian–tracking Services: Since people usually move along the corridor, we extracted the localization results outside rooms in the opposite manner to the experiment for POI detection. Fig 4 shows the mean–distance errors and standard deviation between the estimated location and real location according to the different configurations. All the cases of changing configurations showed worse results than the case of *No diff.*. The performance was improved as the number of APs increased, and the worst results were shown in Time-S and Mixed data, whereas other cases showed relatively small mean–distance errors. In the case where eight APs were used, the mean–distance errors ranged from 2.4–6.4 m depending on the configuration changes. The mean–distance error was reduced to 1.1–3.9 m by using sixteen APs. As shown in Table II, however, the RSS variance problem increased the percentage of *long-tail errors*—the quite unlikely but still existing occurrence of a very large error—which threatens the reliability of the indoor pedestrian-tracking system. The ratio of errors larger than 8 m is increased from 5 % to 11 % when the configuration is changed.

4) Conclusion for Performance Decrease due to the RSS Variance Problem: Based on the experimental results, POI detection is shown to be more robust to the RSS variance problem. With sixteen APs, a probability of correct estimation is always higher than 0.8 regardless of configuration changes. The indoor pedestrian-tracking, however, suffered significant performance degradation due to the RSS variance problem. The average error distance reaches up to 3.9 m, even though sixteen APs were used. Moreover, the percentage of the long-tail error, which threatens the reliability of the system, was increased due to the RSS variance problem.

III. INDOOR PEDESTRIAN-TRACKING SYSTEM WITH PEAK -BASED WI-FI FINGERPRINTING

A. Rationale for Using the RSS Peak

One of the solutions for the RSS variance problem is in identifying a linear transformation from the training phase to the localization phase [6, 10]. Since this method was validated only for the device-diversity issue, we compared the RSS values between the reference data and four test data sets that varied upon device, direction, placement, and time, respectively. As shown in Fig. 5, we observed a linear shift in the RSS values for all cases. However, finding the linear transformation is not practical since it requires an additional learning phase or manual calibration phase. Alternatively, we propose an approach that exploits the location of the maximum RSS, which is preserved in a linear relationship.

Based on the linear relationship between the training and localization phases, we assume that the training data can be mapped to the test data by a linear-transformation function as in Equation (1).

$$y_i^j \approx \alpha \cdot x_i^j + \beta \quad (\alpha > 0) \tag{1}$$

where x_i^j and y_i^j represent the RSS from the j-th AP at location i in the training data and test data, respectively. Then the location of the maximum RSS from AP j is always preserved around the same location as shown in Equation (2).

$$\arg\max_i y_i^j \approx \arg\max_i \left(\alpha \cdot x_i^j + \beta\right) = \arg\max_i x_i^j \tag{2}$$

When a peak RSS signal from a particular AP is detected, we estimate the current position accurately by finding the peak location of the AP in the training data. Since an RSS peak is detected intermittently, we combine the peak-based location estimation with PDR, which uses the accelerometer and digital compass that are built into smartphones [8]. With this

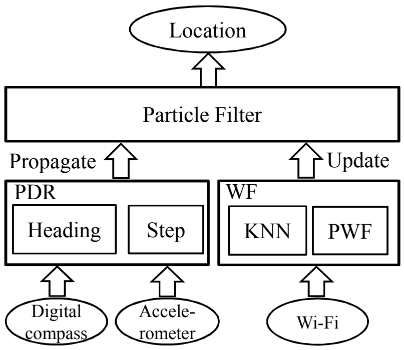


Figure 6. Structure of the proposed tracking system

approach, the RSS variance problem is overcome without resorting to a manual-calibration phase.

B. System Overview

Fig. 6 illustrates the structure of the proposed tracking system. We combined WF with PDR by using a particle filter. The PDR used in the system detects steps and the heading orientation by reading the accelerometer and digital compass. A detailed description of this can be found in [8]. Since PDR provides relative location information, we update the absolute location with WF. WF in the proposed system consists of two components: the KNN-based (KNN) method and the Peak-based Wi-Fi Fingerprinting (PWF) method. KNN is always available but the accuracy is not reliable when the RSS variance problem occurs. PWF provides highly accurate location estimation but is available only when the peak is detected. The proposed system prioritizes the PWF. KNN is performed in two cases: when initialization is required or when a mobile user is in the environment where PWF is unavailable.

C. Peak-based Wi-Fi Fingerprinting

We propose PWF that uses the location of the maximum RSS from a particular AP. The PWF technique has two phases: the training phase and localization phase.

1) Training phase: A site-survey is conducted to builid a radio map in a similar way to the traditional fingerprinting system. Then the system preprocesses the radio map for the PWF as follows:

- The radio map is divided into a number of sections based on the property of the place (e.g. straight corridor).and S_{normal} represents the section where only the KNN-based method is available (e.g. rooms).
- For all the section of S_{peak}, the location index and the RSS value of the maximum RSS of each AP is recorded.

2) Localization phase: Since we do not know whether the currently observed RSS is the maximum RSS or not in real time, we alternatively detect the peak in the series of observed RSS values as follows.

First, the current section s is estimated with the KNN algorithm using the Tanimoto coefficient. The peak-based

Initialization:
A particle is x_t^i, for i=1, …, N and t = 0.
Distribute N particles at the initial position obtained by KNN.
Loop:

1. *Local movement estimation*:
$$x_t^i = f(x_{t-1}^i, u_t)$$
Each particle move is based on the step and heading information u_t.
2. *WF-based location estimation (if possible):*
$$\tilde{w}_t^i = g(x_t^i, L_{wf})$$
Each particle updates its weight based on the current location of the particle and the estimation location L_p.
3. *Filtering*:
$$\tilde{w}_t^i = 0 \; if \; x_t^i \; emulated \; an \; impossible \; movement$$
Then, normalize the weights and resample particles according to the weights.

Figure 7. Pseudo algorithm for the particle filter.

method is suspended when section s is classified as S_{normal}. Second, in order to check if the peak was detected among the n recently observed RSS vectors, the set of RSS for each i-th AP is defined as $R^i = \{rss_1^i, \; rss_2^i, \cdots, rss_{n-1}^i, rss_n^i\}$. For all APs, the peak detection function $Pd(R^i)$ is performed as follows:

$$Pd(R^i) = \begin{cases} true, \begin{pmatrix} \text{if } \; rss_j < rss_{j+1} \text{ for all } j < \frac{n}{2} \\ \text{and if } rss_j > rss_{j+1} \text{ for all } j \geq \frac{n}{2} \end{pmatrix} \\ false \; , else \end{cases} \quad (3)$$

The peak-based method is suspended when the peak detection function returns a false for all APs.

Third, the location is estimated from the location of the maximum RSS from the i th AP recorded in the radio map. The set of fingerprints of section s is denoted as $F_s = \{f_1, f_2, \cdots, f_{m-1}, f_m, k^1, k^2, \cdots, k^{p-1}, k^p\}$ where f represents the location fingerprint, m is the number of fingerprints in the section, k represents the location index of the maximum RSS of each AP, and p is the number of APs in the section. Then the current user location is estimated as the location of $f_{k^i+n/2}$ or $f_{k^i-n/2}$ depending on the direction of movement.

D. Particle Filter with Accelerometer and Digital Compass

Since the KNN and PWF have problems of inaccuracy and intermittence, respectively, we used a particle filter to combine the WF with the PDR. The particle filter estimates the user's movement from sensor readings. The particle filter also simulates uncertainty from the measurement noise by generating a number of particles. Each particle presents a single possible trajectory estimation. Fig. 7 describes the algorithm for the proposed location estimation. Having read sensor u_t, the system updates the current location of each particle x_t. from the PDR. When a WF returns the location L_{wf}, the system evaluates the weight value w_t for each particle by using L_{wf}. Particles close to the location L_{wf} have high weight values and the others do not. Additionally, the weight of particles that move through a wall is set to 0. Based on the weight values of the particles, the system filters out bad

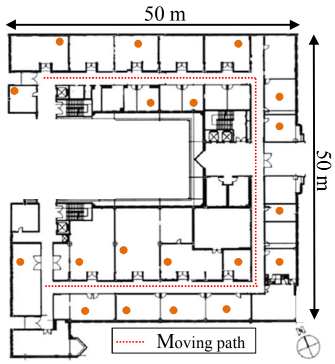


Figure 8. Test environment for the evaluation

TABLE IV. DETAILED SETUP FOR WAR-WALKING

Data	Name	Configurations			
		Device	***Placement***	***Dir.***	***Time***
0	Reference	**Nexus One**	**Messaging**	**Clockwise**	**Daytime 1st week**
1	No diff.	Nexus One	Messaging	Clockwise	Daytime 1st week
2	Device-A	**Xperia Arc**	Messaging	Clockwise	Daytime 1st week
3	Dir-Reverse	Nexus One	Messaging	**Counter-Clockwise**	Daytime 1st week
4	Placement-B	Nexus One	**Bag**	Clockwise	Daytime 1st week
5	Time-S	Nexus One	Messaging	Clockwise	**Night 1st week**
6	Mixed	**Xperia Arc**	**Bag**	**Counter-Clockwise**	**Night 2nd week**

estimations and resamples particles to maintain the total number of particles N.

IV. EVALUATION

A. Experimental Setup

The test environment was an office building, which was 50 m by 50 m and had nineteen APs marked with circles, as shown in Fig. 8. We conducted war-walking; that is, collecting RSS samples by walking along the path marked by a dotted line, with five different configurations that showed significant RSS variances in the preliminary experiments. Each war-walking data set includes radio-signal vectors, ground-truth location information, sensing values from the accelerometer and digital compass, and time information. Table IV shows the detailed setup of each configuration. Data 0 and 1 were obtained by walking clockwise with the HTC Nexus One in its messaging position at daytime during the first week. Data 0 is used for the reference data. Data 2 used the Xperia Arc, which showed the largest RSS difference compared to Nexus One. Data 3 was collected by walking counter-clockwise and data 4 was obtained by carrying the device in the bag. Data 5 was collected at night in the first week and data 6 was the change that occurred over the four categories.

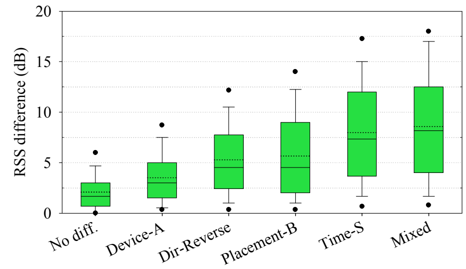


Figure 9. RSS differences with the refrence data (data 0)

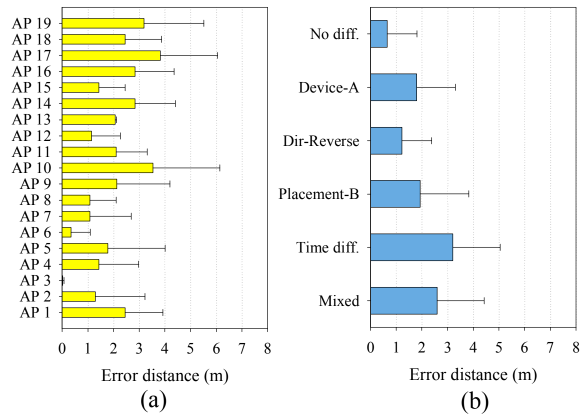


Figure 10. The peak location error according to: (a) the AP, and (b) war-walking configurations.

B. RSS Difference and Preservation of Peak Location

Fig. 9 shows the RSS differences between data 0 and other data. Data 2–6 cause a larger RSS difference than that of the same configuration. The average difference ranges from 3 to 8 dB and the maximum difference reaches about 18 dB depending on the configurations. The device difference (Device-A) showed the smallest RSS variances among the changed configurations (data 2–6). Time-S and Mixed data showed the largest RSS variances up to 17 dB.

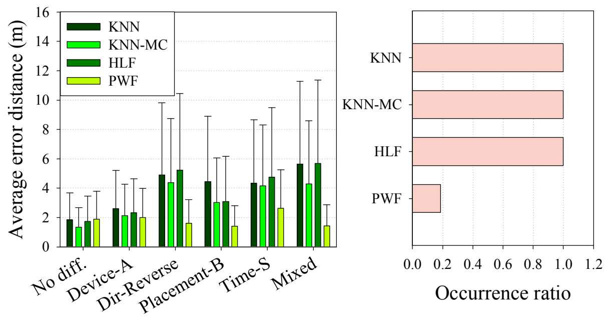
We also compared the location of the RSS peak for each AP. Fig. 10(a) shows the average error distance of each AP's peak location between the reference data and the other data. The average error distance ranges from 0.1 m to 3.18 m depending on the AP. Fig 10(b) shows the error distance of the peak location according to different configurations. Most of peak locations were found within 6 m and the average error distance was 1.2–3.1 m. This result validates that the peak location can be used to overcome the RSS variance problem.

C. Performance of Peak-based Wi-Fi Fingerprinting

We first analyzed the effect of the window size *n* on the proposed PWF method. Table V shows (1) the ratio of true-positive estimation where the peak is detected at the location of the maximum RSS, and (2) the ratio of false-positive estimation where the peak is detected when the RSS is not a maximum value. As window size increases, the true-positive ratio is reduced gradually while the number of false-positive

TABLE V. THE EFFECT OF THE WINDOW SIZE ON PWF

Window size	*True-positive ratio*	*False-positive ratio*
3	0.87	0.31
5	0.73	0.07
7	0.62	0.02
9	0.54	0.01
11	0.44	0.01



(a) (b)

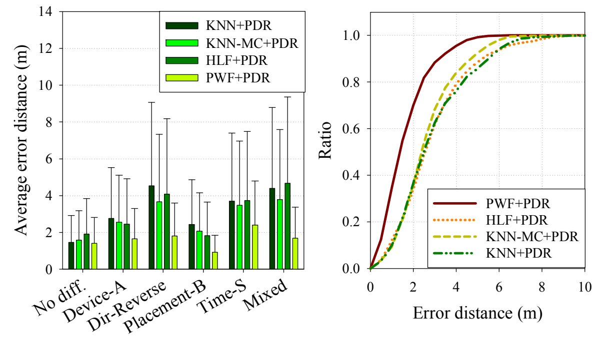
Figure 11. Performance of Wi-Fi fingerprinting algorithms: (a) is the accuracy according to the configurations, and (b) is the occurrence ratio of four different algorithms.

estimations drops rapidly at $n = 5$. We therefore use $n = 5$ in subsequent experiments.

To evaluate the performance of the PWF, we compared PWF with three different algorithms: KNN, KNN after Manual Calibration with linear transformation (KNN-MC) [10] and HLF [5]. We used data 0 as a radio map, and data 1–6 as test data. Fig. 12 compares the average error distance of the different algorithms. From Fig. 11(a) we find that from among the four algorithms, PWF showed the best result regardless of configuration changes. The average error distance was always lower than 2.0 m except for 3.0 m in case of Time-S. We also found that KNN-MC improved the performance of KNN slightly. This means that the linear transformation minimized the sum of the squares of the RSS differences but the gain for individual location fingerprints is insignificant. Since PWF performs localization only when the signal peak is detected, we also calculated the occurrence ratio of each algorithm. As shown in Fig. 11(b), the occurrence ratio of PWF is about 0.2, whereas for other algorithms the ratios are 1.0. To overcome the low occurrence ratio of PWF, the proposed pedestrian-tracking system integrated PWF and PDR with a particle filter.

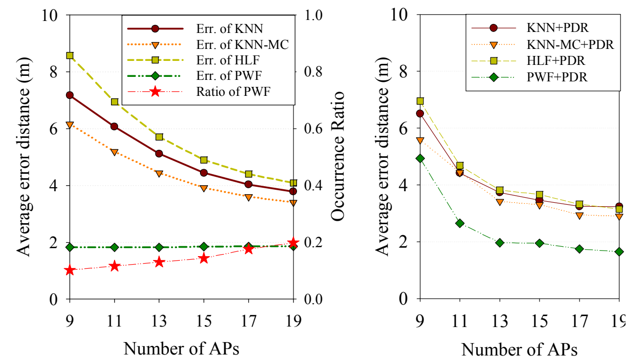
D. Performance of the Proposed Pedestrian-Tracking System

To evaluate the performance of the proposed pedestrian-tracking system (PWF+PDR), we compared the system with three different systems: KNN with PDR (KNN+PDR), KNN-MC with PDR (KNN-MC+PDR), and HLF with PDR (HLF + PDR). PDR provided relative location information based on the sensing values recorded during war-walking. Fig 12(a) shows the performance of different combinations. The average error distance of PWF+PDR is always lower than 2.4 m regardless of configurations. Fig. 12(b) shows the Cumulative Distribution Function (CDF) of the error distances according to the tracking algorithms. PWF+PDR showed good results compared with other algorithms. All the error distances were lower than 6.5 m and the ratio of error distances lower than 3.0 m was 0.88. This means that PWF reduced the percentage of long-tail errors caused by RSS variances, and that the PDR provided the location estimation successfully when PWF was unavailable.



(a) (b)

Figure 12. Accuracy of different pedestrian-tracking systems: (a) mean-distance error according to the configuration changes and (b) CDF of the error distance depending on the systems

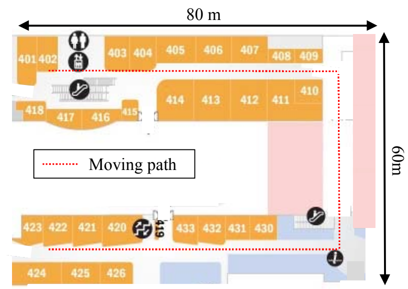


(a) (b)

Figure 13. Performance vs. the number of APs: (a) average error distance of each Wi-Fi localization algorithm and occurrence ratio of PWF and (b) average error distance for pedestrian tracking

E. Performance vs. the Number of APs

The number of APs affects the performance of the WF system. Hence, we analyze the influence of the number of APs on the four different WF systems. Fig. 13(a) compares the error distance according to the number of APs. PWF maintains a high accuracy of 2 m, whereas other algorithms show an average error distance of 4–8 m as the number of APs changes from 19 to 9. Fig. 13(a) also shows the occurrence ratio for PWF. When using 19 APs, PWF performed location estimation with only 20 % of the RSS samples. Moreover, the occurrence ratio drops to 10 % when using 9 APs. In other words, the number of APs does not affect the accuracy of PWF but the occurrence ratio of PWF. Fig. 13(b) shows the average error distances when WF is combined with PDR. Despite the low occurrence ratio, PEAK+PDR shows the best result regardless of the number of APs. The average distance error of PEAK+PDR with 19 APs is 1.6 m, which is improved by 42 % from KNN-MC.



	Office	Mall
Corridor width	2 m	5 m
Material of wall	Concrete	Window
Number of APs	32	18
Number of people on the path	Less than 10	Dozens of people

Figure 14. The test environment in a shopping mall. The difference from the office building is also depicted

F. Case Study: Performance in the Shopping Mall

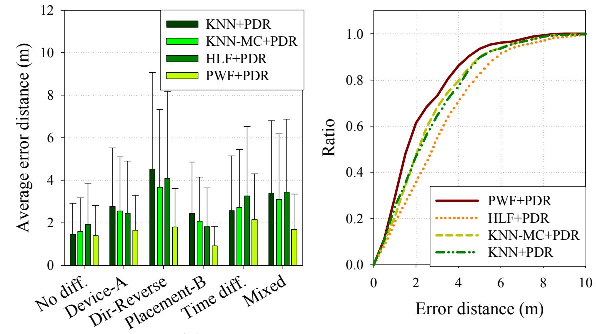
We evaluated the proposed system in a shopping mall, which has different characteristics from the office building, as shown in Fig. 14. Fig. 15(a) compares the accuracy of four different WF systems and the ratios for PWF as well. PWF shows the best result but the error distance is increased and the ratio is decreased compared to the result in the office building. RSS varied significantly due to the crowd in the shopping mall. Consequently, the location of maximum RSS was changed or the system failed to detect the RSS peak more frequently in the localization phase. Fig. 15(b) shows the CDF of error distances depending on the tracking systems. PWF+PDR still showed the best result compared to the other algorithms but the difference from other systems is reduced. This means that the low occurrence ratio in the shopping mall degrades the performance of the proposed tracking system.

V. Related Works

In recent years, active research has been conducted on WF systems. RADAR [3] was the first WF system, which is divided into a training phase and a localization phase. In the training phase, a radio map is constructed by measuring RSSs from existing APs at all locations. In the localization phase, the location is estimated by the K-Nearest Neighbor algorithm that finds the RSS vector that has the closest Euclidian distance to the currently observed RSS vector. Numerous WF systems have been proposed that work to eliminate the disadvantages of RADAR: improving the localization algorithm, reducing the cost of the training phase, and handling the RSS variance problem. We provide an overview of the key research for each improvement.

A. Schemes That Improve the Localization Algorithm

Brunato [14] introduced a location-discovery technique based on a support vector machine. In this system, a regression engine tracks the mobile user and a classification engine determines the room where the user is currently located. Nibble [15] is a probabilistic location service using Bayesian networks to infer the location of a device. Horus [16] reduced computational overhead in the localization



(a) (b)

Figure 15. Performance in the shopping mall: (a) average error distance and the ratio of Wi-Fi-based localization and (b) CDF of the error distances depending on the tracking systems

algorithm by clustering locations sharing a common set of APs. A kernel-based, Wi-Fi positioning system [17] was also proposed to improve location accuracy. In this system, a kernelized distance-calculation algorithm for comparing RSS observations to RSS training records is presented. Kushki et al. [18] proposed a state-space Bayesian filter: a nonparametric information filter to track mobile users in situations where a Kalman filter is not applicable. In this system, a discrimination score filter, i.e., a novel AP selection algorithm was also introduced to improve location accuracy. This filter chooses APs for which RSS values do not vary much over time but vary considerably over space.

B. Schemes That Reduce the Cost of the Training Phase

Oliver et al. [19] proposed a quick and automatic site-survey scheme by tracking the current location with a wearable location system. The organic location system [20] handled off-line training by incorporating a collaborative approach where users collect fingerprints while using their devices. The system mandates active participation by users to collect the RSS vectors of the new locations. The EZ localization system [21] was proposed to enable localization without a training phase. In this system, many RSS vectors collected at unknown locations are reported to the server. The server then runs localization by using constraints in the physics of the wireless propagation. The EZ system, however, assumed that, occasionally, RSS vectors at known locations should be reported. Kim et al. proposed smartphone-based, autonomous war-walking by using an accelerometer and a digital compass [22].

C. Schemes That Handle the RSS Variance Problem

Haeberlen et al. [4] proposed a manual calibration. An RSS map is constructed for each device during the training phase; then a corresponding RSS map is used in the localization phase. This approach is accurate but impractical, since new types of devices are continuously produced. Kjærgaard et al. [23] proposed an automated method that calibrates the RSS variance in different devices during the on-line learning phase. This method solved the scalability issue, but required an additional learning phase and was less accurate than the manual-calibration method. Kjærgaard et al. also proposed a ratio-based localization system [5], which uses the ratio of RSS values between different APs rather than absolute RSS values, while Dong et al. [24] generated a radio map using RSS differences between pairs of APs instead of using absolute RSS values. Tsui et al. [6] proposed an unsupervised learning system that automatically learns the linear-transformation function between two different Wi-Fi devices. The method roughly estimates the current location with the Pearson product-moment correlation coefficient, and then an expectation maximization learning algorithm is applied to train the transformation function. The EZ system [21] performs localization by considering different receiver gains according to the device type. Park et al. [25] reduced the RSS difference between devices by wide smoothing of the signal-strength distribution function.

VI. Conclusion

In this paper, we investigated the RSS variance problem in a smartphone-based WF system. We evaluated the effect of RSS variances between the training phase and the localization phase. We showed that the smartphone-based WF system should consider the device-diversity issue and other issues such as user direction, device placement, and environmental changes over time. The RSS variance problem causes significant performance degradation in indoor pedestrian-tracking systems. The proposed tracking system overcomes the RSS variance problem by using the location of signal peak for the localization. This approach is accurate but has the problem of a low occurrence ratio. We compensated for the defect by integrating the system with PDR, which provides relative location information by using an accelerometer and a digital compass. The experimental results demonstrated that the proposed system is robust to the RSS variance problem in real-world environments.

Acknowledgement

This work was supported by the National Research Foundation of Korea grant funded by the Korean government, Ministry of Education, Science and Technology (No.2011-0000156, No.2011-0015332).

References

[1] J. Hightower and G. Boriello. "Location Systems for Ubiquitous Computing," *IEEE Computer*, vol.34, no. 8, pp. 57–66, 2001.

[2] M. Satyanarayanan, "Pervasive Computing: Vision and Challenges," *IEEE Personal Communications,* vol. 8, pp. 10–17, 2001.

[3] P. Bahl and V. N. Padmanabhan, "RADAR: An In-building RF-based User Location and Tracking System," *In Proc. 19th IEEE Int. Comput. Commun. (InfoCom)*, vol. 2., pp. 775–784, 2000.

[4] A. Haeberlen, E. Flannery, A.M. Ladd, A. Rudys, D.S. Wallach, and L.E. Kavraki, "Practical Robust Localization over Large-scale 802.11 Wireless Networks," *In Proc. of the 10th Annual Int. Conf. Mobile Computing and Networking*, pp. 70–84, 2004.

[5] M.B. Kjærgaard, and C.V. Munk, "Hyperbolic Location Fingerprinting: A Calibration-free Solution for Handling Differences in Signal Strength," *In Proc. of the 6th Int. Conf. Pervasive Computing and Communications*, pp. 110–116, 2008.

[6] A.W. Tsui, Y.H. Chuang, and H.H. Chu, "Unsupervised Learning for Solving RSS Hardware Variance Problem in WiFi Localization," *Mobile Networks and Applications*, vol. 14, no. 5, pp.677–691, 2009.

[7] U. Steinhoff, and B. Schiele, "Dead Reckoning from the Pocket-An Experimental Study," *In Proc. of the 8th Int. Conf. Pervasive Computing and Communications,* pp. 162–170, 2010.

[8] J. Chon, and H. Cha, "LifeMap: Smartphone-based Context Provider for Location-based Services," *IEEE Pervasive Computing*, vol. 10, no.2, pp. 58–67, 2011.

[9] F. Ichikawa, J. Chipchase, and R. Grignani, "Where's the Phone? A Study of Mobile Phone Location in Public Spaces," *In Proc. of the 2nd Int. Conference on Mobile Technology, Applications and Systems,* Mannheim, Guangzhou, pp. 1–8, 2005.

[10] C. Figuera, J.L. Rojo-Álvarez, I. Mora-Jiménez, A. Guerrero-Curieses, M. Wilby, and J. Ramos-López, "Time-Space Sampling and Mobile Device Calibration for WiFi Indoor Location Systems," *IEEE Transactions on Mobile Computing,* vol. 10, no. 7, pp. 913–926, 2011.

[11] K. Frank, B. Krach, N. Catterall, and P. Robertson, "Development and Evaluation of a Combined WLAN and Inertial Indoor Pedestrian Positioning System," *In Proc. of the Int. Technical Meeting of The Satellite Division of the Institute of Navigation*, pp. 538–546, 2001.

[12] W. Oliver and H. Robert, "Pedestrian Localisation for Indoor Environments," *In Proc. of the 10th Int. Conference on Ubiquitous Computing*, pp. 114–123, 2008.

[13] P. Jaccard, "The Distribution of the Flora in the Alpine Zone," *New Phytologist*, vol. 11, no. 2, pp. 37–50, 1912.

[14] M. Brunato, and R. Battiti, "Statistical Learning Theory for Location Fingerprinting in Wireless LANs," *Computer Networks,* vol. *47*, no. 6, pp. 825–845, 2005.

[15] P. Castro, P. Chiu, T. Kremenek, and R. Muntz, "A Probabilistic Room Location Service for Wireless Networked Environments," *In Proc. of the 3rd Int. Conf. Ubiquitous Computing,* pp. 18–34, 2001.

[16] M. Youssef, and A. Agrawala, "The Horus Location Determination System," *Wireless Networks,* vol. 14, no. 3, pp. 357–374, 2008.

[17] A. Kushki, K.N. Plataniotis, and A.N. Venetsanopoulos, "Kernel-based Positioning in Wireless Local Area Networks," *IEEE Transactions on Mobile Computing,* vol. *6*, no. *6*, pp. 689–705, 2007.

[18] A. Kushki, K.N. Plataniotis, and A.N. Venetsanopoulos, "Intelligent Dynamic Radio Tracking in Indoor Wireless Local Area Networks," *IEEE Trans. Mobile Computing,* vol. 9, no. 3, pp. 405–419, 2009.

[19] O. Woodman, and R. Harle. *In Proc. of the 7th Int. Conf. Pervasive Computing and Communications*, pp. 238–255, 2009.

[20] J. Park, B. Charrow, D. Curtis, J. Battat, E. Minkov, J. Hicks, S. Teller, and J. Ledlie, "Growing an Organic Indoor Location System," *In Proc. of the 8th Int. Conference on Mobile Systems, Applications, and Services*, pp. 271–284, 2010.

[21] K. Chintalapudi, A.P. Iyer, and V.N. Padmanabhan, "Indoor Localization without the Pain," *In Proc. of the 16th Annual Int. Conference on Mobile Computing and Networking*, Chicago, pp. 173–184, 2010.

[22] Y. Kim, Y. Chon, and H. Cha, "Smartphone-Based Collaborative and Autonomous Radio Fingerprinting," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 42, no. 1, pp. 112–122., 2012.

[23] M.B. Kjærgaard, "Automatic Mitigation of Sensor Variations for Signal Strength Based Location Systems," *Location-and Context-Awareness*, pp. 30-47, 2006.

[24] F. Dong, T. Chen, J. Liu, Q. Ning, and S. Piao, "A Calibration-free Localization Solution for Handling Signal Strength Variance," *In Proc. of the 2nd ACM Int. Workshop on Mobile Entity Localization and Tracking in GPS-less Environments*, pp. 79–90, 2009.

[25] J. Park, D. Curtis, S. Teller, J. Ledlie, and others, "Implications of Device Diversity for Organic Localization," *In Proc. 30th IEEE Int. Comput. Commun. (InfoCom)*, Shanghai, China, pp. 3182–3190, 2011