

RESEARCH

Open Access

A coherent data filtering method for large scale RF fingerprint Wi-Fi Positioning Systems

Jae-Hoon Kim¹ and Woon-Young Yeo^{2*}

Abstract

The rapid growth of mobile communication and the proliferation of smart phones have drawn significant attention to location-based services (LBS). The Wi-Fi positioning system (WPS) is a newly attractive method as a widely applicable positioning technique in LBS. In WPS, the received signal strength indication (RSSI) data of all Wi-Fi access points (APs) are measured, and stored in a huge database, as a form of radio fingerprint map. Because of the millions of APs in urban areas, radio fingerprint data are seriously contaminated. Therefore, we present a coherent filtering method for radio fingerprint data. All fingerprints used in the developed test bed are harvested from actual radio fingerprint measurements taken throughout Seoul, Korea. This demonstrates the practical usefulness of the proposed methodology.

Keywords: Location based service; Fingerprint; Data filtering; Wi-Fi positioning system

1. Introduction

To improve its relevance, context, and economic value, a location-based service (LBS) coordinates user location with various end-user applications. Despite the many possibilities offered by LBS, its market penetration has been slow. Most early-stage services failed to spread to the mass market. Moreover, monetization of the services is limited to some special purpose markets, such as car map/navigation. The limitations of LBS 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, while the deviation can be of the order of hundreds of meters. The demonstration of Figure 1 shows the layered technical structure for LBS. Among technical layers, the position quality is essentially determined by the positioning system.

The positioning system measures the estimated position of a moving object (usually a mobile handset), and then minimizes the difference between the actual and estimated position. A well-estimated position can reduce the practical and emotional disjunction caused by the position difference.

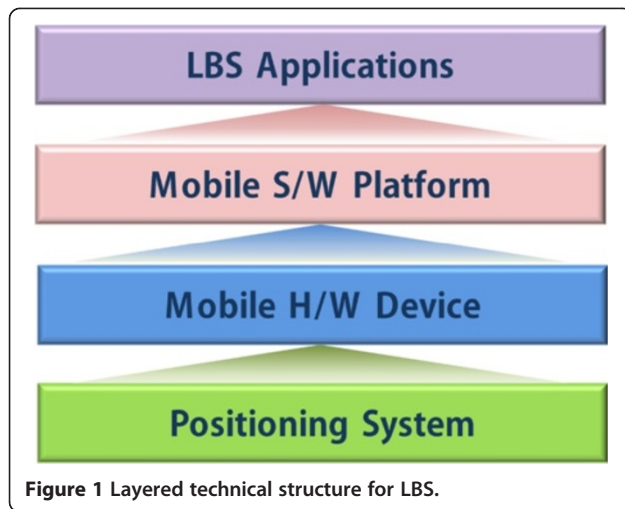
The global positioning system (GPS) provides very precise positioning estimation [1,2]. Owing to current advances in GPS technology, GPS receivers can acquire GPS signals with power levels as low as -160 dBm. However, the GPS signal should be attenuated significantly when it travels through construction materials, or any other obstacle (note that signal strength of at least -145 dBm is needed to acquire ephemeris data; the unobstructed GPS signal strength on Earth is measured at about -130 dBm.) This attenuation of a GPS signal makes it difficult to find a sufficient number of GPS satellites for each receiver in the urban environment (in GPS triangulation, at least three satellites are required to identify a current position). Presently, LBS are mainly used in urban areas, even in indoor environments. Social networking, friend finders, and local search applications need extremely wide coverage of positioning technology with a very short time to first fix (TTFF). LBS applications are shifting from rural areas, highways, and arterial roads, to urban and metropolitan areas.

With the rapid increase in Wi-Fi access points (APs) in metropolitan areas, Wi-Fi can be used as a viable alternative positioning infrastructure [3,4]. Each Wi-Fi AP generates a radio signal with a unique identifier or media access control (MAC) address every second, which enables mobile devices to identify the specific AP. The millions of public/private Wi-Fi APs can be used for

* Correspondence: woonyoungyeo@gmail.com

²Department of Information and Communications Engineering, Sejong University, Seoul, Republic of Korea

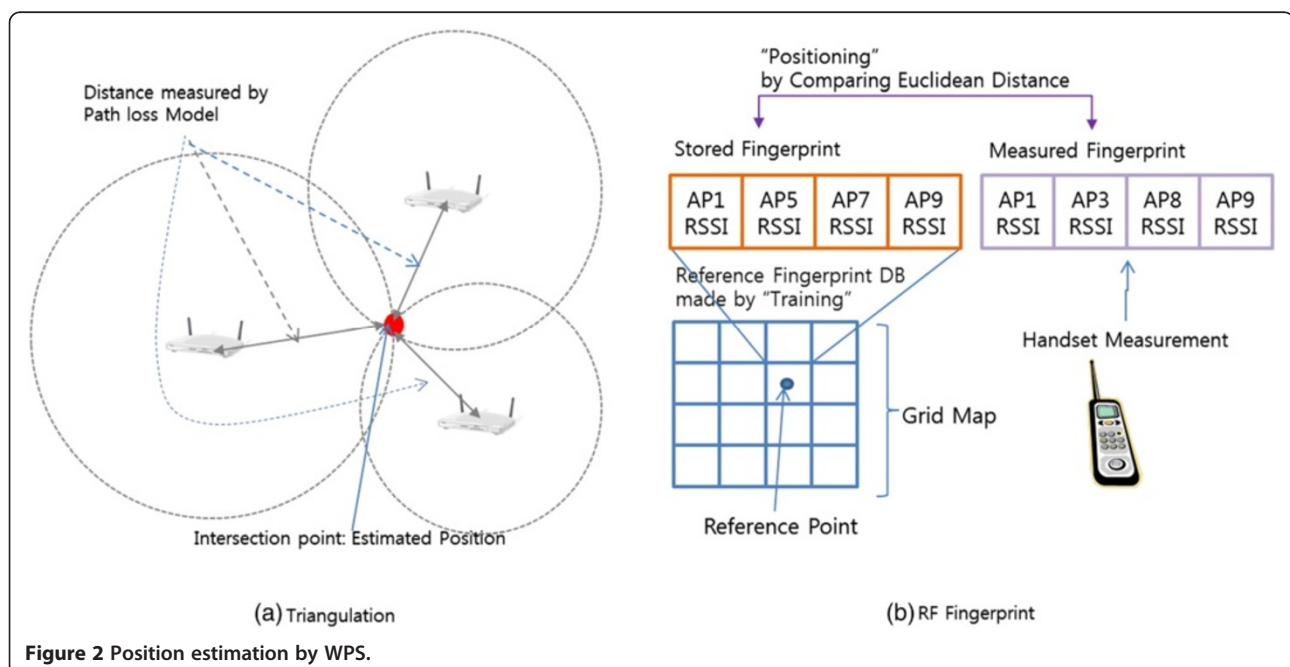
Full list of author information is available at the end of the article



Wi-Fi-based positioning. On the basis of the received signal strength 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 GPS, Wi-Fi positioning works better in urban canyons or indoor environments, than in rural area. It works well in dense metropolitan areas, both outdoors and indoors, owing to its greater-received signal strength and lower attenuation. The two major approaches that implement Wi-Fi positioning are AP triangulation, and radio frequency (RF) fingerprint. Triangulation is simple to implement [3,5,6]. As seen in

Figure 2a, three reference APs with already known coordinates are needed. After measuring the distance from the APs and a target point, three circles can be drawn. The circles intersect at one point, which is the target point. The coordinate of the target point can easily be calculated by the distance from, and the known coordinates, of the APs.

The main difficulty of this approach is measuring the distance from each AP to the target point. The typical path loss models (such as COST231, Okumura-Hata) are generally applied to measure the distance. However, it is extremely difficult to build a good and general model for distance measurement, which coincides with the actual field situation. RF fingerprinting [7,8] consists of two phases - training and positioning - demonstrated in Figure 2b. 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 usually considered the reference points. During the positioning phase, the position of a target point can be identified by comparing its measured fingerprint with the pre-stored reference fingerprint DB. The main advantage of RF fingerprinting is algorithmic simplicity. Simple comparing algorithms, such as pattern matching, can be easily applied to the practical process of position estimation. Then the RF fingerprinting is currently more preferred than triangulation [9]. The most advancement for RF fingerprinting has been searched in the area of position estimation algorithms. A pattern matching algorithm is used to determine the geographical position of a



target point. When a fingerprint pattern of the test point is measured by a portable device (usually a handset), an algorithm compares the measured fingerprint with elements of the reference fingerprint DB. The most well-known pattern-matching algorithm is nearest neighbor (NN) [7]. As an enhanced version of the NN algorithm, the K nearest neighbor (KNN) algorithm can be taken into account [7]. The average of coordinates of k-reference grids can be used to determine the estimated position of the target point (see Figure 3 for a detailed procedure of pattern matching). Various variations, such as smallest polygon [10] and neural networks [11], 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 given the measured fingerprint pattern at the target point. The probabilistic likelihood can be modeled by Histogram [12], Gaussian [11], Log-normal [13], or Kernel [12].

The main challenge related to RF fingerprinting is the creation and maintenance of an up-to-date reference fingerprint DB, which is especially difficult, due to the dynamic character of APs: they are often moved, or in some cases are temporary, with new APs being continuously deployed. The huge numbers of APs, both indoor and outdoor, are generally deployed in the urban environment. These cause serious complexity to the fingerprint DB management. In addition, the RSSI (measured

in dB unit) of each AP is an element of pattern vector. The difference between reference pattern and measured pattern determines the similarity of two patterns. Usually, lots of information (i.e., number of elements in pattern vector) provides confidence to similarity estimation. However, we should consider the measurement error and characteristics of radio signal strength. Because of environmental interference, the measurement error is inevitable. We can observe usually ± 3 dB or higher signal strength fluctuation. The important thing is that the unit 'decibel (dB)' has log scale. That is, the difference between -75 and -87 dB means not just '12 ($= -75$ to -87)' degree separation. But, the signal power of -87 dB is $1/16$ (i.e., 6.25%) of that of -75 dB. Same fluctuation on electromagnetic field gives extremely higher effect to low RSSI. Small signal power change on low RSSI makes a large change on the dB scale. AP information with low RSSI should be filtered (i.e., zero or very low weight on the AP with low RSSI) to guarantee estimation quality. A set of data filtering methods should be applied as a key management framework in a complex fingerprint DB.

The most popular common data filtering method is the Kalman filter [14]. The historically measured data can be unified with newly measured data, to eliminate data noise. However, the Kalman filter can just be applied to statistical outliers, by comparing the historic and current data. It is too complicated to apply to practical large-scale fingerprint DB, and also, it cannot make any common cutover threshold for consistent

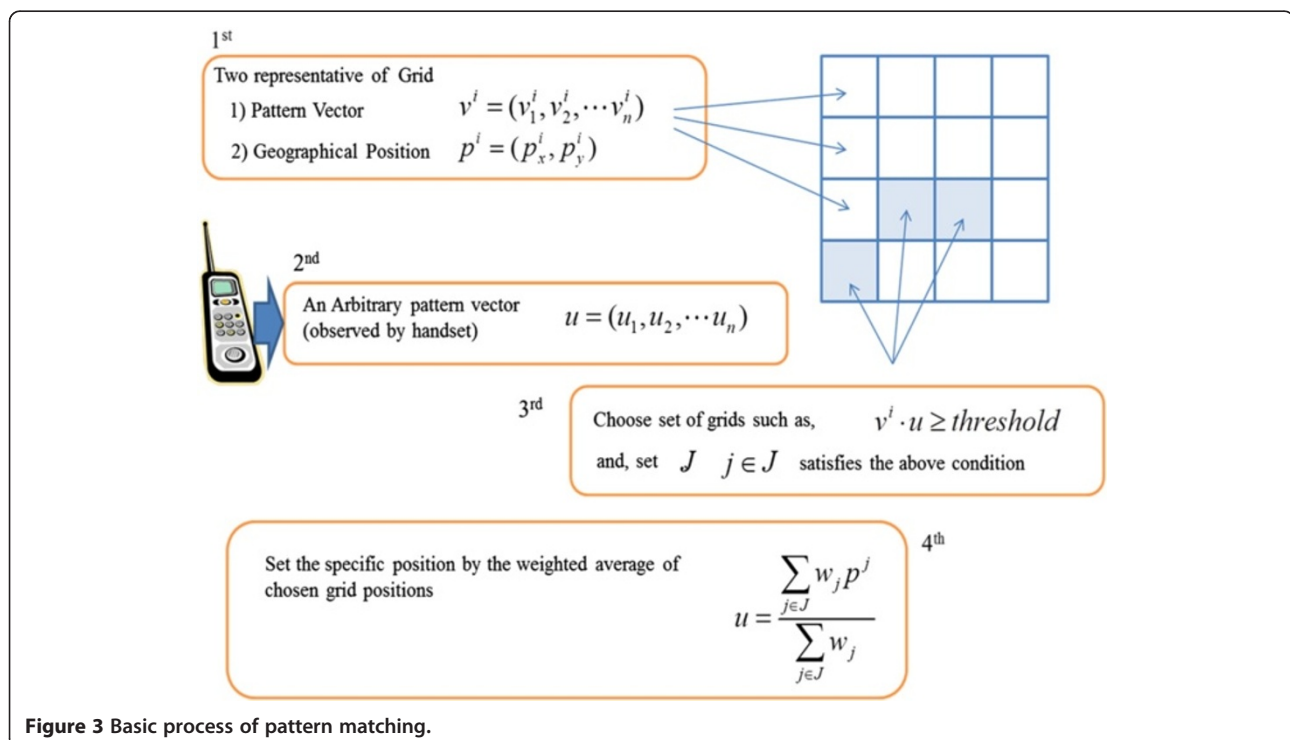


Figure 3 Basic process of pattern matching.

filtering in RF fingerprinting (the importance of consistent filtering in fingerprint WPS is described in Section 3).

Therefore, we propose a coherent data-filtering framework for standard RF fingerprint DB. The sufficiently valid fingerprint data are maintained by the proposed filtering framework. The number of APs for each reference fingerprint is maintained, with an effective level. The entire framework is described by an integer programming model. Moreover, we simultaneously propose a practical procedure for filtering, as a form of dynamic algorithm using an iterative function. As a result of a coherent filtering framework, we can make a standard RF fingerprint DB, which is full of effective and valid data sets. All fingerprint data used in the developed test bed are harvested from actual radio fingerprint measurements taken throughout Seoul, Korea. This demonstrates the practical usefulness of the proposed framework.

2. RF fingerprint map

A reference RF fingerprint DB constitutes the fundamental basic information of the proposed WPS. Most position estimation systems that use Wi-Fi APs require prior knowledge of the Wi-Fi RF fingerprint. The usual Wi-Fi fingerprints are collected in the form of Table 1.

The Wi-Fi fingerprint consists of 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 (i.e., BSSID, SSID, RSSI) are stored with its position (i.e., MES_X, MES_Y). These fingerprint data should be stored and handled as a DB map. A conventional reference fingerprint DB map consists of many grids. Each grid has RF fingerprint data consisting of AP identifiers (usually the MAC address of an AP), and the signal strength of each AP (see Figure 4).

In general, fingerprint data for reference DB are collected by wardriving. Wardriving is a data acquisition method for a position estimation system using Wi-Fi, and it is suitable for collecting fingerprint data over a

wide range [15]. We collected the entire fingerprint data of the Seoul Gangnam urban area, which contains about 110,000 grids. Figure 5 shows a sample of a reference fingerprint DB map, with grid representation.

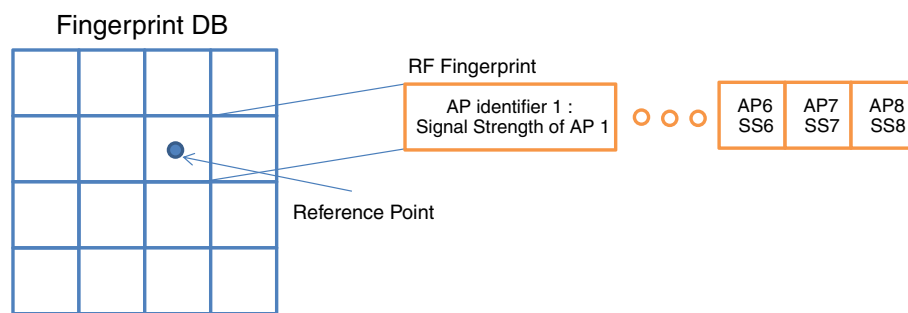
An up-to-date reference fingerprint DB should be maintained for precise position estimation. The segmented rescanning can be a promising way to update reference DB. The entire area is segmented into unit areas. Each unit area has distinguishable geographical characteristics, such as commercial, residential, and industrial or so. Then, a sample set of reference points is selected in a respective unit area. In general, the number of sample reference points is very small compared to total reference points, and the sample reference points are scattered uniformly throughout the unit area. We monitor the fingerprints of sample reference points periodically. The proportion of sample reference points, which have significantly different fingerprint pair for two time-consecutive monitoring epochs, is the determinant of area rescanning. If the proportion of significantly different fingerprint pairs is higher than a pre-specified ratio, we can conclude that the stored fingerprints of a unit area are not 'valid' (i.e., data is outdated), and we need to replace all fingerprints of the unit area by rescanning. Otherwise, the fingerprints of the unit area are verified as still valid, and remained to the reference fingerprint DB. Note that RF fingerprint DB is not stored in each handset in real applications. All applications send requests of position estimation to a location server which contains entire RF fingerprint DB.

A reference fingerprint DB map gives two types of information: 'the number of APs per grid', and 'the coverage of an AP'. The number of APs per grid determines the size of fingerprint DB map. The size of the map has a strong relationship with both the calculation speed and accuracy of position estimation. A set of APs that have relatively low RSSI values has limited effect for position estimation. The valuable fingerprint data are obtained from the APs that have higher RSSI values. The higher RSSI value means closer location from a reference point of the fingerprint DB map. Figure 6 shows the coverage of an AP. The coverage is shown as a set of detected grids of the AP. By the change of RSSI cutover threshold from -90 to -55 dB, the coverage of an AP shrinks to a smaller range.

The proper cutover threshold filters out the ineffective APs, decreases the number of detected APs per grid, and then finally restricts the coverage of the APs. The restricted AP coverage guarantees the higher likelihood for fixation of position estimation. Therefore, the determination of a cutover threshold for each AP is the essential point for reference fingerprint DB management. The key of the proposed coherent data filtering framework is the determination of the cutover threshold.

Table 1 Example of the Wi-Fi AP fingerprint

BSSID	SSID	MES_X	MES_Y	RSSI (dB)
00:01:36:1f9c:d2	Skilifeap_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



3. Coherent data filtering and compensation

exceeds hundreds of megabytes. This huge volume of fingerprint data has a significant negative effect on both the running speed of the positioning estimation algorithm, and reference fingerprint DB map maintenance. Moreover, the dimension of fingerprint should be restricted, for practical pattern matching type position estimation algorithms (the dimension of a fingerprint is



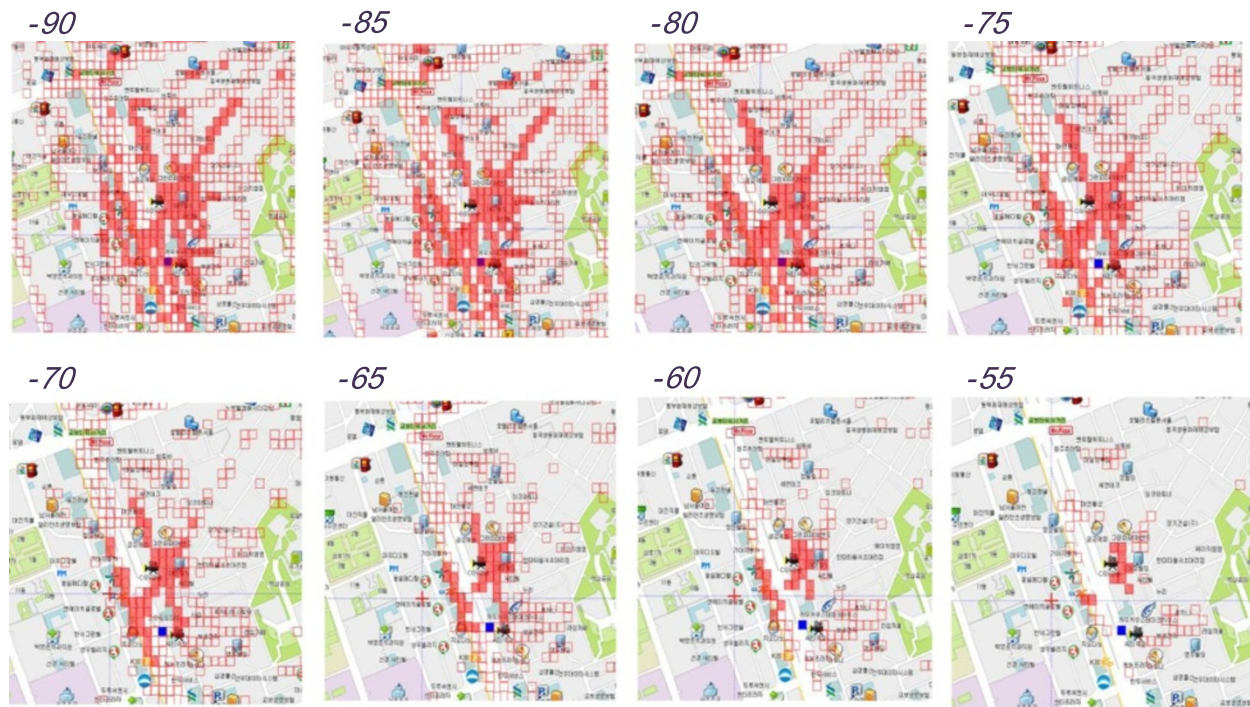


Figure 6 Coverage change of an AP.

the number of (AP identification and RSSI value) pairs, as shown in Figure 7). For the entire domestic national data collection and efficient position estimation, the filtering mechanism should be widely applied, in any form whatsoever.

Figure 7 shows the general filtering structure of the proposed method. To maintain the consistency of filtering, cut-over RSSI thresholds are applied to both the 'reference fingerprint DB' and 'fingerprint measured by handset in positioning stage'. The simultaneous application of common cutover thresholds gives consistency to the fingerprint pattern matching in the actual positioning stage. The entire

framework to determine the proper cutover threshold can be mathematically described as the following integer programming models (1), (2), (3), (4), and (5).

$$\text{Minimize } \sum_j |g_j - m_j| \quad (1)$$

$$\text{S.T. } m_j = \sum_i y_{ij} \quad (2)$$

$$y_{ij} \leq o_{ij} \quad (3)$$

$$(o_{ij} - y_{ij})r_{ij} \leq y_{ij}r_{ij} \quad (4)$$

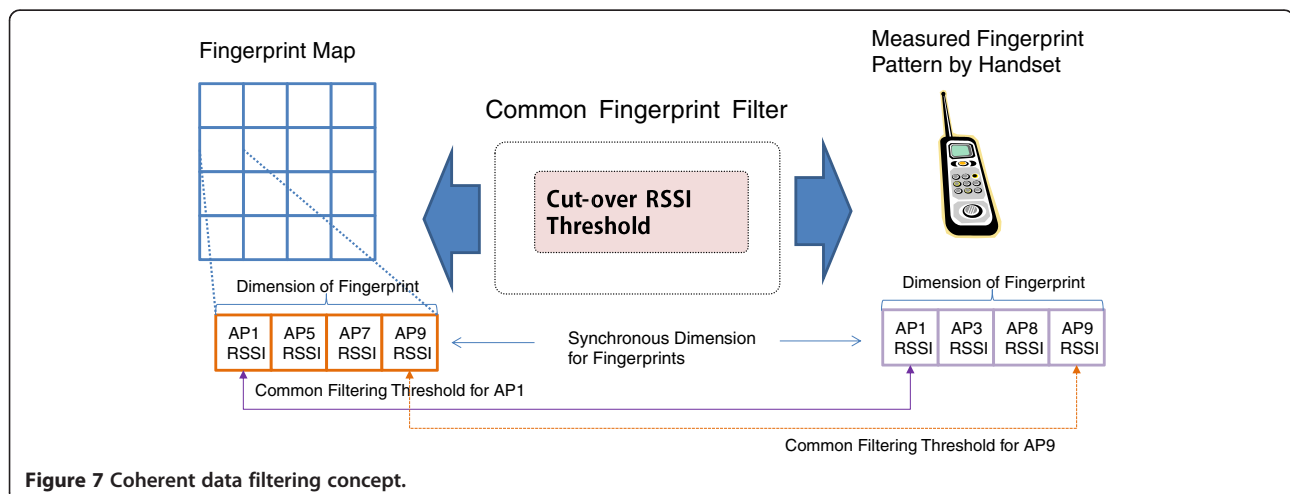


Figure 7 Coherent data filtering concept.

$$y_{ij}, o_{ij} = \{0 \text{ or } 1\} \quad (5)$$

where g_j is the desired number of APs in grid j . Note that g_j determines the numerical dimension in a position estimation algorithm. A typical position estimation algorithm uses a pattern-matching method. The synchronous and proper dimensioning of pattern matching is important for the significance and running speed of estimation. The practical value of g_j can be fixed as differentiated numbers, according to the regional characteristics, such as residential or commercial areas (i.e., smaller values for residential and larger ones for commercial). m_j denotes the number of APs offered for grid j in the reference fingerprint DB map, after filtering. Thus, the objective function (1) minimizes the difference between the desired and offered number of APs, for all grids in the reference DB map. Equation 2 represents the calculation of m_j . y_{ij} is determined as 1, if AP i in grid j survives after the filtering, otherwise as 0. The summation of y_{ij} for all i is in a grid j (i.e., $\sum y_{ij}$) can make the value of m_j . o_{ij} means the original existence of AP i in grid j . When AP i is originally detected in grid j , the value of o_{ij} is determined as 1, otherwise as 0. The inequality (3) shows the existence of AP i in grid j before and after filtering. The candidate APs for filtering should be selected among the originally detected APs by the inequality (3). The inequality (4) guarantees that all survived AP RSSIs are greater than the RSSIs of filtered APs. $(o_{ij} - y_{ij})$ has value 1 (i.e., $(o_{ij} - y_{ij}) = 1$), if AP i in grid j is filtered. Thus, $(o_{ij} - y_{ij})r_{ij}$ means RSSIs of filtered APs. $(o_{ij} - y_{ij})r_{ij}$ should be less than the $y_{ij}r_{ij}$, the RSSIs of survived APs. r_{ij} is the measured RSSI of AP i in grid j . For arithmetical consistency, we slightly modify the RSSI of AP in our model, in the form of $r_{ij} = -1/\text{RSSI}_{ij}$ (RSSI_{ij} is the actual measured value of RSSI for AP i in grid j (unit: dB)).

The proposed mathematical model determines the cutover RSSI threshold for each AP in each grid. Then, it can select optimal significant APs for each grid. The

determined RSSI threshold guarantees the minimum difference between the desired and offered number of APs. However, for coherent application of common cutover threshold to both the 'reference DB map' and 'measured fingerprint by handset', the practical cutover threshold should be given for each AP as a single value, not for each grid (a mobile handset does not know its current grid position in the positioning stage). The determined RSSI cutover threshold by the mathematical model can make multiple thresholds for a single AP. Figure 8a shows the discrepancy of RSSI thresholds for a specific AP (i.e. -85 and -88 dB for AP14) in the mathematical modeling approach.

To get a practical cutover RSSI threshold for each AP, we should determine a single value threshold for a single AP, such as -88 dB for AP14 in Figure 8a, or -85 dB for AP14. This single-value threshold determination cannot maintain the optimality of the aforementioned mathematical model, but it can give the common cutover threshold to both the reference DB map building and handset positioning. Because of the dynamic interaction between adjacent grids, as presented in Figure 8b, a single-value threshold should be determined, under the harmonization among grids. The decrement or increment of single-value threshold can make for a different offered number of APs among grids. We suggest a dynamic type of algorithm to determine a harmonized single-value threshold for each AP.

$$\arg_{\text{RSSI}^{\text{AP}i}} \min f(\text{RSSI}^{\text{AP}i}) = \left\{ \text{RSSI}^{\text{AP}i} \mid f(\text{RSSI}^{\text{AP}i}) = \left(\sum_{j(k)} |g_{j(k)} - m_{j(k)}(\text{RSSI}^{\text{AP}i})| \right) \right\} \quad (6)$$

Equation 6 shows the single-value RSSI threshold determination for AP i . $|g_{j(k)} - m_{j(k)}(\text{RSSI}^{\text{AP}i})|$ has the same

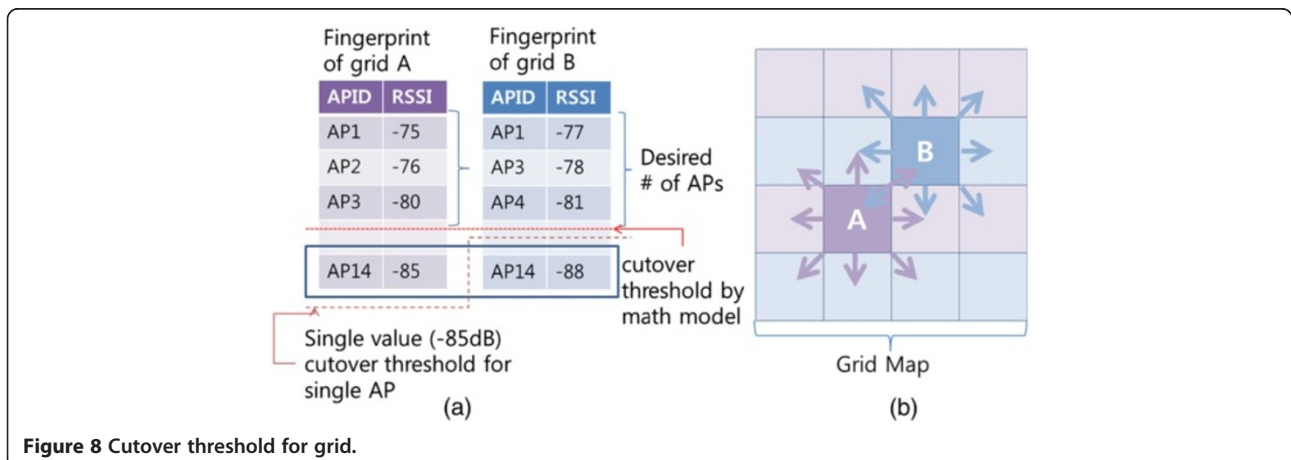


Figure 8 Cutover threshold for grid.

meaning in function (1), except that $m_{j(k)}(\text{RSSI}^{\text{AP}_i})$ denotes the offered number of APs for grid $j(k)$ under the given $\text{RSSI}^{\text{AP}_i}$ (a grid $j(k)$ is in a grid set j which has AP_i as a detected AP in its fingerprint; k is the ordering index for each grid in a set j). $\text{RSSI}^{\text{AP}_i}$ is the determined single-value RSSI threshold for AP_i . The $\text{RSSI}^{\text{AP}_i}$ guarantees the minimum difference between the desired and offered number of APs for grids j . The $\text{RSSI}^{\text{AP}_i}$ values are determined by an iterative Equation (7):

$$f_{n+1}(\text{RSSI}_{n+1}^{\text{AP}_i}) = \left| g_{j(n+1)} - m_{j(n+1)}(\text{RSSI}_{n+1}^{\text{AP}_i}) \right| + f_n^*(\text{RSSI}_n^{\text{AP}_i}) \quad (7)$$

The function f_n^* in Equation 7 denotes the summation of difference between the desired and offered number of APs from the 1st to the n th grid, under the given current RSSI threshold for AP_i , $\text{RSSI}_n^{\text{AP}_i}$. The f_{n+1} can be obtained by the summation of 'the n th function value (i.e. $f_n(\text{RSSI}_n^{\text{AP}_i})$)' and 'difference between the desired and offered number of APs for the $n+1$ th grid (i.e., $|g_{j(n+1)} - m_{j(n+1)}(\text{RSSI}_{n+1}^{\text{AP}_i})|$ '. The difference between the desired and offered number of APs for the $n+1$ th grid is calculated under the RSSI threshold of AP_i in the $n+1$ th grid. Now, we compare the $f_{n+1}(\text{RSSI}_{n+1}^{\text{AP}_i})$ and $f_{n+1}(\text{RSSI}_n^{\text{AP}_i})$. If the gap between $f_{n+1}(\text{RSSI}_{n+1}^{\text{AP}_i})$ and $f_{n+1}(\text{RSSI}_n^{\text{AP}_i})$ is lower than a pre-specified range, the current RSSI threshold for

AP_i is maintained at the same value for the n th grid, (i.e., $\text{RSSI}_{n+1}^{\text{AP}_i} = \text{RSSI}_n^{\text{AP}_i}$). Otherwise, we configure the new current RSSI threshold as $\text{RSSI}_{n+1}^{\text{AP}_i}$. Note that the RSSI threshold (i.e., $\text{RSSI}_n^{\text{AP}_i}$ for all i and n) is ideally obtained from the integer programming models (1) ~ (5). However, we have a candidate RSSI range, namely $\text{RSSI}_n^{\text{AP}_i} = \{\min\text{RSSI}_n^{\text{AP}_i}, \dots, \max\text{RSSI}_n^{\text{AP}_i}\}$, for calculation convenience. The complexity of the proposed dynamic algorithm is $O(nC)$, n is the number of neighboring grids of target grid and C is the scale factor of RSSI measure. To specify all APs in a grid (if we assume k APs per grid), we need maximum k implementations of the dynamic algorithm of $O(nC)$ complexity (i.e., total $O(nkC)$ complexity of implementations). The selection of APs in online positioning stage is totally based on the common cutover RSSI threshold of specific AP. The process of online positioning stage is relatively simple. Just applying the common cutover RSSI threshold can give the simplicity in practical application in online positioning process. Figure 9 shows the sample grid area for calculation of the proposed filtering method. A total of nine neighbor grids are selected for applying the filtering method. We assume the desired number of APs as a single value of 10, for simplicity (the determination of desired number of APs (i.e., g_j) is described in the Appendix).

By applying the filtering method, we obtain the following results (Table 2), which show the offered numbers of

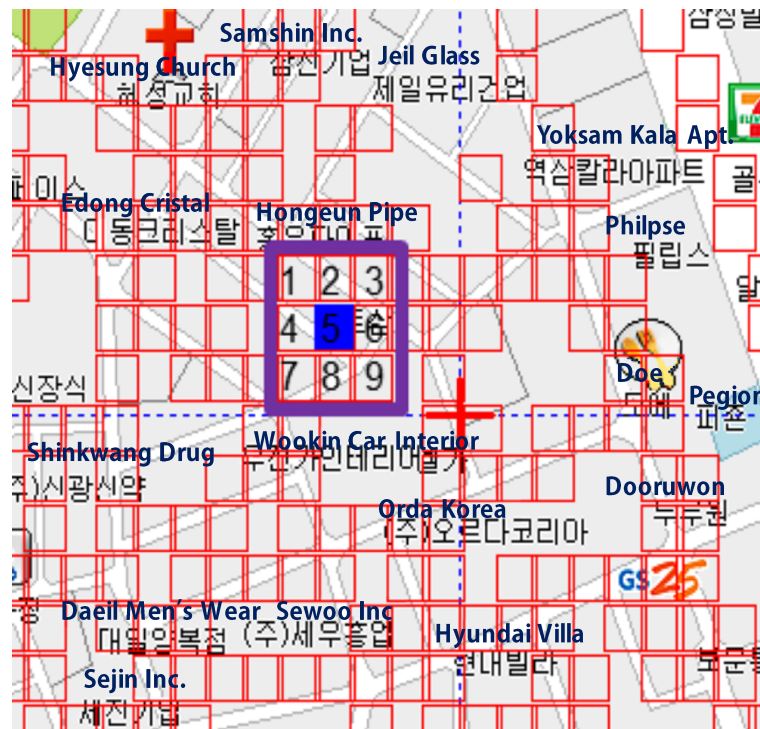


Figure 9 Sample grid set for single value cutover threshold calculation. Note: All Korean characters indicate names of places. The imported map uses Korean geographical interface.

Table 2 Results for sample filtering

Grid#	1	2	3	4	5	6	7	8	9
Before filtering ^a	25	19	18	21	13	18	7	8	20
$ g_j - m_j $	15	9	8	11	3	8	3	2	10
After filtering ^b	11	10	9	15	9	9	5	6	15
$ g_j - m_j $	1	0	1	5	1	1	5	4	5

^aAverage difference 7.67, average offered APs 16.56; ^bAverage difference 2.56, average offered APs 9.889.

APs for each grid. The average offered numbers of APs are very close to the desired number. The difference is reduced from 7.67 to 2.56.

After the filtering by single-value cutover RSSI threshold, we perform a fingerprint compensation process in the building of reference DB, based on a coherence test. The collected fingerprint data have empirical fluctuations on RSSI caused by environmental factors. The human movement makes a short-range fluctuation of RSSI (The human body is a sort of radio wave absorber. Approximately 3 dB is attenuated throughout the single human body). We can compensate imperfectly measured RSSI values by the compensated ones. Figure 10 shows the concept of compensation. First, we find the temporal fluctuation on RSSI. Because of human movement, RSSI

can be fluctuated temporally. Thus, we collect RSSI measurement data from several different time bands at each collection point. If the difference of RSSI measurements is greater than a certain level, we select vertical and horizontal neighbor grids and apply a smoothing technique. By the curve fitting (linear or exponential) with neighboring RSSI values, the two newly compensated RSSI values (i.e., vertical and horizontal) are obtained according to both the vertical and horizontal axis".

Using the filtering and compensation mechanism, the volume of fingerprint data is significantly reduced, and the data quality is highly enhanced, for precise position estimation. The common RSSI threshold for both the reference fingerprint DB and handset-measured fingerprint gives effective dimension of the fingerprint, which imparts good estimation quality, with sufficiently fast running speed for position estimation algorithms.

4. Numerical results

To show the applicability of the proposed filtering and compensation methods, we collected all the fingerprint data from the Seoul Gangnam urban district, which contains about 110,000 grids. A single scanning process usually generates approximately 600,000 fingerprint data. This is a huge amount of data, and a relatively large area is not

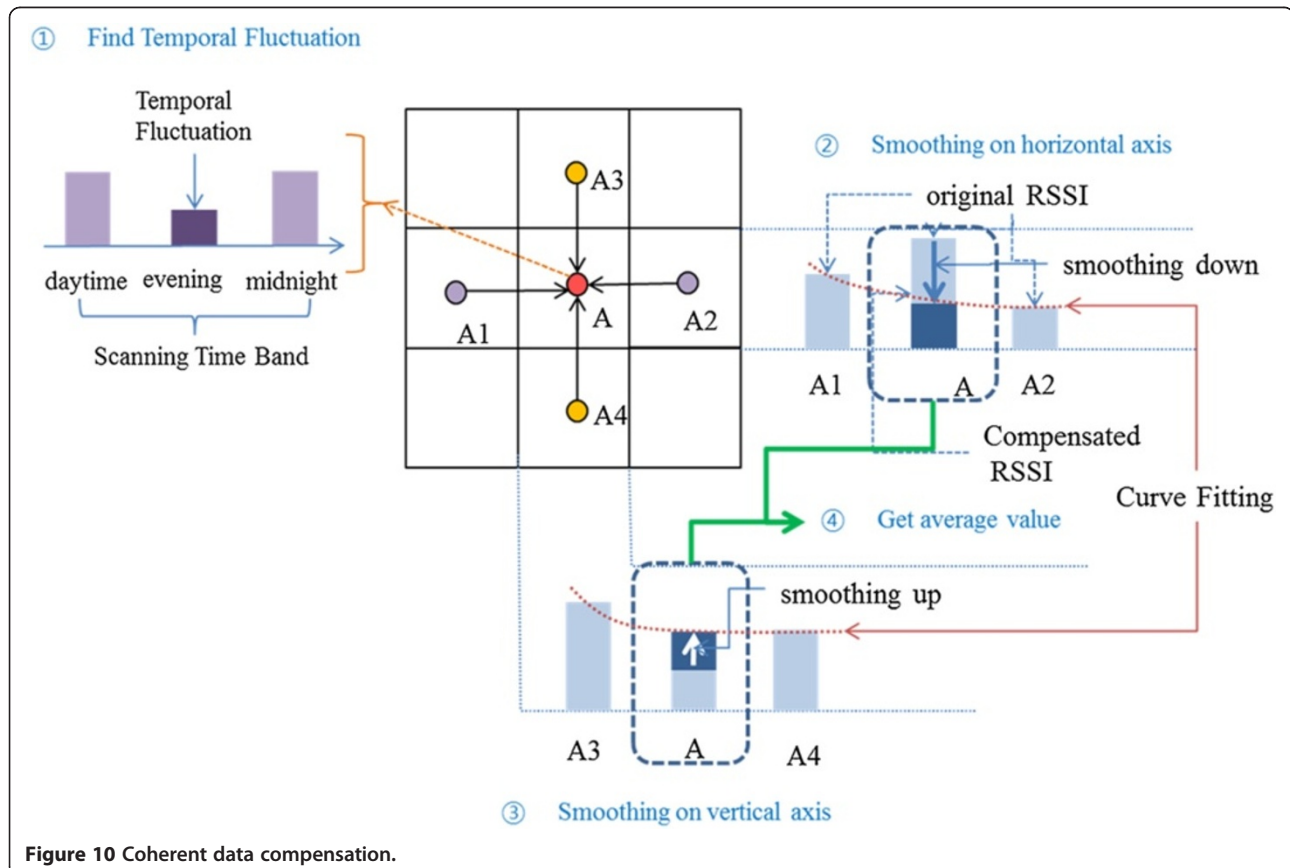


Figure 10 Coherent data compensation.

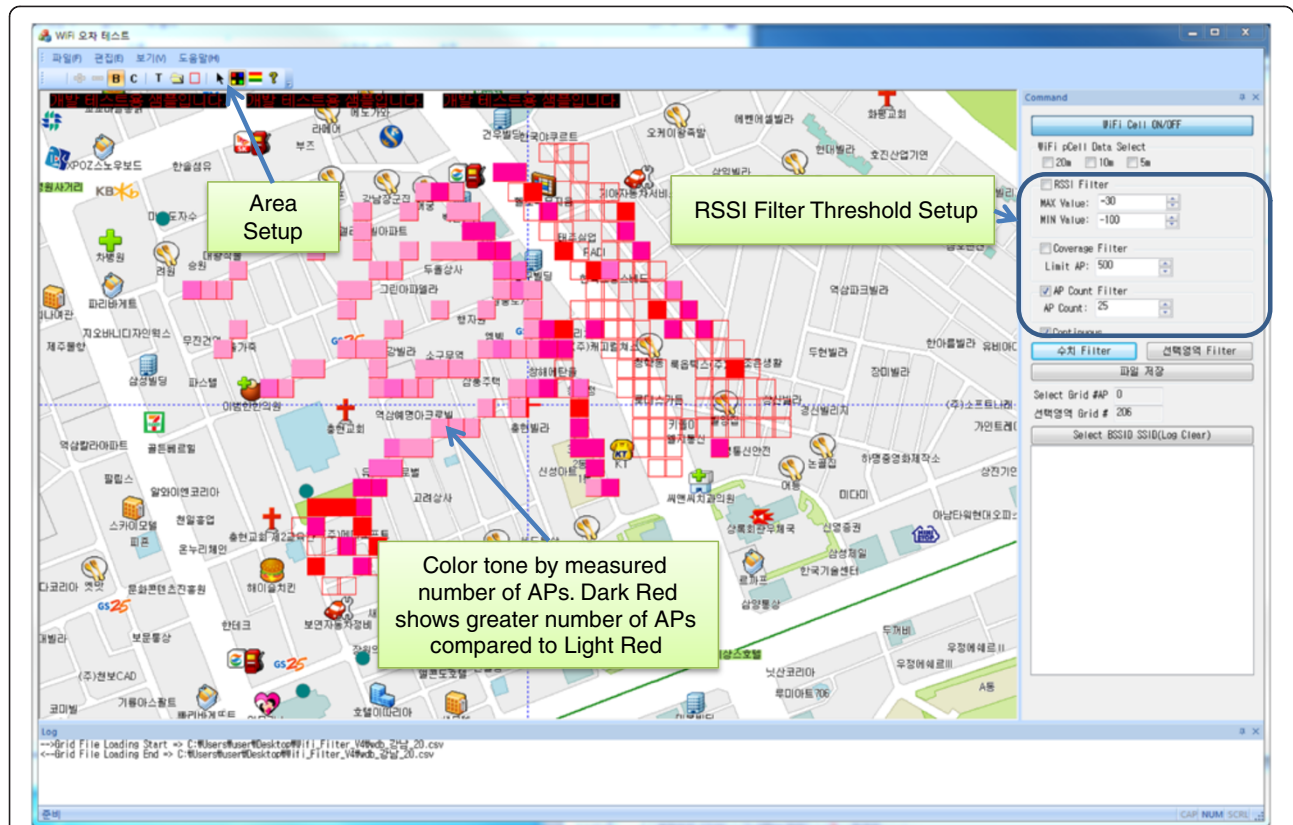


Figure 11 Snapshot of analyzing tool. Note: All Korean characters indicate names of places. The imported map uses Korean geographical interface.

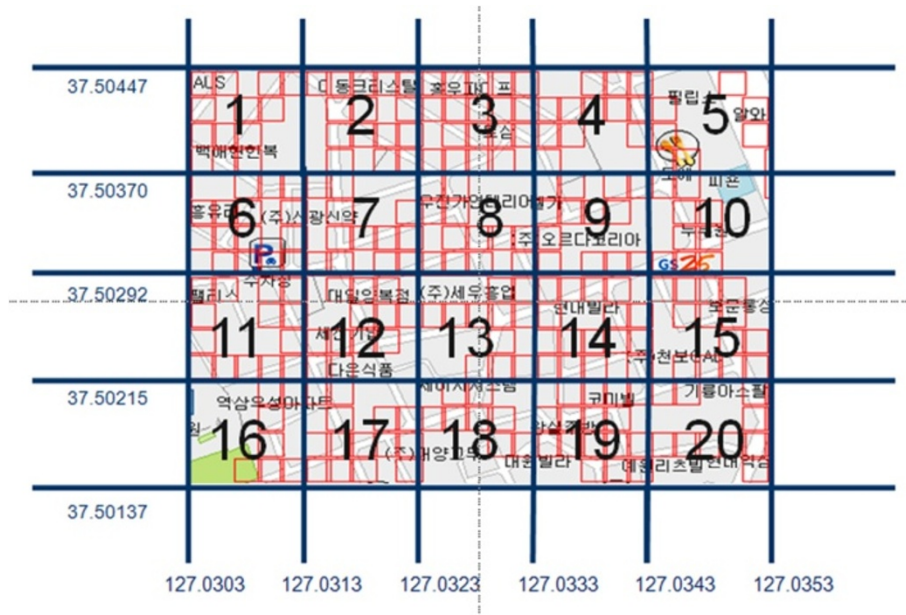


Figure 12 Test district.

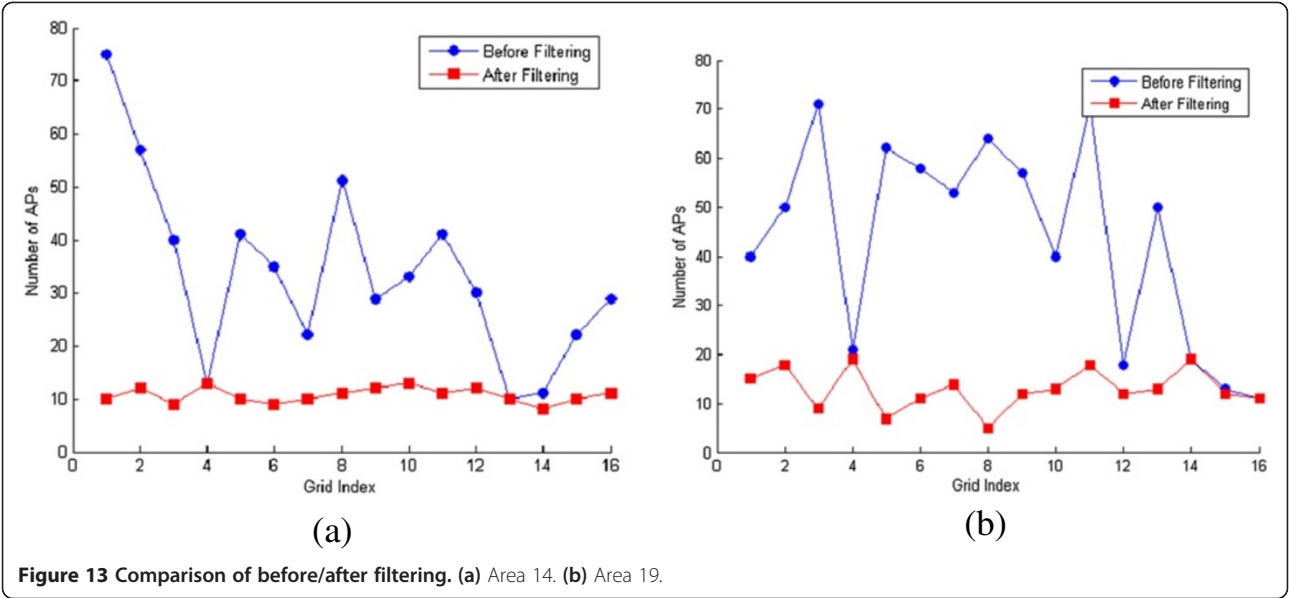


Table 3 Single-value thresholds for APs

AP index	Cutover threshold (dB)	AP index	Cutover threshold (dB)	AP index	Cutover threshold (dB)
952	−95	976	−79	1,000	−60
953	−69	977	−76	1,001	−85
954	−40	978	−73	1,002	−95
955	−95	979	−81	1,003	−95
956	−84	980	−79	1,004	−85
957	−79	981	−84	1,005	−77
958	−85	982	−81	1,006	−81
959	−76	983	−77	1,007	−76
960	−78	984	−75	1,008	−75
961	−81	985	−78	1,009	−83
962	−67	986	−54	1,010	−95
963	−95	987	−40	1,011	−95
964	−84	988	−95	1,012	−95
965	−95	989	−79	1,013	−95
966	−79	990	−79	1,014	−76
967	−61	991	−81	1,015	−55
968	−61	992	−95	1,016	−95
969	−85	993	−95	1,017	−76
970	−64	994	−61	1,018	−87
971	−80	995	−84	1,019	−76
972	−79	996	−83	1,020	−76
973	−79	997	−83	1,021	−85
974	−81	998	−79	1,022	−51
975	−95	999	−95	1,023	−42

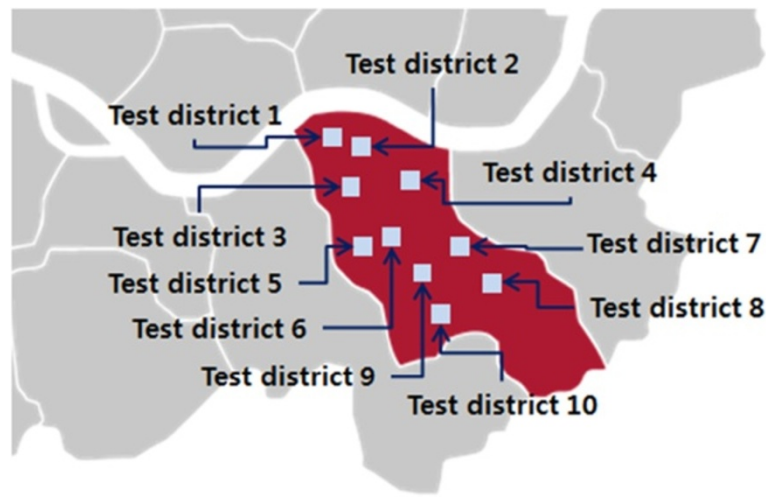


Figure 14 Test districts in Gangnam, Seoul.

suitable for testing and enhancing the details of the filtering method. Thus, we constructed a Windows-based performance analysis tool, as shown in Figure 11. The analysis tool shows the density of APs by different coloring (Light red means relatively light density of APs; dark red means relatively high AP density.). We applied the proposed filtering method and evaluated its performance in a relatively restricted area as the first step.

The test area shown in Figure 12 is a square district (320 m \times 500 m) in Gangnam, Seoul. This district is classified as a commercial area in Seoul. It includes many commercial buildings and dense foot traffic. There are 248 grids. A total of 1,267 APs are detected. Each grid has one fingerprint, which has 26.6 APs and their RSSI measurement value, on average. A total of 6,566 (AP identification

and RSSI value) pairs are used to build the reference fingerprint DB.

The proposed filtering method greatly reduces the volume of data: from 6,566 measurement pairs (26.6 APs per grid on average), to 2,751 (10.9 APs per grid on average). Approximately 60% of the measurement data are filtered out by the proposed method. Figure 13 shows the difference of the originally measured APs and offered APs by the filtering method, for areas 14 and 19 of Figure 13. The offered numbers of APs are in the range of the desired number of APs per grid (we set the desired number of APs to 10 for the test).

Table 3 shows the single-value RSSI thresholds for APs. We list a part of the whole list of thresholds for reference.

Table 4 Position estimation error for test districts

Test district	Commercial/ residential	Average error without filtering (m)	Average error with unified RSSI filter (m)	Average error with coherent filtering and compensation (m)	
				Coherent filtering only	Coherent filtering and compensation
1	65:35	64	63	52	49
2	25:75	49	47	42	41
3	60:40	72	72	53	53
4	30:70	56	55	43	42
5	70:30	81	79	57	55
6	40:60	47	46	42	41
7	55:45	59	58	44	41
8	70:30	74	71	54	51
9	75:25	76	73	61	58
10	50:50	62	61	50	49
Average	-	64.0	62.5	49.8	48
Enhancement rate	-	-	2%	22%	25%

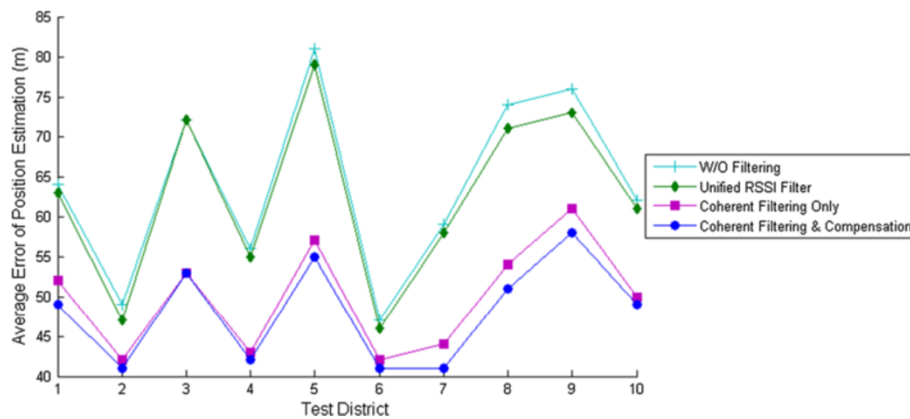


Figure 15 Error difference between before/after filtering.

Next, we extended coherent filtering and compensation in a large area. Ten test districts (see Figure 14) in Seoul Gangnam were selected to prove the applicability of the proposed filtering method. The area of Gangnam district is 39.55 km². The range of area for test districts is 0.10 ~ 0.17 km². Each district has 50 ~ 60 target points for position estimation.

We apply the well-known and widely used the KNN pattern matching algorithm [7]. The results prove the effectiveness of the proposed method, in various diversified environments of an urban area. For comparison purposes, Table 4 includes the results of 'W/O filtering' and the 'unified RSSI filter'. In the unified RSSI filter, we apply the same RSSI cutover threshold for all APs, of -85 dBm.

The results show the effect of the proposed method. An average of 22% enhancement (for coherent filtering only) and 25% enhancement (for coherent filtering and compensation) are measured for 10 different districts. Most of the enhancement depends on coherent filtering. The contribution of compensation is somewhat limited. We also attach

the ratio of land utilization (commercial/residential) for test districts. Each test district has its differentiated ratio for commercial and residential areas. The enhancement is diversified with the range (5 ~ 24 m). We can observe the relatively higher enhancement for commercial-oriented districts. The commercial-oriented districts have approximately 18-m (range 12 ~ 24 m) enhancement, whereas approximately 8-m (range 5 ~ 13 m) enhancement for residential-oriented districts. Figure 15 directly shows the effectiveness of the proposed method by a graphic chart.

Note that the majority of test points belong to the outdoor environment. The automatic scanning vehicle has an access problem to the indoor environment. The most of indoor fingerprints are collected by human power. Thus, our experiment has a limitation for the applicability on indoor environment. However, our proposed framework are applicable both on indoor and outdoor environments. The radio signal fluctuation and structure complexity are more serious in an indoor environment. The proposed coherent filtering framework has relative advantage on the complex

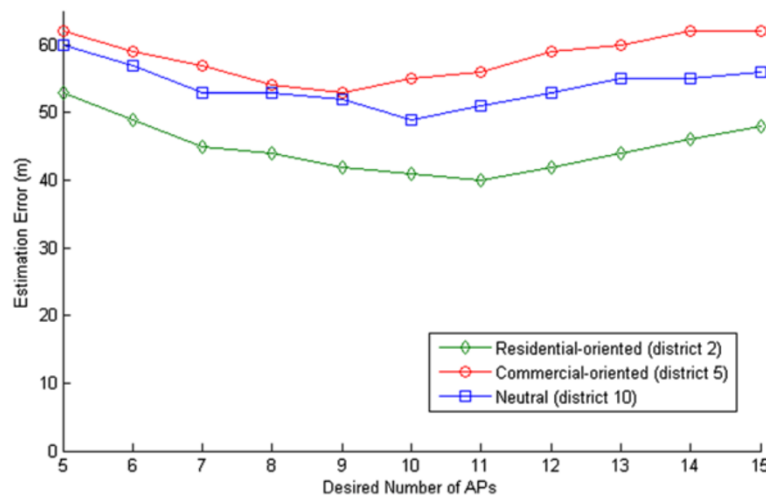


Figure 16 Determination of desired number of APs.

and fluctuated environment (see the comparison between commercial and residential). We carefully expect the effective application to the indoor environment.

5. Conclusion

The rapid growth of mobile communication and the proliferation of smart phones 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, in estimating an AP's exact position. Moreover, significant technical advances are not shared publicly by solution providers. RF fingerprint WPS is an alternative valuable way to penetrate the positioning solution provider market. Even by indiscriminate fingerprint collection, providers can build a fingerprint DB and apply a simple pattern-matching algorithm for position estimation. However, to build a competitive fingerprint WPS solution, we should focus on fingerprint data management, and precise estimation algorithms. The essential factor of radio fingerprint map is the data integrity of RSSI. Because of millions of APs in the urban area, RSSI measurement data are seriously contaminated. Therefore, we present a coherent filtering method for RSSI measurement data. In our method, we built a new fingerprint filtering method. Based on the single cutover threshold and data coherency, collected fingerprints are filtered and compensated. A new fingerprint data filtering for position estimation can strengthen the advantages of RF fingerprint WPS. Compared to the existing approaches for fingerprint filtering, our method achieves a better performance, in both average error of estimation, and deviation of errors. Furthermore, all the fingerprint data were harvested from the actual measurement of RF fingerprints in Seoul's Gangnam district. We built an effectively filtered fingerprint DB for the entire area of Seoul and applied position estimation. These trials show the practical usefulness of the proposed methodology.

Appendix

The higher number of g_j provides more information for precise position estimation. On the other hand, it also generates larger deviation (i.e., the deviation is gradually increased according to increment of g_j). The following statistical implication shows the negative effect (i.e., large deviation on position estimation) by the increment of g_j .

The similarity of two patterns is usually determined by the Euclidian distance of two different patterns. That is, the statistical difference between two fingerprints is based on the square of the Euclidean distance ($d^2(i, j)$) between two fingerprint pairs (f_i, f_j) as $d^2(i, j) = (f_i - f_j)^2$, where $f_i = \{RSSI_{AP1}^i, RSSI_{AP2}^i, \dots, RSSI_{APn}^i\}$, $f_j = \{RSSI_{AP1}^j, RSSI_{AP2}^j, \dots, RSSI_{APn}^j\}$. Each value of $RSSI_{APk}^i$ (RSSI for APk 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 (i.e., $d^2(i, j) \sim \chi^2(n)$). Generally, the $\chi^2(n)$ has a mean n and variance $2n$. The value of n is determined by g_j , then, deviation of position error is increased by the increment of g_j (i.e., if n is increased, variance ($2n$) is also increased).

Now, we have to find the tradeoff relation between 'deviation caused by higher g_j ' and 'information density obtained by higher g_j '. We adopt an empirical experiment to find the proper number of g_j . Figure 16 shows three representative illustrations for position estimation error.

This figure shows the position estimation error for three test districts: (1) residential-oriented district (district 2 in Table 4), (2) commercial-oriented district (district 5 in Table 4), and (3) neutral district (district 10 in Table 4). Despite of small difference, significant range of g_j can be found; 9 ~ 11 APs are sufficient numbers for position estimation. From the knowledge of empirical experiments, we can apply proper g_j to practical reference DB management. Note that this empirical experiment cannot give an absolute value of g_j for all cases. The repetitive experiments are required to guarantee the effectiveness of the empirical method.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

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 (2013-025572).

Author details

¹Department of Industrial Engineering, Ajou University, Suwon, Republic of Korea. ²Department of Information and Communications Engineering, Sejong University, Seoul, Republic of Korea.

Received: 7 June 2013 Accepted: 12 January 2014

Published: 21 January 2014

References

1. Y Masumoto, *Global positioning system*, US Patent 5,210,540, 11, May 1993
2. JM Watters, *Combining GPS with TOA/TDOA of cellular signals to locate terminal*, US Patent 5982324, 14, May 1998
3. S Wireless, *Estimation of positioning using WLAN access point radio propagation characteristics in a WLAN Positioning System*, US Patent 7515578, World Intellectual Property Organization, 22, Nov 2007
4. J del Prado Pavon, *Link adaptation strategy for IEEE 802.11 WLAN via received signal strength measurement*, in *Wireless Communication and Networking Conference*, (New Orleans, 16–20 March 2003)
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>. Accessed 16 January 2008

6. S Wireless, *Location beacon database and server, method of building location beacon database, and location based service using same*, US 62310804, World Intellectual Property Organization, 29 Oct 2004
7. B Li, J Salter, AG Dempster, C Rizo, Indoor positioning techniques based on Wireless LAN, in *The First IEEE International Conference on Wireless Broadband and Ultra Wideband Communications*, (Sydney, Australia, 13–16 March 2006)
8. B Li, JJ Quader, AG Dempster, On outdoor positioning with Wi-Fi. *Journal of GPS* **7**, 1 (2008)
9. S Das, T Teixeira, SF Hasan, Research issues related to Trilateration and fingerprinting methods. *Int J Res Wireless Systems* **1**, 33–35 (2012)
10. D Pandya, R Jain, E Lupu, Indoor location estimation using multiple wireless technologies, in *Proceedings of the 15th IEEE International Symposium on Personal Indoor and Mobile Radio Communication (PIMRC)*, (Beijing, China, 7–1 Sept 2003)
11. R Battiti, TL Nha, A Villani, *Location-aware computing: a neural network model for determining location in wireless LANs*, Technical Report DIT-02-0083 (University of Trento, 2002). <http://eprints.biblio.unitn.it/233/>
12. A Haeberlen, E Flannery, AM Ladd, A Rudys, DS Wallach, LE Kavraki, Practical robust localization over large-scale 802.11 wireless networks, in *Proceedings of the 10th annual international conference on Mobile computing and networking (MobiCom'04)*, (Philadelphia, 26 Sept - 1 Oct 2004)
13. K Kaemarungsi, *Distribution of WLAN received signal strength indication for indoor location determination*, 2006 1st International Symposium on Wireless Pervasive Computing, Phuket, Thailand, 16–18, Jan 2006
14. G Welch, G Bishop, *An Introduction to the Kalman Filter* (University of North Carolina, Technical Report, 2006)
15. H Yoshida, S Ito, N Kawaguchi, Evaluation for pre-acquisition methods for position estimation system using Wireless LAN, in *Proceedings of the Third International Conference on Mobile Computing and Ubiquitous Networking (ICMU 2006)*, (London, UK, 11–13 Oct 2006), pp. 148–155

doi:10.1186/1687-1499-2014-13

Cite this article as: Kim and Yeo: A coherent data filtering method for large scale RF fingerprint Wi-Fi Positioning Systems. *EURASIP Journal on Wireless Communications and Networking* 2014 **2014**:13.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com