

## Research Article

# A Design of Irregular Grid Map for Large-Scale Wi-Fi LAN Fingerprint Positioning Systems

Jae-Hoon Kim,<sup>1</sup> Kyoung Sik Min,<sup>2</sup> and Woon-Young Yeo<sup>3</sup>

<sup>1</sup> Department of Industrial Engineering, Ajou University, 206 Worldcup-ro, Yeongtong-gu, Suwon 443-749, Republic of Korea

<sup>2</sup> International Cooperation Group, Korea Internet & Security Agency, 109 Jungdae-ro, Songpa-gu, Seoul 138-803, Republic of Korea

<sup>3</sup> Department of Information and Communications Engineering, Sejong University, 98 Gunja-dong, Gwangjin-gu, Seoul 143-747, Republic of Korea

Correspondence should be addressed to Woon-Young Yeo; [wyyeo@sejong.ac.kr](mailto:wyyeo@sejong.ac.kr)

Received 7 May 2014; Revised 13 August 2014; Accepted 27 August 2014; Published 15 September 2014

Academic Editor: David Akopian

Copyright © 2014 Jae-Hoon Kim et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The rapid growth of mobile communication and the proliferation of smartphones have drawn significant attention to location-based services (LBSs). One of the most important factors in the vitalization of LBSs is the accurate position estimation of a mobile device. The Wi-Fi positioning system (WPS) is a new positioning method that measures received signal strength indication (RSSI) data from all Wi-Fi access points (APs) and stores them in a large database as a form of radio fingerprint map. Because of the millions of APs in urban areas, radio fingerprints are seriously contaminated and confused. Moreover, the algorithmic advances for positioning face computational limitation. Therefore, we present a novel irregular grid structure and data analytics for efficient fingerprint map management. The usefulness of the proposed methodology is presented using the actual radio fingerprint measurements taken throughout Seoul, Korea.

## 1. Introduction

A location-based service (LBS) coordinates user location with various end-user applications to improve relevance, context, and economic value. Despite the many possibilities offered by an LBS, its market penetration has been slow. Most early-stage services have failed to proliferate a mass market. Moreover, monetization of the services is limited to a few special-purpose markets, such as car map/navigation. The limitations of LBSs are related mainly to the insufficient precision of position estimation. The general mean error of position estimation is in the order of many tens of meters and the deviation can be in the order of hundreds of meters. With the rapid increase in Wi-Fi LAN access points (APs) in metropolitan areas, Wi-Fi can be used as a viable alternative positioning infrastructure [1, 2]. Each Wi-Fi AP continuously generates a radio signal with a unique identifier or media access control (MAC) address, which enables mobile devices to identify a specific AP. The millions of public and private Wi-Fi APs can be used for Wi-Fi-based positioning. Based

on the Received Signal Strength Index (RSSI) from each valid AP and embedded algorithms, the typical accuracy of Wi-Fi positioning is in the order of tens of meters in metropolitan areas, which is more accurate than other cellular positioning technologies, because Wi-Fi APs are more closely spaced than cellular network base stations. The TTFF can be as short as 100 ms. Compared with a GPS [3, 4], Wi-Fi positioning works better in urban canyons than in rural areas. It works well in dense metropolitan areas, both outdoors and indoors, owing to its greater received signal strength and lower attenuation. In many Wi-Fi positioning methods, AP triangulation and radio frequency (RF) fingerprinting provide the basic scientific significance to current positioning methodologies. Triangulation is simple to implement [1, 5, 6]. As seen in Figure 1(a), three reference APs with already known coordinates are required. After measuring the distance from the APs and a target point, three circles can be drawn. The circles intersect at a single target point. The coordinate of the target point can be easily calculated by the distance from the known coordinates of the APs.

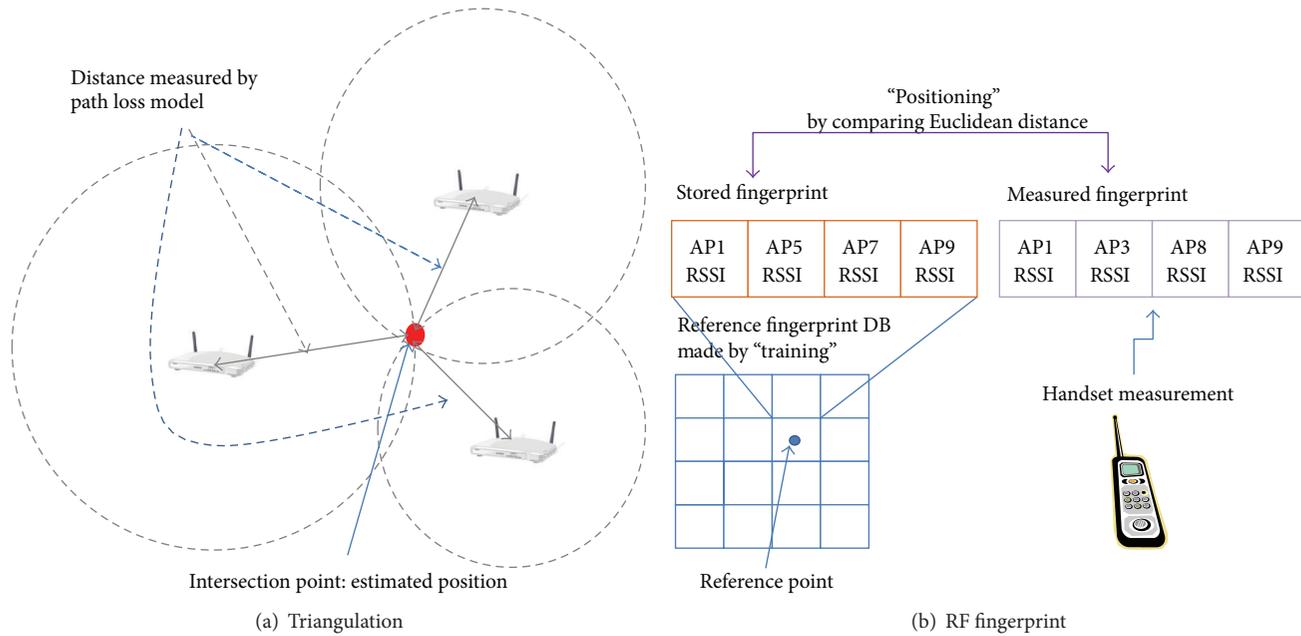


FIGURE 1: Position estimation by WPS.

The significant difficulty of this approach is the distance measurement from each AP to the target point. Typical path loss models such as COST231 [7] and Okumura-Hata [8] are generally applied to measure the distance. However, it is extremely difficult to build a good, general model for distance measurement, which coincides with the actual field situation. RF fingerprinting [9, 10] consists of two phases—training and positioning—demonstrated in Figure 1(b). In the training phase, a reference fingerprint database (DB) is constructed. The reference DB contains the signal strength measurements of the APs at all reference points. Usually, the entire area should be divided into a set of grids and the centers of grids are considered as the reference points. During the positioning phase, the position of a target point can be identified by comparing its measured fingerprint with the prestored reference fingerprints DB. The main advantage of RF fingerprinting is algorithmic simplicity. Simple comparing algorithms, such as pattern matching, can be easily applied to a practical process of position estimation. Then the RF fingerprinting is more preferred than triangulation [11].

Most of advancements for RF fingerprinting have been searched in position estimation algorithms. The most well-known pattern matching algorithm is nearest neighbor (NN) [9]. As an enhanced version of NN algorithm, K nearest neighbor (KNN) algorithm can be taken into account [9]. The average of coordinates of k-reference grids can be used to determine the estimated position of a target point. Various variations, such as smallest polygon [12] and neural networks [13], are applied in the framework of KNN pattern matching. Another type of algorithm for positioning adopts a probabilistic framework. The idea of the probabilistic framework is to compute the conditional probabilistic density function (pdf) of an estimated position of the target point. The probabilistic likelihood can be modeled by histogram [14],

Gaussian [13], Lognormal [15], or Kernel [14]. In addition, a hybrid method of pattern matching and probabilistic frame was invented [16]. Using the overlapped probabilistic existence maps of APs, they calculated the most promising position of a target point. Another significant study for RF fingerprinting is data filtering [17]. To keep the integrity of fingerprint data, data filtering schemes, such as the Kalman-filter [18] or probabilistic filtering approach with machine learning [14], are applied to Wi-Fi positioning for fingerprint data management.

In spite of such advancement of position estimation algorithms or data filtering methods, the enhancement of positioning precision currently faces computational limitation. The error bound still remains to tens of meter. The recent algorithmic advances just present small and limited improvement. Therefore, we focus on the more fundamental frame of RF fingerprint WPS: the structure of reference fingerprint DB. The conventional large-scale fingerprint DBs have a regular square map structure. The entire area of a geographical region is divided into many regular square grids [10]. The measured fingerprints are allocated to each grid. If multiple fingerprints are measured in a single grid, the multiple fingerprints are merged into a single fingerprint. Shin and Cha [19] construct a topological map with regular Wi-Fi signal calibration points, assign semantically meaningful labels into the map, and estimate the semantic location of the user based on the current Wi-Fi observation. Chan et al. [20] and Kim et al. [21] build autonomous and collaborative RF fingerprint localization systems. They both use an indoor regular grid formation by anonymous mobile users who automatically collect data in daily life without purposefully surveying an entire building. Brunato and Battiti [22] use many regression-based algorithms to estimate position over the regular grids in indoor environment. Nafarieh and Ilow

[23] build a testbed with off-the-shelf equipment and the corresponding applications over the regular grid formation. The generality of a regular grid formation is also shown in metropolitan area applications. The study of Taipei [24] contains city-scale data samples. They use uniform calibration points in the city. In the work of Seattle [25], they collect a wide-scale training trace for the entire Seattle area. The trace data are stored in a radio map which has regular trace routes and a grid formation.

The regular grid fingerprint DB has a critical drawback to estimate precise position. In a general RF fingerprint WPS, the estimated position of a target point, that is, the position estimation point requested by a handset, is usually calculated by the weighted average of the fingerprint positions with the estimated probability. The regular grid structure merges the collected fingerprints to fit the regular grid segmentation, and the merged fingerprint positions are arranged to the center points of the grids. Then, the multiple candidate grids are selected by a position estimation algorithm. The center points of candidate grids are used to calculate the estimated position of a target point. That is, the center points of grids have a significant role in the position estimation. Figure 2(a) shows the misleadingness for center points in a regular grid fingerprint DB. In the case of a regular grid structure, the center point is assumed to be a representative point of merged fingerprints. Thus, the target point is easily misestimated. The target point  $T$  is definitely geographically close to the point  $A$ , but the measured fingerprint by a handset in  $T$  is more similar to the fingerprint of  $b$  than the fingerprint of  $a$ . Here, the fingerprint of  $a$  is allocated to the center point  $A$  and the fingerprint of  $b$  is allocated to the center point  $B$ . Then, the grid, which has center point  $B$ , is selected as one of candidate grids for a position estimation algorithm. The irregular grid setup in Figure 2(b) eliminates the aforementioned incorrect grid selection. Each measured fingerprint is fully analyzed: some of the fingerprints can be clustered and integrated under the statistical significance test. Then, the finally validated fingerprint by the significance test has its statistical significance among the other fingerprints in a reference DB.

The area of a grid in an irregular grid structure presents additional important information for position estimation. Each grid shows the dominant coverage of the specific fingerprint that has distinct statistical significance. The dominant coverage directly means the error bound of position estimation. The usual RF fingerprinting estimates the position of a target point as a center of grid. But, the actual position of the target point can be located anywhere in the selected grid. The coverage of fingerprint (i.e., grid) gives important information for error limit.

In this paper, we propose a design of irregular grid structure for RF fingerprinting and a statistical significance test framework for fingerprint data validation. The sufficiently valid fingerprint data are established by the proposed significance test and the grid area of each fingerprint is maintained with an effective level with an irregular form. As a result of the proposed significance test framework, we can create a standard RF fingerprint DB that consists of an effective, valid dataset. All fingerprint data used in the developed testbed are

harvested from actual radio fingerprint measurements taken throughout Seoul, Korea. This demonstrates the practical usefulness of the proposed framework.

## 2. Irregular Grid Segmentation of Fingerprint DB

Grid size is one of the main concerns of RF fingerprint WPS. Almost all the fingerprint data are collected by automatic scanning (usually using a vehicle) [26]. Then, the fingerprint data should be aligned to each grid in a reference fingerprint DB. Figure 3 shows three representative fingerprint alignment methods.

The most desirable situation is the uniform assignment case. Fingerprint collection in the training phase is uniformly performed on a grid map. A single, effective fingerprint can be assigned to each grid. Then, the estimation quality can be guaranteed. However, in most cases, fingerprint collection cannot be performed uniformly. The collected fingerprints are scattered: partially dense and partially scarce. Moreover, fingerprints are probably not collected at the center of grids because it is not easy to collect fingerprint data from the exact center of a regular grid. Even for the uniform assignment illustrated in Figure 3, the difference between aligned and actual collection point produces additional alignment error. Figure 4 shows the performance for three types of conventional single-size regular segmentation:  $20\text{ m} \times 20\text{ m}$ ,  $10\text{ m} \times 10\text{ m}$ , and  $5\text{ m} \times 5\text{ m}$ . To test grid segment variations, we organized three types of grid maps from the same raw fingerprint data. The collected raw fingerprint data can be simply merged into a single grid. Further, small-size grid segmentation can contain many fingerprint data holes. To compare the performance among the grid segmentations, we select 27 target points for position estimation and then we applied three representative estimation algorithms: KNN [9], probabilistic frame [14], and a hybrid method of pattern matching with probabilistic frame [16]. The bar chart in Figure 4 presents the ratio of grid segmentation which has the minimum estimation error (i.e., in case of KNN, 40.7% of target points have the best estimation quality in  $20\text{ m} \times 20\text{ m}$  segmentation, 33.3% in  $10\text{ m} \times 10\text{ m}$ , and 25.9% in  $5\text{ m} \times 5\text{ m}$ ).

Figure 4 strongly intends the performance indifference among regular grid segmentations. It is hard to find any relation between segmentation size and position estimation quality. This indifference comes from the ineffectiveness of a single-size regular grid segmentation and alignment method (i.e., simple merging and data hole marking). The simple merging or data hole marking for regular grid segmentations does not give any sufficient compensation effect on the collected fingerprint data. Therefore, we propose a totally novel structure for fingerprint DB map: variable and irregular grid segmentation.

Figure 5 shows the irregular grid segmentation by clustered merging. We replace simple merging with the clustered merging by the significance test, which ensures effective alignment of collected fingerprints to their respective grids. Fingerprints collected using any scanning devices can be

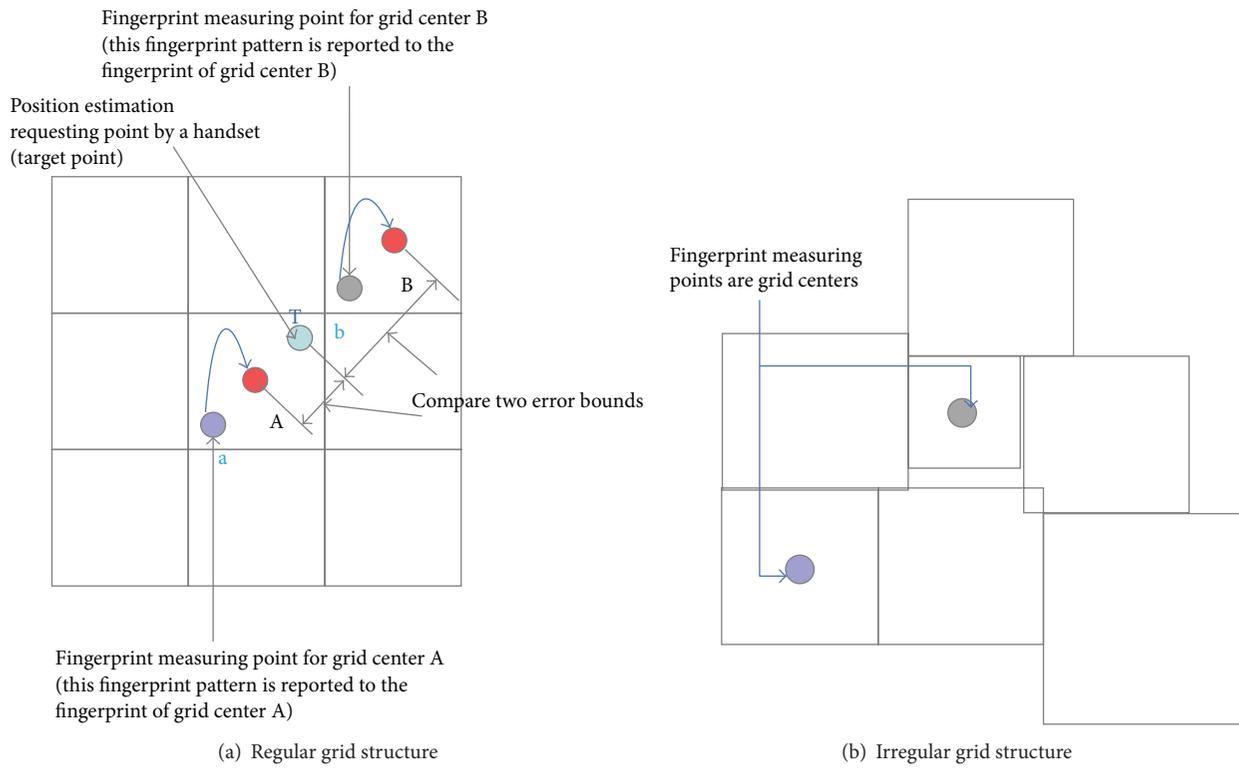
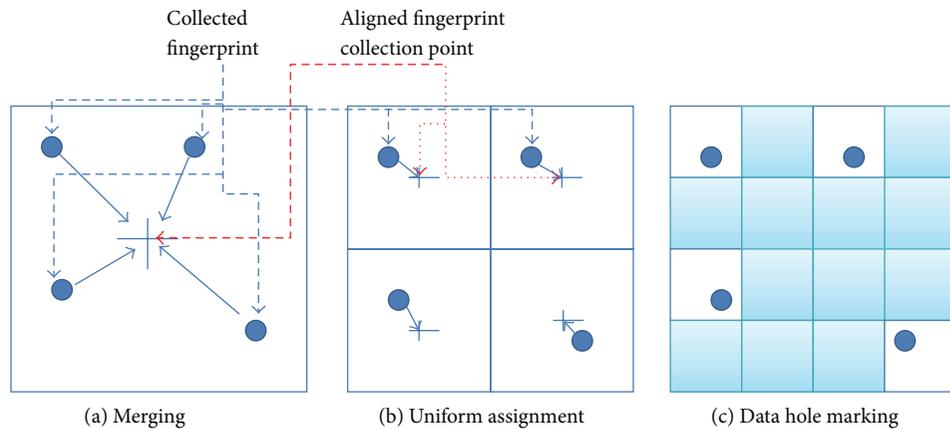


FIGURE 2: Comparison of regular and irregular forms of grids.



Grid segmentation level	Less segmented	Appropriate segmented	Over-segmented
Data density	Coarse data	Effective data for each grid	Scarce data
Estimation quality	Comparably low	Achievably high	Not determined

FIGURE 3: Grid segmentation level and fingerprint data alignment.

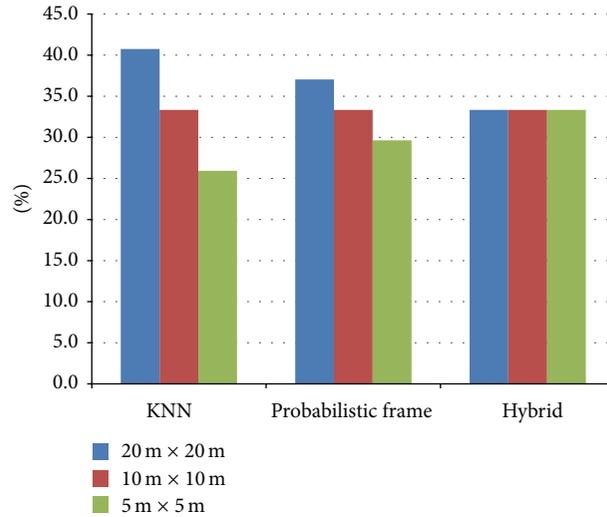


FIGURE 4: Performance comparison for grid segmentations.

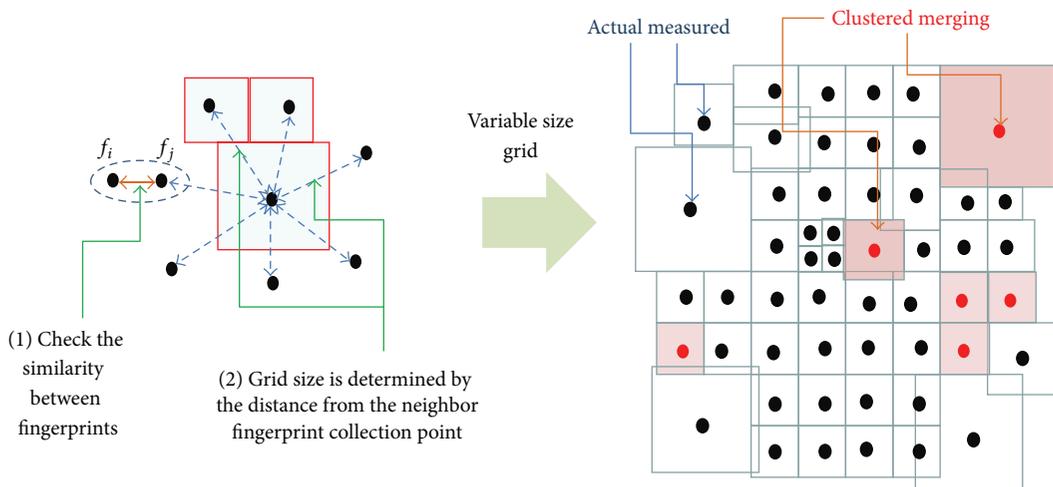


FIGURE 5: Making irregular grid segmentation.

appropriately operated (i.e., similarity test described in Section 3) and assigned to flexible-size grids. The grid size of a fingerprint is determined by the geographical relationship with the neighbor fingerprints. We can find the average distance to the neighbor fingerprint collection points and assign the proper grid size to each fingerprint. The valid fingerprint data stored in irregular grids guarantee the efficiency of data management and enhanced accuracy of position estimation in RF fingerprint WPSs. Additionally, the flexibility of the grid segmentation can be a useful strategy for the fingerprint collection process: a dense segmentation in commercial areas and a light segmentation in residential or rural areas. Next, we build the clustered merging by significance test framework to establish valid fingerprint data.

### 3. Clustered Merging by Significance Test

The usual fingerprint collection devices are operated automatically: they measure fingerprint along the collection

routes and store the measured fingerprints to a reference fingerprint DB without data calibration. Because of the automatic collection process, we can find lots of close fingerprint groups. The fingerprints in each group are collected at very close points with very similar fingerprint patterns. These fingerprints in a close group are hard to be assigned to the separated grids. Any position estimation algorithms have their own inherent error bound. Each of the small-size grids, within the inherent error bound, does not have its statistical differentiation. The clustered merging for the close fingerprint groups can generate statistically distinct fingerprints. Each fingerprint generated by the proposed clustered merging has sufficient statistical validity to each grid. The study of Kaufman and Rousseeuw [27] demonstrated the recent researches for group clustering. In an ordinary clustering, the member of cluster has a scalar value or a simple vector form (the elements of vector have solid and deterministic values). An RF fingerprint has also a vector form shown in Figure 1. But all elements of a fingerprint

vector are random variables. An RF fingerprint, a vector of random variables, has mathematical difficulty to be applied to an ordinary clustering. Therefore, we develop a special statistical tool for fingerprint clustering: significance test on fingerprints.

The proposed significance test is performed using the geographically close fingerprint group. The statistical difference between two fingerprints is based on the square of the Euclidean distance ( $d^2(i, j)$ ) between the two fingerprint pairs ( $f_i, f_j$ ). The Euclidean distance is a very well-known metric to measure the difference between two vectors. Feng et al. [28] adopted the Euclidean distance to measure the difference (or similarity) between two fingerprints with vector form as follows:

$$d^2(i, j) = (f_i - f_j)^2, \quad (1)$$

where  $f_i = \{\text{RSSI}_{\text{AP}_1}^i, \text{RSSI}_{\text{AP}_2}^i, \dots, \text{RSSI}_{\text{AP}_n}^i\}$ ,  $f_j = \{\text{RSSI}_{\text{AP}_1}^j, \text{RSSI}_{\text{AP}_2}^j, \dots, \text{RSSI}_{\text{AP}_n}^j\}$ .

Each value of  $\text{RSSI}_{\text{AP}_k}^i$  ( $\text{RSSI}$  for  $\text{AP}_k$  in fingerprint  $f_i$ ) is a random variable and has a measurement error that tends to follow a normal distribution. Thus, each element of vector  $f_i - f_j$  also follows a normal distribution. By transforming the elements of vector  $f_i - f_j$  to the standard normal distribution,  $d^2(i, j)$  tends to follow a chi-square distribution with a degree of freedom  $n$ ; that is,  $d^2(i, j) \sim \chi^2(n)$ . Generally, the  $\chi^2(n)$  has a mean  $n$  and variance  $2n$ . Moreover, the  $\sqrt{2d^2(i, j)}$  is approximately normally distributed with a mean  $\sqrt{2n - 1}$  and unit variance [29]. Based on the aforementioned statistical characteristic, we can determine a difference between the two fingerprints ( $f_i, f_j$ ). When the calculated  $\sqrt{2d^2(i, j)}$  value is greater than a certain threshold,  $f_i$  and  $f_j$  are determined to be two statistically different fingerprints (the calculation of a practical threshold is described in Section 4). However, this significance test between only two fingerprints has limitations when it is applied to the practical clustered merging. The significance test should be applied on a fingerprint group basis in an entire area. (Example groups are illustrated in Figure 6.)

Figure 6 shows examples of clustered groups ( $F_1 \sim F_4$ ). Each clustered group has geographical proximity. The fingerprint members in the same clustered group are determined by the following inclusion test:

$$\text{Inclusion test: } \sqrt{2d^2(f_i, F_l \setminus f_i)} \leq \text{inclusion threshold}, \quad (2)$$

where  $d^2(f_i, F_l \setminus f_i) = \text{square of Euclidian distance between } f_i \text{ and } \text{mean}(F_l \setminus f_i) \text{ and } \text{mean}(F_l \setminus f_i) \text{ denotes an artificial fingerprint with elements that are the mean values of elements for all fingerprints in } F_l, \text{ excluding } f_i. \text{ A fingerprint } f_i \text{ is set to be a member of the clustered group } F_l \text{ when } \sqrt{2d^2(f_i, F_l \setminus f_i)} \text{ is less than an inclusion threshold. An inclusion threshold is applicable as same as aforementioned two fingerprints cases. Some fingerprints can be considered to be clustered with multiple groups (see the clustered group } F_2 \text{ and its proximity$

in Figure 6). We perform a separation test for discriminating the clustered group partitions. Consider

$$\text{Separation test: compare } d^2(f_i, F_{l1}) \text{ and } d^2(f_i, F_{l2}). \quad (3)$$

If  $f_i$  is a candidate member of both clustered groups  $F_{l1}$  and  $F_{l2}$ , we compare a square of Euclidian distance between  $f_i$  and  $\text{mean}(F_{l1})$  (also,  $\text{mean}(F_{l2})$ ).  $f_i$  is added to a closer group in view of the Euclidian distance.

The aforementioned inclusion and separation test guarantee the strict group partition for clustered merging. However, as there are many fingerprints to be tested, we develop a practical clustered merging procedure that has an order of tested fingerprints. See Algorithm 1.

At the beginning of the procedure, all fingerprints belong to *unmarked set*  $S$ . A fingerprint  $f_i$  is selected in the set  $S$ . Then, the significance test is performed between the selected fingerprint and one of its neighbor fingerprints ( $j \in \text{neighbor}(i)$ ). Note that the selected fingerprint and its neighbor fingerprints share geographical proximity.

## 4. Numerical Results

To demonstrate the applicability of the proposed method, we collected whole fingerprint data from the Seoul Gangnam urban district. The WPS practices in actual city were performed in some previous works. In work of Cheng et al. [30], the 30-minute restricted scanning tried to approve the practicability of WPS. The work of Yoshida et al. [26] shows a 100 m  $\times$  100 m test district with 873 APs. The study in Sydney [10] has a 500 m  $\times$  800 m test district with 1300 APs and 172 reference grids. We significantly extended the applicability of fingerprint WPS to a metropolitan area of Seoul (the area of Seoul Gangnam is 39.55 km<sup>2</sup>) [31]. A single scanning process for Gangnam district usually generates 0.6 million fingerprint data. The scanning process was performed three times to collect fingerprint data. The total volume of collected fingerprint data exceeded 1.8 million fingerprints. This is a huge amount of data in a relatively large area. Figure 7 shows a simplified diagram for fingerprint collection by a scanning vehicle. A scanning vehicle runs through the metropolitan area to make an entire RF fingerprint DB in a form of a geographical map. For efficient fingerprint collection, a fingerprint collector segments an entire area into multiple fractions and builds efficient scanning routes for each fraction. The most popular method of building scanning routes is Chinese Postman Routing. Chinese Postman Routing is a very well-known postman tour or route inspection method of finding the shortest closed path or circuit that visits every edge of an (connected) undirected graph. This method can be used to obtain the optimal Eulerian circuit (a closed walk that covers every edge once).

The scanning vehicle includes a diagnostic machine (DM) with GPS receiver. The collected fingerprint data are stored in a temporary storage in a notebook PC connected to a DM. After the single collection route, the whole fingerprint data are transferred to a central storage for reference DB map. The usual Wi-Fi fingerprints are collected in the form of Table 1.

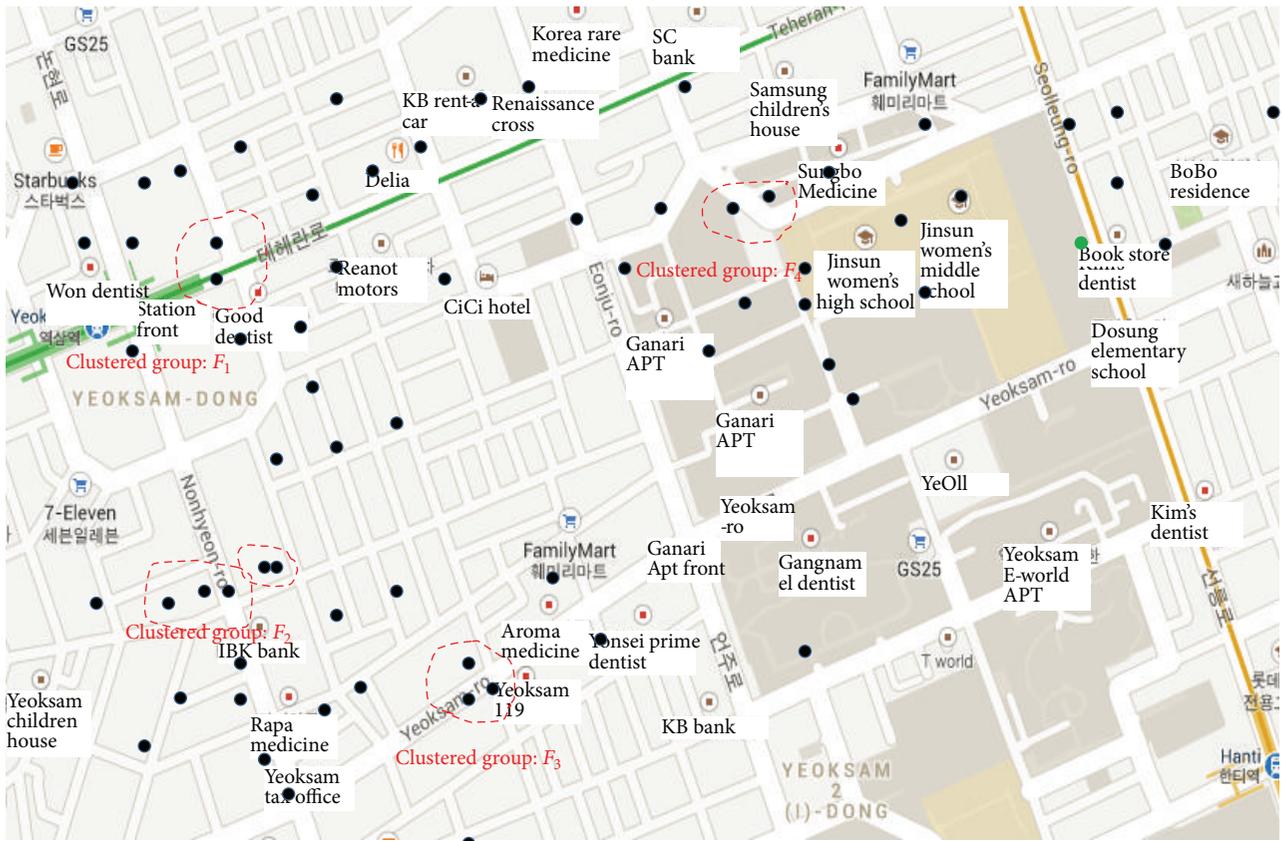


FIGURE 6: Clustered groups for the significance test.

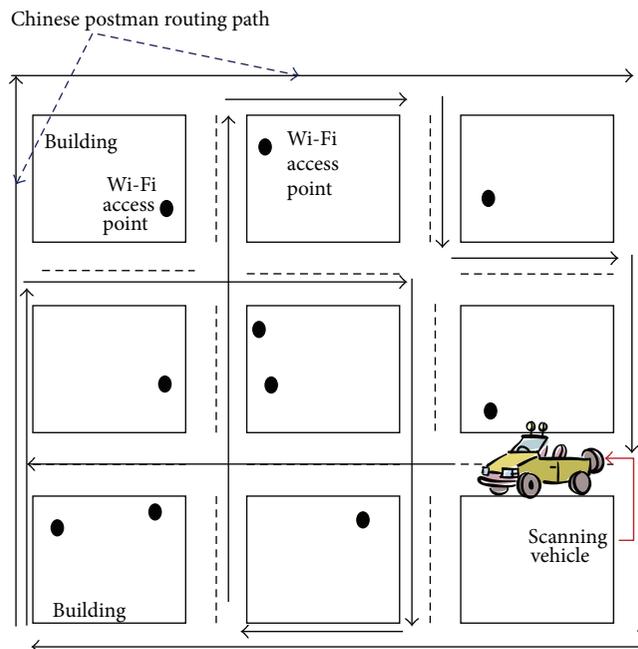
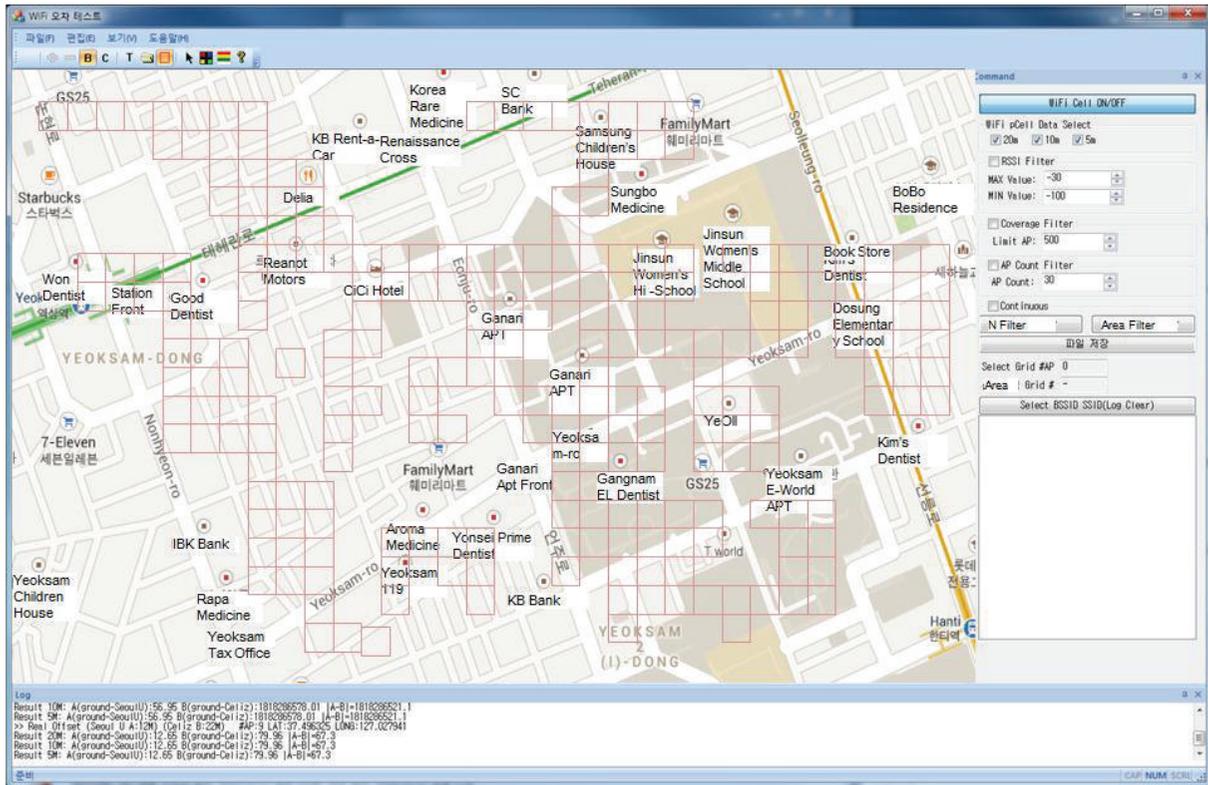


FIGURE 7: Scanning of RF fingerprints.



Note. All Korean characters mean names of places. The imported map uses Korean geographical interface

FIGURE 8: Example of single-size regular grid segmentation.

TABLE 1: Example of a Wi-Fi AP fingerprint.

BSSID	SSID	MES_X	MES_Y	RSSI (dB)
00:01:36:1f:9c:d2	sklifeap_4	37.5036	127.0336	-80
00:01:36:24:66:60	primebc-ap2	37.5060	127.0436	-69
00:01:36:25:2a:ea	Hpsetup	37.4956	127.0302	-79
00:01:36:25:2a:eb	SK_WLAN	37.4956	127.0302	-79
00:01:36:26:24:27	KWI-B2200T-	37.5038	127.0274	-87
00:01:36:26:24:28	D-1201	37.5038	127.0274	-83
00:01:36:27:4d:54	Tectura Corporation	37.5084	127.0434	-89
00:01:36:2a:84:f9	Default	37.4962	127.0302	-81

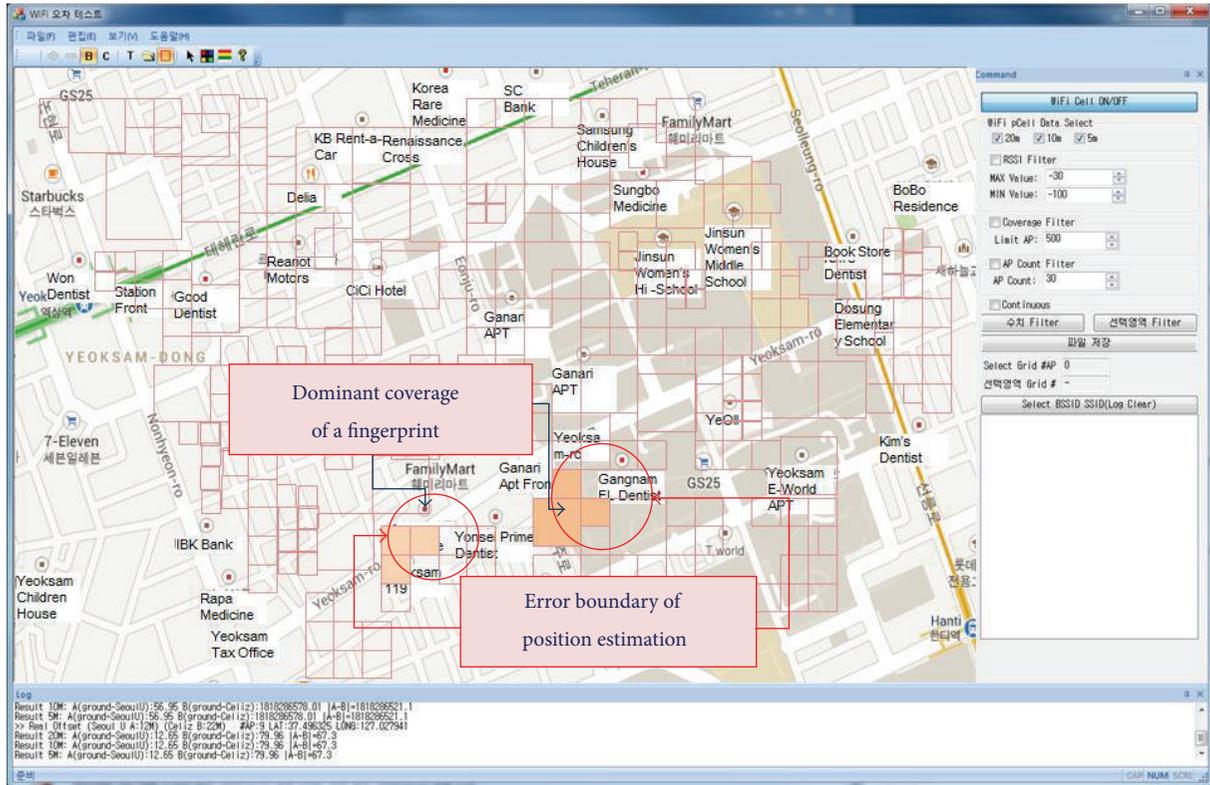
A Wi-Fi fingerprint consists of the following: a base station identification (BSSID; i.e., MAC address), service set identification (SSID), measurement  $x$ -axis (MES\_X, i.e., longitude), measurement  $y$ -axis (MES\_Y, i.e., latitude), and Received Signal Strength Index (RSSI). When an AP is detected by an automatic scanning device, the fingerprint data, that is, BSSID, SSID, and RSSI, are stored with their position, that is, MES\_X and MES\_Y.

As the first step, we applied the proposed data alignment method and evaluated its performance in a relatively restricted area for visualization (see the window-based performance analysis tool as shown in Figure 8). The test area is a square district (550 m  $\times$  350 m) in Seoul Gangnam. This district is classified as a commercial area that includes many

commercial buildings and dense foot traffic. In total, 1,682 APs were detected and 459 fingerprint data were measured. Figure 8 contains a single-size regular grid (20 m  $\times$  20 m) segmentation. There were 241 grids of regular size, and 52.3% of the sample area was covered by single-size grids.

Next, we applied irregular grid segmentation by clustered merging. Figure 9 shows the new segmentation of the sample area shown in Figure 8.

The proposed fingerprint data alignment method reconstructs entire grids in the sample area. In total, 289 grids were generated with irregular size. The total area covered by the proposed method was 72.1% of the sample area. The threshold value, which was applied to the significance and inclusion tests, was determined by the statistical characteristic of the



Note. All Korean characters mean names of places. The imported map uses Korean geographical interface

FIGURE 9: Example of irregular grid segmentation.

```

While unmarked set  $S \neq \emptyset$ 
  Select  $i \in S$ 
   $S = S - \{i\}$ 
  while  $neighbor(i) \neq \emptyset$ 
    Select  $j \in neighbor(i)$ 
     $neighbor(i) = neighbor(i) - \{j\}$ 
    do significance ( $f_i, f_j$ )
    if significance ( $f_i, f_j$ ) is "positive",
      set  $j \in F(i), S = S - \{j\}$ 
  proc significance ( $f_i, f_j$ )
  calculate square of Euclidean distance  $d^2(i, j) = (f_i - f_j)^2$ 
  If  $\sqrt{2d^2(i, j)} > threshold$ ,
    report "positive"
    
```

ALGORITHM 1

fingerprint difference. The fingerprint difference, that is,  $\sqrt{2d^2(i, j)}$  in Section 3, tends to follow a normal distribution with mean  $\sqrt{2n} - 1$  and unit variance; that is,  $\sigma = 1$ . Ideally, we can determine the two fingerprints, whose difference (i.e.,  $\sqrt{2d^2(i, j)}$ ) is greater than zero, as two separated vectors. However, all elements and their mathematical variations are random variables because of their measurement imperfectness. That is, the mean  $\sqrt{2n} - 1$  is observed as an average difference between two fingerprints exactly speaking,  $\sqrt{2} \times$

(Euclidean distance of two fingerprints). We can specify a certain relative difference level from an average difference to determine similarity between two fingerprints. For example, when an average difference is observed as a value  $a$ , we can set  $ka$  ( $0 < k < 1$ ) as a threshold for significance test. Now, finding a normal distribution of  $\sqrt{2d^2(i, j)}$ , we can use  $\sqrt{2n} - 1 - m$  as a useful threshold value; that is, relative difference level can be controlled by  $m$ . By the simple specification of  $m$ , we can make a statistically meaningful threshold. If  $m = 1$ , then the two fingerprints, which have a

TABLE 2: Grids difference for 10 districts.

District	1	2	3	4	5	6	7	8	9	10
Regular	256	312	261	292	304	297	214	223	196	269
Irregular	287	332	296	338	341	332	248	236	221	286

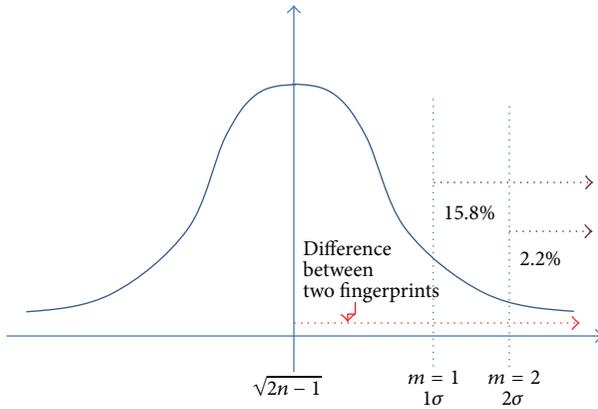


FIGURE 10: Statistical threshold specification.

difference belonging to a lower 15.8% of the whole difference distribution, are determined to be statistically different. To test the effectiveness of irregular grid segmentation, we set  $m = 2$ , that is, lower 2.2% of the total difference distribution, and compared the regular grid segmentation (see Figure 10).

Figure 11 shows detailed results of the position estimation for 20 sample target points. The  $x$ -axis denotes the estimation error in meter and the  $y$ -axis denotes the cumulative number of sample target points. The irregular grid segmentation generated superior estimations for all the target points for all the applied estimation algorithms (e.g., for KNN, 18 points have the error bound within 55 m in case of irregular segmentation, but only 12 points within 55 m error bound in case of regular segmentation). Three representative estimation algorithms were applied: KNN [9], probabilistic frame [14], and a hybrid method of pattern matching with probabilistic frame [16]. Figure 11(d) shows the performance difference among estimation algorithms. The hybrid method has slightly higher performance compared to KNN or probabilistic method.

Honestly, a part of contribution is due to the increment of grids. 19.9% increased grids of irregular grid segmentation (the difference between regular and irregular grids is 48) make more precise position estimation. However, we could observe much higher extra enhancement. Even considering the 19.9% increment of grids, the 37.5% precision enhancement is observed for hybrid positioning method, 34.7% for KNN, and 35.1% for probabilistic.

Next, we extended our irregular grid segmentation using the significance test in a large area. Ten test districts (see Figure 12) in Seoul Gangnam were selected to prove the applicability of the proposed irregular grid segmentation. The area of Gangnam district is  $39.55 \text{ km}^2$ . The range of area for

test districts is  $0.10 \text{ km}^2 \sim 0.17 \text{ km}^2$ . Each district has 50 ~ 60 target points for position estimation.

Figure 13 shows the estimation results which prove the effectiveness of the proposed method in various diversified environments of an urban area. For comparison purposes, Figure 13 contains the estimation error of the regular grid segmentation for three representative estimation algorithms (KNN, probabilistic, and hybrid). We measured on average 30% enhancement for KNN, 29% for probabilistic, and 26% for hybrid position estimations.

To show the effectiveness of the proposed irregular segmentation itself, the numbers of grids for 10 districts are presented in Table 2.

The average increment of grids was 11.2%. However, 26.9% precision enhancement was observed for hybrid positioning method, 25.7% for KNN, and 30.1% for probabilistic. The presented comparison between “increment of grids” and “precision enhancement” shows the significant advantage of irregular grid segmentation. The simple increment of grids cannot guarantee such a significant enhancement.

Additionally, we were able to show the radius of the error boundary (see Figure 14). The position estimation algorithm deployed in the experiment selected the best matching grids in multiple candidate grids, as shown in Figure 9. The circular boundary of the candidate grids was considered as the estimation error boundary. The proposed irregular grid segmentation was better able to illustrate the dominant coverage of each fingerprint. Then, the radius of error boundary can be useful for the qualification of position estimation.

Note that the majority of test points belong to the outdoor environment. The automatic scanning vehicle has a problem to access indoor environment. Most of the indoor fingerprints are collected by human power. Thus, our experiment has a limitation for the applicability on the indoor environment. However, our proposed segmentation is applicable both on indoor and on outdoor environment. The radio signal fluctuation and structure complexity are more serious in the indoor environment. The proposed irregular segmentation has relative advantage on the complex and fluctuated environment. We carefully expect the effective applying to the indoor environment.

## 5. Conclusion

The rapid growth of mobile communication and the proliferation of smartphones have drawn significant attention to location-based services. One of the most important factors in the vitalization of LBS is the accurate position estimation of a mobile device. Traditional triangulation has an inevitable weakness when estimating the exact position of an AP. Moreover, significant technical advances are not shared publicly

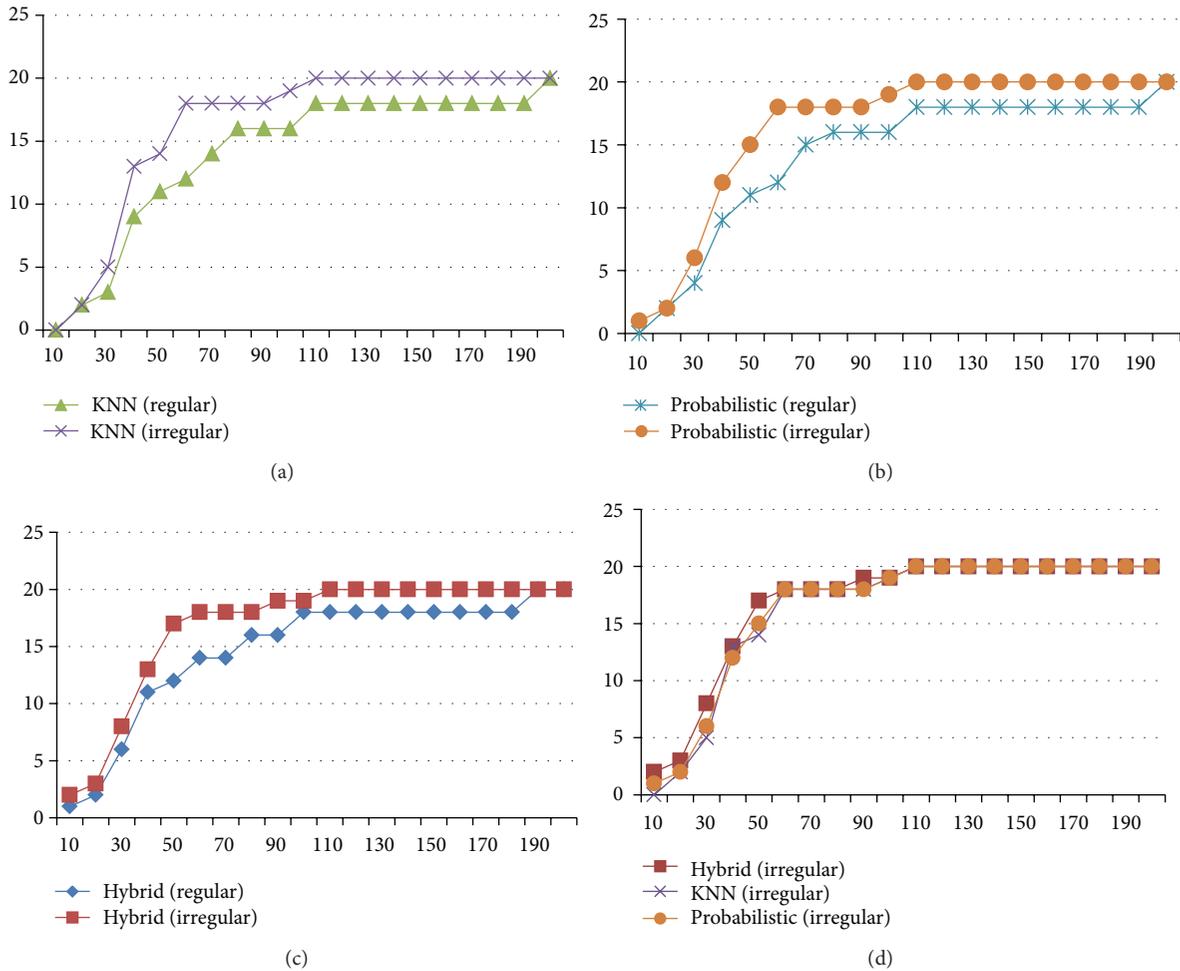


FIGURE 11: Position estimation error for sample points.

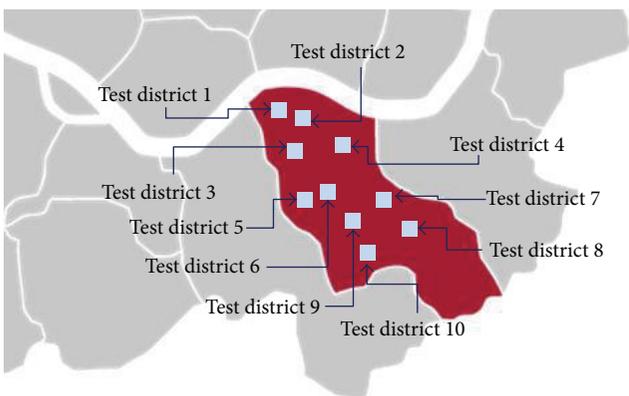


FIGURE 12: Test districts in Seoul Gangnam.

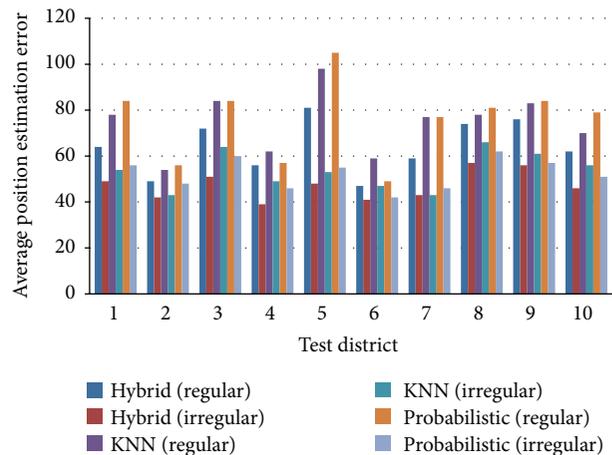


FIGURE 13: Comparison of regular and irregular grid segmentation.

by solution providers. An RF fingerprint WPS is another valuable way to penetrate the positioning solution provider market. Even using indiscriminate fingerprint collections, providers can build an approximate fingerprint DB and apply a simple pattern-matching algorithm for position estimation.

However, to build a competitive fingerprint WPS solution, we should focus on the fingerprint data management and precise estimation algorithms. Much of enhancement on fingerprint

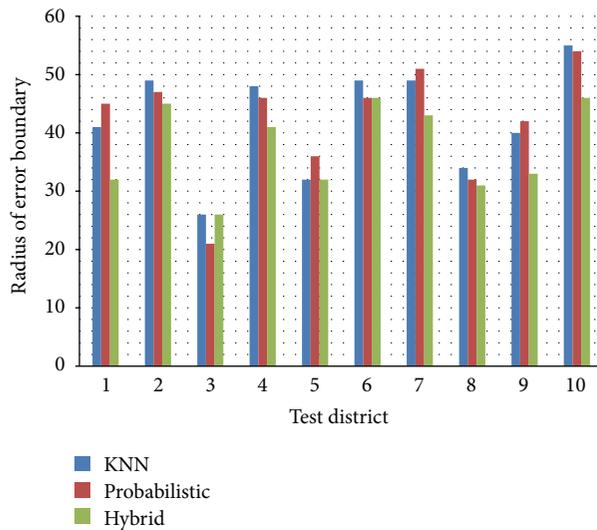


FIGURE 14: Comparison of radius of error boundary.

WPS focuses on an estimation algorithm itself. However, the improvement by estimation algorithms faces limitation. The more essential factor of fingerprint WPS is a structure of radio fingerprint map. Because of the geographical complexity of urban areas, similar, even duplicated, fingerprint data are collected at close points. Therefore, we presented a data clustering method and irregular grid segmentation. Based on the statistical significance test for fingerprints, collected fingerprints were merged. Each fingerprint grid had an irregular form to cover the geographical area. The proposed new fingerprint data management for position estimation can strengthen the advantages of the fingerprint WPS. Compared with conventional fingerprint data alignment approaches, our method achieves better performance in both average error of estimation and deviation of errors. Furthermore, all the fingerprint data were harvested from an actual measurement of RF fingerprints from the Gangnam district, Seoul. We built an irregular fingerprint map for an entire area of Seoul and applied position estimation. These trials show the practical usefulness of the proposed methodology.

## Conflict of Interests

The authors declare no conflict of interests.

## Acknowledgments

This research work is supported by SK Telecom in South Korea. All collected data are obtained using the facility of SK Telecom. This work was also supported by the National Research Foundation of Korea (NRF) grant funded by the Korean Government (2011-0011825).

## References

[1] Skyhook Wireless, "Estimation of positioning using WLAN access point radio propagation characteristics in a WLAN

positioning system," US Patent 7515578, World Intellectual Property Organization, 2007.

- [2] J. P. Pavon and S. Choi, "Link adaptation strategy for IEEE 802.11 WLAN via received signal strength measurement," in *Proceedings of the IEEE International Conference on Communications (ICC '03)*, vol. 2, pp. 1108–1113, Anchorage, Alaska, USA, 2003.
- [3] Y. Masumoto, "Global Positioning System," US Patent 5,210,540, 1993.
- [4] J. M. Watters, "Combining GPS with TOA/TDOA of cellular signals to locate terminal," US Patent no. 5982324, May 1998.
- [5] Skyhook Wireless, "Skyhook Wireless technology used in revolutionary iPhone and iPod touch," 2008, <http://www.businesswire.com/news/home/20080116005162/en/Skyhook-Wireless-Technology-Revolutionary-iPhone-iPod-touch>.
- [6] Skyhook Wireless, "Location beacon database and server, method of building location beacon database, and location based service using same," US 62310804, World Intellectual Property Organization, 2004.
- [7] R. Mardeni and T. S. Priya, "Optimised COST-231 Hata models for WiMAX path loss prediction in suburban and open urban environments," *Modern Applied Science*, vol. 4, no. 9, 2010.
- [8] A. Medeisis and A. Kajackas, "On the use of the universal Okumura-Hata propagation prediction model in rural areas," in *Proceedings of the IEEE 51st Vehicular Technology Conference Proceedings (VTC '00)*, vol. 3, Tokyo, Japan.
- [9] B. Li, J. Salter, A. G. Dempster, and R. Rizos, "Indoor positioning techniques based on wireless LAN," in *Proceedings of the 1st IEEE International Conference on Wireless Broadband and Ultra Wideband Communications (AusWireless '06)*, Sydney, Australia, March 2006.
- [10] B. Li, I. J. Quader, and A. G. Dempster, "On outdoor positioning with Wi-Fi," *Positioning*, vol. 1, no. 13, 2008.
- [11] S. Das, T. Teixeira, and S. F. Hasan, "Research issues related to trilateration and fingerprinting methods," *International Journal of Research in Wireless Systems*, vol. 1, pp. 33–35, 2012.
- [12] D. Pandya, R. Jain, and E. Lupu, "Indoor location estimation using multiple wireless technologies," in *Proceeding of the 14th IEEE 2003 International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '03)*, vol. 3, pp. 2208–2212, September 2003.
- [13] R. Battiti, T. L. Nhat, and A. Villani, "Location-aware computing: a neural network model for determining location in wireless LANs," Technical Report DIT-02-0083, University of Trento, Trento, Italy, 2002.
- [14] T. Roos, P. Myllymäki, H. Tirri, P. Misikangas, and J. Sievänen, "A Probabilistic Approach to WLAN User Location Estimation," *International Journal of Wireless Information Networks*, vol. 9, no. 3, pp. 155–164, 2002.
- [15] A. Haeberlen, E. Flannery, A. M. Ladd, A. Rudys, D. S. Wallach, and L. E. Kavvaki, "Practical robust localization over large-scale 802.11 wireless networks," in *Proceedings of the 10th Annual International Conference on Mobile Computing and Networking*, pp. 70–84, Philadelphia, Pa, USA, 2004.
- [16] J. H. Kim, H. I. Choi, D. S. Choi, and S. Y. Kang, "A 2-stage hybrid position estimation framework," *Wireless Networks*, vol. 20, no. 6, pp. 1541–1556, 2014.
- [17] J. H. Kim and W. Y. Yeo, "A coherent data filtering method for large scale RF fingerprint Wi-Fi Positioning Systems," *Eurasip Journal on Wireless Communications and Networking*, vol. 2014, article 13, 2014.

- [18] G. Welch and G. Bishop, "An introduction to the Kalman Filter," Technical Report 2006, University of North Carolina, Wilmington, NC, USA, 2006.
- [19] H. Shin and H. Cha, "Wi-Fi fingerprint-based topological map building for indoor user tracking," in *Proceedings of the 16th IEEE International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA '10)*, pp. 105–113, August 2010.
- [20] E. C. L. Chan, G. Baciuc, and S. C. Mak, "Using wi-fi signal strength to localize in Wireless sensor networks," in *Proceedings of the International Conference on Communications and Mobile Computing*, pp. 538–542, Yunnan, China, January 2009.
- [21] Y. Kim, Y. Chon, and H. Cha, "Smartphone-based collaborative and autonomous radio fingerprinting," *IEEE Transactions on Systems, Man and Cybernetics C: Applications and Reviews*, vol. 42, no. 1, pp. 112–122, 2012.
- [22] M. Brunato and R. Battiti, "Statistical learning theory for location fingerprinting in wireless LANs," *Computer Networks*, vol. 47, no. 6, pp. 825–845, 2005.
- [23] A. Nafarieh and J. Ilow, "A testbed for localizing wireless LAN devices using received signal strength," in *Proceeding of the 6th Annual Communication Networks and Services Research Conference (CNSR '08)*, pp. 481–487, Halifax, Canada, May 2008.
- [24] A. W. T. Tsui, W.-C. Lin, W.-J. Chen, P. Huang, and H.-H. Chu, "Accuracy performance analysis between war driving and war walking in metropolitan Wi-Fi localization," *IEEE Transactions on Mobile Computing*, vol. 9, no. 11, pp. 1551–1562, 2010.
- [25] M. Y. Chen, T. Sohn, D. Chmelev et al., "Practical metropolitan-scale positioning for GSM phones," in *Ubiquitous Computing*, vol. 4206 of *Lecture Notes in Computer Science*, pp. 225–242, Springer, Berlin, Germany, 2006.
- [26] H. Yoshida, I. Seigo, and N. Kawaguchi, "Evaluation of pre-acquisition methods for position estimation system using wireless LAN," in *Proceedings of the 3rd International Conference on Mobile Computing and Ubiquitous Networking*, pp. 148–155, London, UK, October 2006.
- [27] L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*, John Wiley & Sons, 2009.
- [28] C. Feng, W. S. A. Au, S. Valaee, and Z. Tan, "Received-signal-strength-based indoor positioning using compressive sensing," *IEEE Transactions on Mobile Computing*, vol. 11, no. 12, pp. 1983–1993, 2012.
- [29] R. A. Fisher, *Statistical Methods for Research Workers*, Oliver and Boyd, 1954.
- [30] Y.-C. Cheng, Y. Chawathe, A. Lamarca, and J. Krumm, "Accuracy characterization for metropolitan-scale Wi-Fi localization," in *Proceedings of the 3rd International Conference on Mobile Systems, Applications, and Services (MobiSys '05)*, pp. 233–245, Seattle, Wash, USA, June 2005.
- [31] J. H. Kim and W. Y. Yeo, "Selective RF fingerprint scanning for large-scale Wi-Fi positioning systems," *Journal of Network and Systems Management*, 2014.



**Hindawi**

Submit your manuscripts at  
<http://www.hindawi.com>

