FIMO: A Novel WiFi Localization Method

Yao Zhou, Leilei Jin, Cheqing Jin*, and Aoying Zhou

Shanghai Key Laboratory of Trust Worthy Computing,
Software Engineering Institute
East China Normal University, Shanghai 200062, China
sei_zy2006@126.com, lljin_ecnu@sina.com, {cqjin,ayzhou}@sei.ecnu.edu.cn

Abstract. With the development of technology and the proliferation of mobile computing devices, people's need for pervasive computing is rapidly growing. As a critical part of pervasive computing, Location Based Service (LBS) has drawn more and more attention. Localization techniques that report the real-time position of a moving object are key in this area. So far, the outdoor localization technologies (i.e, GPS) are relatively mature, while the indoor localization technologies are still under improvement. In this paper, we propose a novel WiFi localization method, called FIMO (FInd Me Out). In this method, we take the instability of signal strength and the movement of objects into consideration when determining the location based on Fingerprint. Experimental results show the proposed method is capable of estimating a moving object's location precisely.

Keywords: LBS, localization, indoor, WiFi.

1 Introduction

For a long time, people expect the electronic devices around them can have perception towards the environment.In early 2012, IDC (Internet Data Center) predicted that around 1.1 billion smart connected devices would be sold in 2012, and this number would even be doubled in 2016[1]. As one of the important applications of ubiquitous computing system, localization application has drawn more and more attention. In general, these systems often need to provide information or services according to user's current location. For example, a customer who is shopping in a new mall may need to know where his /her favorite shop is and how to reach it. Advertisements can be recommended to users based on their locations etc..

Nowadays, outdoor localization technologies, such as GPS, have become quite popular in daily life. Recent years also witness the increasing demand for indoor localization. Several indoor localization techniques have been developed in recent years, including Bluetooth [3], RFID [8] and WiFi [1]. Bluetooth and WiFi based approaches rely on Received Signal Strength (RSS) for localization, while RFID based approaches determine the location of the moving object by reading the

---

* Corresponding author.

[1] See http://www.199it.com/archives/29224.html

active RFID tag. WiFi localization has many advantages, such as ubiquitous coverage, scalability, no additional hardware required, extended range, no line of sight restrictions and free measurement etc.

We can see that WiFi is a good choice for localization in indoor space. However, the accuracy of basic approaches is low due to the instability of WiFi signal strength. Some projects combined WiFi with other devices such as RFID [5] [13], Bluetooth [2] [12] to improve the accuracy. Although the accuracy is improved, additional devices are required, and will cause extra burden to the setup process. Besides, very few approaches have taken advantage of continuous monitoring of users' locations and considered the features of WiFi signal strength itself.

In this paper, we propose a novel approach to estimate a moving object's location, by taking into consideration the instability of signal strength and the movement of objects. This method is based on the Fingerprint approach in estimating the location of a moving object.

The remainder of this paper is organized as follows. In Section 2, we survey related work in localization technologies. In Section 3, we discuss our research methodology. In Section 4, we present and analyze our approach. Section 5 reports a comprehensive experimental study. Finally, we present our conclusions and provide some future research directions in Section 6.

2   Related Work

The existing work on indoor localization can be divided into two main groups: signal propagation approaches and fingerprinting approaches.

2.1   Signal Propagation (SP) Approaches

Signal propagation approaches tend to as exponential attenuation models for WiFi signals and use the path loss[2] to determine location based upon distance from the Access Points (APs), whose locations are assumed known [10]. Although the signal propagation models may work well in open space, they cannot be used directly in indoor environment since there exists many kinds of obstructions. In other words, we need to consider additional path loss caused by the physical obstructions between the APs in indoor localization and make the models much more complex than in the open area. The model proposed by Seidel and Rappaport [11] is the most popular one till now. In this model, the walls made of different materials are assigned different attenuation factors for better precision. SP models can therefore be generalized to locations without available reference data. Furthermore, it only requires storage for the location of each AP and a simple description of its signal attenuation. However, since WiFi signals are unstable, attenuation is almost never radially symmetric. In addition, other complex factors, such as multipath, different construction materials, and various objects in the building, ever decrease the precision of the above methods.

As a conclusion, signal propagation approaches are easy to implement, but the performance is low.

---

[2] See http://en.wikipedia.org/wiki/Path_loss

2.2 Fingerprinting (FP) Approaches

FP approaches ignore attenuation and instead compute likelihoods from location-specific statistics compiled from reference data. FP needs to create a signal strength map in off-line manner in order to start localization. When a user submits a location request, his current signal strength values are compared with those in the database. The system then compute a most probable location for the user. Although FP techniques require reference data to setup, reference data is only needed to be collected once in a new building and the accuracy of FP is better than SP.

FP approach was developed by Microsoft Research RADAR [1]. In their paper, they proposed two ways to match signal strength values. One is to return the position with the nearest neighbor in signal space; the other is to return the average of the coordinates of k neighbors in signal space as an estimate. In general, it was found that WiFi signal strength at the same location changes with time, temperature and objects around it. Thus, it is hard to achieve good accuracy based on pure matching of signal strength.

Most existing methods estimate the user's location only by using of the current signal strengths. Ho et. al. proposed a method to retrain the system when the user finds that the estimation location is wrong [7]. N. Hernandez et. al. proposed a WiFi localization system based on fuzzy techniques. It is only suitable for the small-scale variations which happen when the user moves over a small distance (in the range of wavelength) [6]. Baniukevic et. al. proposed an indoor localization algorithm by combining WiFi with Bluebooth [2]. It uses WiFi signal strength to localization and adjusts the coordinates of the location using Bluetooth at reference points. The methods proposed by D. L. Lee and Q. X. Chen have considered the movement of objects, but these methods can only identify the room where the user is in [4] [9].

Our algorithm is a type of Fingerprint approach. Existing approaches determine the location according to signal strength received directly. However, the signal strength is unstable in the indoor environments. Meanwhile, few of them consider the continuous movement of the user. Our approach records the signal strength received at the mobile user continuously, eliminates the locations that are unlikely to be reachable from the previous location by considering the walk speed and then determines the user's location in the possible locations by matching the signal strength with weights (which are got by considering the instability of signal strength. To be discussed in Section 4.2).

3 Research Methodology

In this section, we describe the experimental testbed and data collection.

3.1 Experimental Testbed

Our experimental testbed is located at the ground floor of a 3-storey building. Figure 1 shows the layout of the floor (765.43 m^2). Five Access Points (APs) are
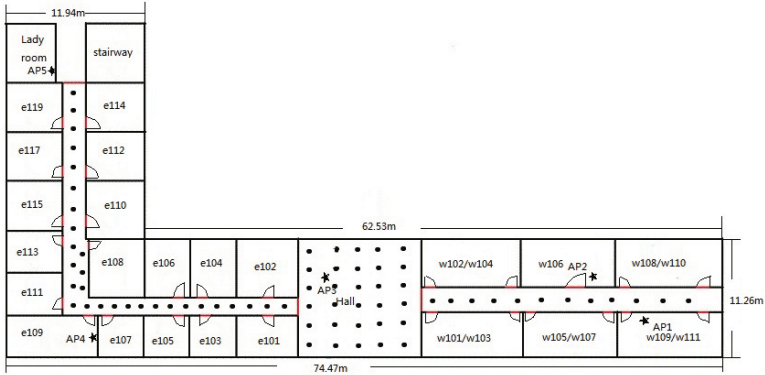
Fig. 1. Map of the floor where the experiments were conducted. The stars show the locations of the 5 access points. The black dots denote locations were empirical signal information was collected.

deployed at different locations, denoted as AP_1, AP_2, AP_3, AP_4 and AP_5. Each AP is a Totolink N300R Router. In our tests, a mobile host is a laptop computer equipped with Intel WiFi Link 5100 AGN running Windows 7. The APs provide overlapping coverage over the entire floor together.

3.2   Data Collection

Data collection phase is critical in this research methodology. We use Native WiFi API[3] to record the signal strength values. The reference data set and the testing data sets are created as follows:

- *reference data set*: We select 71 reference points, as shown in Figure 1. At each reference point, we repeatedly recorded the signal strength for 40 times, and used the average as the signal strength of this point.
- *testing data set 1*: This data set contains the signal strength values received by the user when he was walking *uniformly* from the location near AP_5 to the location near AP_1.
- *testing data set 2*: This data set contains the signal strength values received by the user when he was walking *with variable velocity* from the location near AP_5 to the location near AP_1.

In the first data set, each record is described as $(x, y, \overrightarrow{SS})$, where x and y represent the location, and $\overrightarrow{SS}$ is an array of signal strengths for five APs. In the last two sets, the form of the each record tuple is $(t, x, y, \overrightarrow{SS})$, where t represents the timestamp to record the signal strength, and the definition of the rest attributes are as same as the above definition.

---

[3] The Native Wifi API is supported by Windows platform. It automatically sets component configures, connects to, and disconnects from wireless networks. Native Wifi can store profiles on the networks it interacts with in the form of XML documents.

4   Our Algorithm

In this section, we describe our algorithm in detail. Ideally, we expect the signal strength of an AP remains almost stable with time going on. Figure 2 shows the signal strength levels at different distances. We can find that almost all of the signal strength values received 1m away from the AP are the same, while the stand deviation of the signal strength values received 10m away from the AP is around 3%. When the distance reaches 35m, the stand deviation can be around 6%. From this, we conclude that the signal strength keeps fluctuating all the time and the ones with higher value is more stable. We also test the stability of signal strengths received by a moving object, as shown in Figure 2. In this figure, a user walks from AP5 to AP1 slowly, we can observe that the signal strengths would have greater fluctuation when the user is moving. How to use these unstable signal strengths to localization is challenging to us.
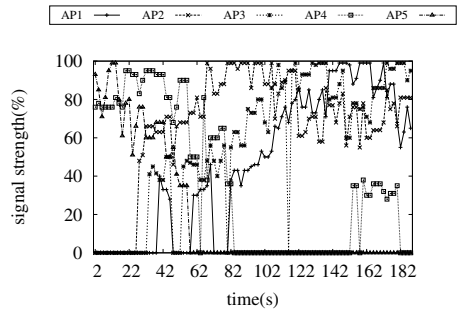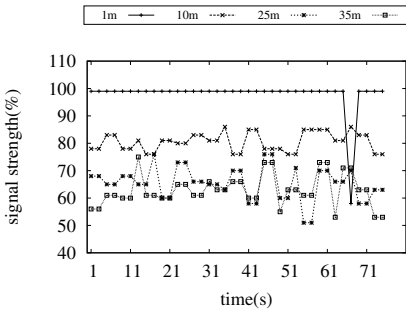
Fig. 2. Signal Level at different distances
Fig. 3. The signal strength values received the user along work

Figure 2 and 3 show that the signal strength would reflect distance more accurately if the fluctuation of the signal strength could be smoothed. Besides, in Figure 2, it's easy to see that stronger signal strength are more stable. And the degree of fluctuation of the signal strength changes with time. So we can be convinced that the signal strength with different values and the signal strength received at different time should have different credibility.

Due to these particular features, we propose an approach, called FIMO (FInd Me Out), for indoor localization. FIMO continuously records signal strength values from each AP at the user's location at fixed intervals when the user is walking around. When the user submits a location request, FIMO will process the data and then determine the location of the user using Algorithm 1. FIMO backtracks the signal strengths received by the user which are recorded in the database and put them into *ssArray* (Line 2). It smoothes the signal strength values in *ssArray* chronologically and puts the Smoothed Values into *smoothArray* (Line 3), as discussed in Section 4.1. After calculating the weight of

every signal strength value using the Original Value (The signal strength values that haven't been smoothed) (Line 5), as discussed in Section 4.2, FIMO returns the location to the user by matching the reference points (Line 7-14, Section 4.4) in the possible range (Line 6, Section 4.3).

Algorithm 1. FInd Me Out (FIMO)

Require: l: threshold of the records tracked back; n: the number of APs; $userID$: the ID of the moving object;

Ensure: location of $userID$: loc

1: Initialize $isFirst \leftarrow true$ and $minSim \leftarrow 0$;
2: Set $ssArray$ as the latest l records received by $userID$;
3: $smoothArray \leftarrow SMOOTH(ssArray)$;
4: // smooth the signal strength values in chronological order;
5: Set $w_1...w_n$ based on $SSArray$;
6: Set $rpArray$ as reference points which locate at the possible range;
7: **for** each $RP_c \in rpArray$ **do**
8:    $sim_c \leftarrow SIM(\overrightarrow{w}, smoothArray.get[0].\overrightarrow{SS}, RP_c.\overrightarrow{SS})$;
9:    // compute the similarity value between the signal strengths received right now and reference points;
10:   **if** $(isFirst == true || sim_c < minSim)$ **then**
11:      $minSim \leftarrow sim_c$;
12:      $loc \leftarrow RP_c$;
13:   **end if**
14: **end for**
15: **return** loc;

4.1 Smoothing the Signal Strength

Since the signal strength is unstable, we smooth the signal strength values before matching them. As shown in Figure 3, all of values have different degree of fluctuation. We need to smooth the fluctuation while keeping the trend of signal strength. A moving average[4] is commonly used with time series data to smooth short-term fluctuations and highlight longer-term trends or cycles. So we calculate the moving averages of all values to smooth their fluctuations. Figure 4(a) shows that this method is effective.

However, there is a special situation when the fluctuation is not continuous, just as shown in Figure 4(b) . If we just calculate the moving averages, the smoothed values may be inaccurate. Under the situation, we detect the outliers and deal with them through the following process before calculating the moving averages of all values.

1. **Detecting the outliers:** The system scans the signal strength values according to the time sequence. Assume that θ represents the threshold of normal value, and the system acquires four sequential values, a, b, c and d. If $(b - a > \theta$ and $b - c > \theta$ and $c - d \leq \theta)$ or $(a - b > \theta$ and $c - b > \theta$

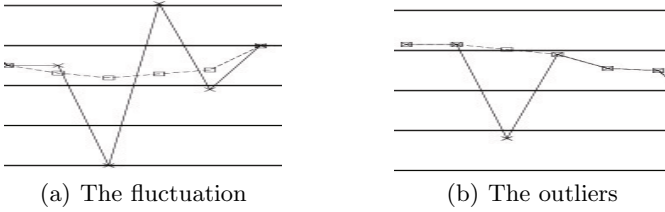[4] See http://en.wikipedia.org/wiki/Moving_average

(a) The fluctuation   (b) The outliers

Fig. 4. Smoothing the signal strength

and $d - c \leq \theta$), b will be considered as an outlier. All of a, b, c and d are measured in percentage. As the threshold for the outlier, θ is set to be 20% in our experiments. The system will regard the value as an outlier (as shown in Figure 4(b)) and proceed to step 2.

2. **Dealing with the outliers.** The system replaces the outlier b with the average of a and c. Figure 4(b) shows the result of smoothing.
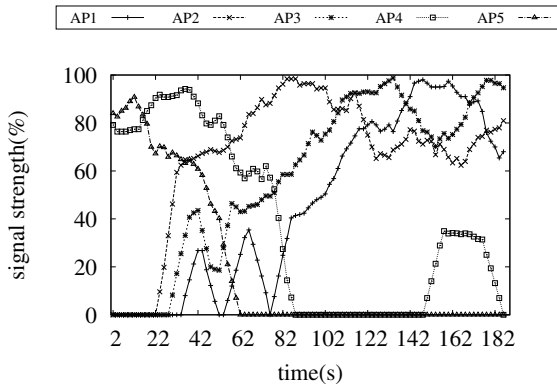

Fig. 5. The signal strength values received by the user along work after smoothing

Figure 5 shows the signal strengths after smoothing. Compared with Figure 3 which is before smoothing, it seems that the quality of the data has improved obviously after smoothing. We call the values after smoothing the Smoothed Value and the values having not been smoothed the Original Value.

4.2 Calculating the Weights

Figure 2 shows stronger signal strength tends to be more stable. To a certain extend, stronger signal strength has higher credibility. There are a lot of moving objects in indoor spaces, and the number and behaviors of them are constantly changing. These will cause the fluctuation of signal strength. Significant fluctuation will have material impact upon the credibility of the signal strength.

Therefore, even at the same location, different signal strength has different credibility, and we should give different weights to different signal strength when matching the signal strength in the database. Considering the two factors affecting the credibility of signal strength mentioned above, we propose a method to calculate the weights. The rule is to give the signal strength with the value which is higher and more stable higher weight. We can use the Original Value (o_i) of AP_i's signal strength to represent the credibility of AP_i's signal strength.

Then it comes to the question of how to measure the stability of signal strength. It is generally known that the sample standard deviation shows how much variation from the average (or expected value). A low sample standard deviation indicates that the data points tend to be very close to the mean, *vice verse*. However, the average value is not the expected value because the signal strength varies with distance, so we cannot simply use the sample standard deviation here. Although the trend of the signal strength should be described as curve function, we just focus on the stability of signal strength in a short time (such short time is deemed as a time window), and then it can be seen as linear. We regard the time period containing four continuous signal strength values as a time window in our experiments, and the length of the time window can be changed. If the length is too long, the items could not be fit with the linear function. And if it is too short, the linear function fitting out could not reasonably reflect the trend of the changes of signal strength. The method of least squares[5] is a standard approach to the approximate solution of overdetermined systems. Least Squares means that the overall solution minimizes the sum of the squares of the errors made in the results of every single equation. So we use Least Squares to find out the linear function which could describe the distribution of signal strength in a short time. Then we can get the expected signal strength value ($e_{i,j}$) received from AP_i at time t_j with the linear function.

Using the expected signal strength values, the sample standard deviation of several signal strength values (σ_i) in the same period can be calculated with Equation (1) below.

$$\sigma_i = \sqrt{\frac{1}{W-1}\sum_{j=1}^{W}(o_{i,j} - e_{i,j})^2} \tag{1}$$

where W indicates the length of the time window, and $o_{i,j}$ is the Original Value of AP_i at time t_j.

To sum up, the weight of every signal strength value of AP_i would be w_i coming from Equation (2).

$$w_i = \frac{o_i}{\sigma_i} \cdot \frac{1}{\max_k(o_k/\sigma_k)} \tag{2}$$

---

[5] See http://en.wikipedia.org/wiki/Least_squares

4.3   Considering the Walk Speed

The points on the circle centered at AP would have the same signal strength in free space. And the user might receive the same signal strength values at different locations. Therefore, the user's position might be located at a location which is far away from the actual one. In order to avoid this situation, we should consider the behavior of the user. In the indoor space, walking is the most common way of moving and the speed will has an upper limit. Thus, we could use the speed of the user and the last location of the user recorded to limit the possible candidate locations.

Specific studies have found pedestrian walking speed ranging from 1.25 m/s to 1.32 m/s for older individuals to 1.48 m/s to 1.51 m/s for younger individuals[6]. The highest normal pedestrian walk speed is about 1.5 m/s. Of course, with speed testing equipment, the accuracy of the speed can be improved. Here we use 1.5 m/s as the pedestrian walk speed to limit the candidates. We take the last location of the user as the center of circle and the maximum distance the user could walk from the last location as radius. The locations in the circle are the candidates. But the last location of the user recorded in database might not be the actual location which the user was, so we should extend the range of candidate locations properly because each record interval is the same. We use 2 times of the maximum distance as the radius of circle to limit the candidate locations. The process of calculation can be described as $r = 2vT$, where r is the radius of the circle containing candidate locations, v is the pedestrian walk speed (here we use 1.5m/s) and T is the time from the last record to this record.

The reference points located in the circle with the radius of r, which is centered at the last location of the user are the candidate locations of the user's current location. Thus, the location of the user must be one of the candidate locations.

4.4   Determining the Location

Similar with the base Fingerprint approach, we use the reference point with the most similar signal strength values as the user's location. We use the Equation (3) to compute the degree of similarity.

$$sim_c = \sum_{i=1}^{n} \left(w_i * (s_i - r_{c,i})^2 \right) \tag{3}$$

where sim_c is the similarity between the signal strength values received by the user and the signal strength values of the cth reference point which is a candidate location, n is the number of APs which can be detected by the user's device, w_i is the weight of AP_i's signal strength value, s_i is the Smooth Value of AP_i's signal strength value and $r_{c,i}$ is the AP_i's signal strength value of the cth reference point.

The reference point with the signal strength closest to that received by the user (sim_c is the minimum) would be returned to the user as the estimated location .
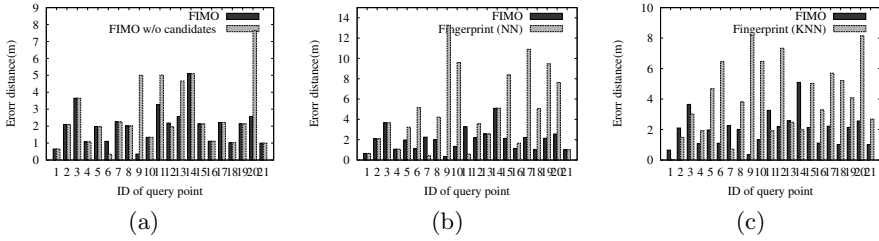
---

[6] See http://en.wikipedia.org/wiki/Walking

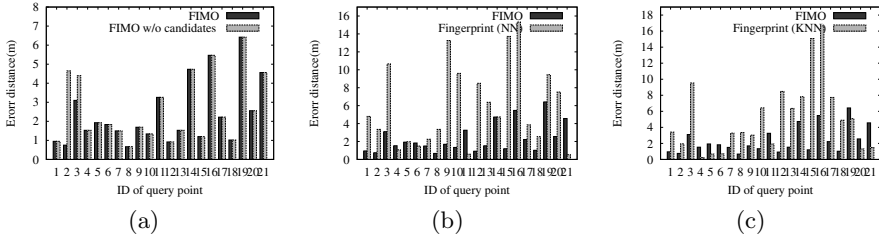**Fig. 6.** The query points of the first user

Fig. 7. The query points of the second user

5   Experiments

We have conducted series of experiments on the testbed described in Section 3.1. The testing data sets are also introduced in Section 3.2. We use four different approaches to locate the user. Two approaches are ours, one is FIMO introduced in Section 4, the other is FIMO without considering the walk speed. We use the Fingerprint using the nearest neighbor (Fingerprint(NN))and the Fingerprint using the multiple nearest neighbors (Fingerprint(kNN)) introduced in Section 2.2 as comparisons.

Figure 6 and 7 show the results of localization. Here, x-axis represents the ID of querying requirement, and the y-axis represents the error. In Figure 6(b)-(c) and 7(b)(c), we can observe that FIMO is much better than the Fingerprint(NN) and Fingerprint(kNN). In FIMO, considering the walk speed is the only step that might requires other device or estimation. In Figure 6(a) and 7(a), we can see the FIMO without considering the walk speed of the user already has good accuracy, but the FIMO considered the walk speed could generate better results.

Table 1 shows the sums of squares among the groups[7] (SSA) in Figure 6 and 7. The statistics suggests that the accuracy of FIMO has improved significantly comparing with the Fingerprint(NN) and Fingerprint(kNN). The FIMO is proved to have the best accuracy with respect to all the parameters (i.e. the average, medium and maximum values,as shown in Table 2). The errors in distances of all query points are less than 7m, and the points erred within 3m are more than 70%.

---

[7] SSA is a measure of how much the means differ from one another. Its conceptualized a little differently, because it is thought of as the variation of each mean from the mean of the total sample.

Table 1. Sum of Squares Among the Groups (SSA)

	FIMO vs. FIMO w/o Candidates	FIMO vs. Fingerprint (NN)	FIMO vs. Fingerprint (kNN)
1	3.79	78.71	43.69
2	0.64	136.69	86.33

Table 2. The average, median and maximum of the error distances

	FIMO	FIMO w/o Candidates	Fingerprint (NN)	Fingerprint (kNN)
average	2.17m	2.59m	5.34m	4.62m
median	1.95m	1.99m	4.04m	3.61m
maximum	6.42m	7.64m	15.33m	16.66m

6   Conclusions and Future Works

In this paper, we presented FIMO, a WiFi-based localization method. In this approach, the system records the signal strength values the user receives while he is moving, and determines the user's location by backtracking the signal strength values and the user's movement. Experimental reports show that the performance is significantly improved over existing methods.

In this paper, our work merely applies to specific environmental. We are considering to expand the application to more complicated situations, such as multiple floors or models of access points etc. Since FIMO is based on fingerprint approach, a lot of data is required to be collected for reference points in advance. If the signal strength values of some reference points could be estimated, the setup would be less time consuming. The curve fitting might be a practicable method. We believe that it is a direction worth exploring in the future.

References

1. Bahl, P., Padmanabhan, V.N.: Radar: An in-building rf-based user location and tracking system. In: Proc. of INFOCOM, pp. 775–784 (2000)
2. Baniukevic, A., Sabonis, D., Jensen, C.S., Lu, H.: Improving wi-fi based indoor positioning using bluetooth add-ons. In: Proc. of MDM, pp. 246–255. IEEE Computer Society, Washington, DC (2011)

3. Bargh, M.S., de Groote, R.: Indoor localization based on response rate of bluetooth inquiries. In: Proc. of MELT, pp. 49–54. ACM, New York (2008)
4. Chen, Q., Lee, D.-L., Lee, W.-C.: Rule-based wifi localization methods. In: Proc. of EUC, pp. 252–258. IEEE Computer Society, Washington, DC (2008)
5. Chen, Y.-C., Chiang, J.-R., Chu, H.-H., Huang, P., Tsui, A.W.: Sensor-assisted wi-fi indoor location system for adapting to environmental dynamics. In: Proc. of MSWiM, pp. 118–125. ACM, New York (2005)
6. Hernández, N., Herranz, F., Ocaña, M., Bergasa, L.M., Alonso, J.M., Magdalena, L.: Wifi localization system based on fuzzy logic to deal with signal variations. In: Proc. of ETFA, pp. 317–322. IEEE Press, Piscataway (2009)
7. Ho, W., Smailagic, A., Siewiorek, D.P., Faloutsos, C.: An adaptive two-phase approach to wifi location sensing. In: Proc. of PERCOMW, pp. 452–456. IEEE Computer Society, Washington, DC (2006)
8. Jin, G.-Y., Lu, X.-Y., Park, M.-S.: An indoor localization mechanism using active rfid tag. In: Proc. of SUTC, pp. 40–43. IEEE Computer Society, Washington, DC (2006)
9. Lee, D.L., Chen, Q.: A model-based wifi localization method. In: Proc. of InfoScale, pp. 40:1–40:7. ICST (2007)
10. Letchner, J., Fox, D., Lamarca, A.: Large-scale localization from wireless signal strength. In: Proc. of the National Conference on Artificial Intelligence, pp. 15–20. The MIT Press (2005)
11. Seidel, S.Y., Rapport, T.S.: 914 MHz path loss prediction model for indoor wireless communications in multi-floored buildings. In: Proc. of Antennas Propagation. IEEE Trars (1992)
12. Wang, R., Zhao, F., Luo, H., Lu, B., Lu, T.: Fusion of wi-fi and bluetooth for indoor localization. In: Proc. of MLBS, pp. 63–66. ACM, New York (2011)
13. Yeh, L.-W., Hsu, M.-S., Lee, Y.-F., Tseng, Y.-C.: Indoor localization: Automatically constructing today's radio map by irobot and rfids. In: Proc. of Sensors, pp. 1463–1466 (2009)