Crowdsourcing-Based Indoor Propagation Model Localization Using Wi-Fi

Yongliang Sun[1,2](✉), Jian Wang[2], Wenfeng Li[2], Rui Jiang[1], and Naitong Zhang[2,3]

[1] School of Computer Science and Technology, Nanjing Tech University, Nanjing, China
syl_peter@163.com
[2] School of Electronic Science and Engineering, Nanjing University, Nanjing, China
[3] Communication Research Center, Harbin Institute of Technology, Harbin, China

Abstract. To save labor and time costs, crowdsourcing has been used to collect received signal strength (RSS) for building radio-map of Wi-Fi fingerprinting localization with common users' mobile devices. However, usually a great number of crowdsourcing data should be collected to calculate a satisfactory localization result. Therefore, we proposed a crowdsourcing-based indoor propagation model (PM) localization system in this paper. Our system only needs to collect crowdsourcing data at a few locations called crowdsourcing points, which can be easily finished in a short time. The system first eliminates RSS outliers in crowdsourcing data and then optimizes PM parameters using the processed data. Furthermore, the processed data is also used to estimate a distance between a user and the nearest crowdsourcing point for coordinate correction. Experimental results show that our system is able to achieve a comparable performance and the mean error of PM localization method is reduced from 7.12 m to 3.78 m.

Keywords: Crowdsourcing · Wi-Fi localization · Outlier processing · Propagation model · Coordinate correction

1 Introduction

With the development and popularization of mobile devices, demand for location-based services (LBS) has been increasing rapidly. Owing to limitations of satellite and cellular network-based localization systems in indoor environments, various indoor localization systems based on different techniques, such as Wi-Fi, infrared and ultrasound, have been developed [1]. Among them, the localization systems using Wi-Fi are favored because Wi-Fi has been widely deployed for communications and its received signal strength (RSS) can be easily measured by commonly available mobile devices [2]. Several localization methods using Wi-Fi have been proposed like fingerprinting, propagation model (PM), time of arrival (TOA), time difference of arrival (TDOA), and angle of arrival (AOA) [2–4].

Q. Chen et al. (Eds.): ChinaCom 2016, Part II, LNICST 210, pp. 587–596, 2018.
DOI: 10.1007/978-3-319-66628-0_56

Compared with TOA, TDOA and AOA, fingerprinting method has been extensively researched because it only needs software update and outperforms the other methods under non-line-of-sight (NLOS) environments. However, it requires a process of building a database called radio-map with location-labeled RSS samples. Real-time RSS samples are matched with the samples in the radio-map for coordinate estimation using fingerprinting algorithms like K-nearest neighbors (KNN), weighted KNN (WKNN) and artificial neural network (ANN) [3]. Usually the radio-map is established by professionals and the process involves intensive labor and time costs, so the application of fingerprinting method is limited to some degree. Regarding PM method, it needs no radio-map, but its performance heavily relies on the PM that is employed to calculate distances between a user and different access points (APs) for trilateration localization. So the performance of PM method is usually far from satisfactory.

Since crowdsourcing offers a new solution to data collection at low cost, it is very suitable for constructing the radio-map of fingerprinting method. Instead of professionals, crowdsourcing employs common users, namely crowdsourcing participants, to collect RSS and location information [5]. So far, several crowdsourcing-based fingerprinting localization systems have been proposed. Some of them labeled RSS samples with location information using indoor electronic maps [5,6]. Mirowski *et al.* deployed a number of two-dimensional code labels in their experimental environment, with which crowdsourcing participants are able to obtain location information of collected RSS samples [7]. Wu *et al.* recorded numerous trajectories of crowdsourcing participants with smartphone sensors and then matched the trajectories with RSS samples using multidimensional scaling (MDS) [8].

However, one problem of radio-map establishment through crowdsourcing is nearly no reliable localization coordinates can be obtained until enough crowdsourcing data are collected. To solve this problem, we apply crowdsourcing to PM method that requires no radio-map and propose a crowdsourcing-based indoor PM localization system using Wi-Fi. The proposed system only needs location-labeled RSS samples collected at a few crowdsourcing points (CPs). These RSS data can be easily collected in a short time and usually are not enough for fingerprinting method.

To improve the quality of crowdsourcing data, we eliminate RSS outliers with quartile method that does not involve RSS data distribution. Then the processed crowdsourcing data are used to optimize PM parameters for estimating more accurate distances between a user and APs, which improves the performance of trilateration localization. Because RSS data and location coordinates of CPs are known, the distance between a user and CP can be estimated and then used to correct coordinates calculated by trilateration localization. Before the collected crowdsourcing data are enough for building the radio-map of fingerprinting method, our proposed PM localization system is able to achieve a comparable performance. To the best of our knowledge, so far no localization system that employs crowdsourcing data to optimize PM parameters and also to correct trilateration localization results has been proposed.

The rest of this paper is organized as follows: Sect. 2 introduces the related works of our proposed crowdsourcing-based indoor PM localization system. In Sect. 3, a general frame of the proposed system is given and every part of it is described in details. The experimental setup, results and analyses are presented in Sect. 4. Finally, Sect. 5 concludes the paper.

2 Related Works

2.1 RSS Outlier Processing

Usually a certain amount of RSS samples should be collected by crowdsourcing participants. However, it will be boring for crowdsourcing participants to collect RSS data for a long time. Some researchers stated that the time for data collection at each CP should be less than one minute [9]. Thus, we collect RSS samples at every CP for one minute with a sampling rate of 2 RSS samples per second. Most statistical outlier detection methods, such as Grubbs, Chauvenet and three-sigma criterions, are based on the assumption that is data should follow Gaussian distribution. Because the RSS data of such amount from each AP are difficult to fit Gaussian distribution as shown in Fig. 1, we detect RSS outliers with quartile method that has no requirement for data distribution.

Assume that a total of K APs are deployed in an indoor environment and L RSS data from AP k denoted as $[r_1^{(k)}, r_2^{(k)}, \cdots, r_L^{(k)}],\ k \in (1, 2, \cdots, K)$ are collected, then these collected RSS data are sorted in a non-descending order as $[r_{(1)}^{(k)} \leq r_{(2)}^{(k)} \leq \cdots \leq r_{(L)}^{(k)}],\ k \in (1, 2, \cdots, K)$. We find the upper-quartile Q_{Up} and lower-quartile Q_{Low} of the sequence that split off the highest 25% and lowest 25% of data from the other ones, respectively. Then the inter-quartile range Q_{IQR} can be calculated by:

$$Q_{\mathrm{IQR}} = Q_{\mathrm{Up}} - Q_{\mathrm{Low}}. \tag{1}$$

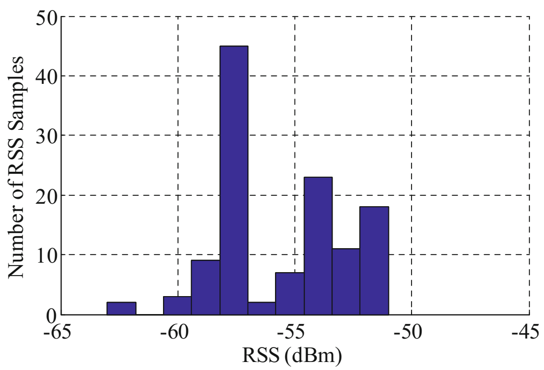

Fig. 1. Distribution of 120 RSS data at a crowdsourcing point.

With the inter-quartile range Q_{IQR} computed by (1), the boundaries for defining RSS outliers can be determined. If an RSS value $r_{(i)}^{(k)}$, $i \in (1, 2, \cdots, L)$ from AP k satisfies: $r_{(i)}^{(k)} < Q_{\mathrm{Low}} - 1.5Q_{\mathrm{IQR}}$ or $r_{(i)}^{(k)} > Q_{\mathrm{Up}} + 1.5Q_{\mathrm{IQR}}$, then it is recognized as an RSS outlier.

Compared with other RSS outlier detection methods, this quartile method does not require RSS data distribution information and has a low computation complexity. Also, RSS outliers have less influence on quartile calculation.

2.2 Propagation Model Optimization with Crowdsourcing Data

In our indoor office environment, we model the radio path loss between a user and AP with site-general model [10], which is given by:

$$P_{\mathrm{L}} = 20\log_{10}f + N\log_{10}d + P_f(n) - 28, \tag{2}$$

where P_{L} is the path loss in dB; f is the frequency in MHz; d is the distance between a user and AP in meters ($d > 1$ m); $P_f(n)$ is the floor penetration loss factor in dB; n is the number of floors between a user and AP; N is the distance power loss coefficient, which equals 30 at 2.4 GHz in office environments.

Because all the APs we utilize are deployed on the same floor as our experimental area, we remove $P_f(n)$ from (2). Let $P_{\mathrm{T}}^{(k,j)}$ and $P_{\mathrm{R}}^{(k,j)}$ be transmission power of AP k and received power measured by a crowdsourcing participant's mobile device at CP j, respectively. Then we rewrite (2) as:

$$P_{\mathrm{T}}^{(k,j)} - P_{\mathrm{R}}^{(k,j)} = 20\log_{10}f + N^{(k,j)}\log_{10}d^{(k,j)} - X^{(k,j)}, \tag{3}$$

where $X^{(k,j)}$ and $N^{(k,j)}$ are the PM parameters we need to optimize. $P_{\mathrm{T}}^{(k,j)}$ and $P_{\mathrm{R}}^{(k,j)}$ in dBm can be derived from AP configuration and RSS data, respectively. The distance between the crowdsourcing participant at CP j and AP k can be calculated by:

$$d^{(k,j)} = 10^{\dfrac{P_{\mathrm{T}}^{(k,j)} - P_{\mathrm{R}}^{(k,j)} - 20\log_{10}f + X^{(k,j)}}{N^{(k,j)}}}. \tag{4}$$

Let the known location coordinates of AP k and CP j be $(x_{\mathrm{AP}}^{(k)}, y_{\mathrm{AP}}^{(k)})$ and $(x_{\mathrm{CP}}^{(j)}, y_{\mathrm{CP}}^{(j)})$, respectively. Then the real distance between them is calculated by:

$$d_{\mathrm{Real}}^{(k,j)} = \sqrt{(x_{\mathrm{AP}}^{(k)} - x_{\mathrm{CP}}^{(j)})^2 + (y_{\mathrm{AP}}^{(k)} - y_{\mathrm{CP}}^{(j)})^2}. \tag{5}$$

Parameters X and N should be optimized to approach the minimum value of differences between the real and estimated distances described by (6). The problem can be considered as an unconstrained nonlinear multivariable optimization.

$$(\widehat{X}^{(k,j)}, \widehat{N}^{(k,j)}) = \underset{(\widehat{X}^{(k,j)}, \widehat{N}^{(k,j)})}{\arg\min} \left| d_{\mathrm{Real}}^{(k,j)} - d^{(k,j)} \right|. \tag{6}$$

All real and estimated distances between different CPs and APs are used for optimizing the two PM parameters. Then mean values $\widehat{\overline{X}}$ and $\widehat{\overline{N}}$ of the two optimized parameters are calculated, respectively.

2.3 Coordinate Correction with Crowdsourcing Data

Because location coordinates of CPs are known, if a distance between a user and the nearest CP is estimated, then the distance can be used to correct the user's localization coordinates. Assume that J CPs are selected and CP j is the nearest to a user who is at location i, as shown in Fig. 2. Let $P_{\mathrm{R}}^{(1,i)}$ and $P_{\mathrm{R}}^{(1,j)}$ be the received powers from AP 1 measured at location i and CP j, respectively. Then the received power $P_{\mathrm{R}}^{(1,i)}$ can be written as:

$$P_{\mathrm{R}}^{(1,i)} = P_{\mathrm{T}}^{(1,i)} - 20\log_{10}f - \bar{\hat{N}}\log_{10}d^{(1,i)} + \bar{\hat{X}}. \tag{7}$$

Using the similar equation for $P_{\mathrm{R}}^{(1,j)}$, we can have:

$$P_{\mathrm{R}}^{(1,j)} - P_{\mathrm{R}}^{(1,i)} = \bar{\hat{N}}\log_{10}\frac{d^{(1,i)}}{d^{(1,j)}}. \tag{8}$$

As shown in Fig. 2, the distance $d^{(i,j)}$ between location i and CP j should not be less than $|d^{(1,i)} - d^{(1,j)}|$ given by:

$$d^{(i,j)} \geq |d^{(1,i)} - d^{(1,j)}| = \left| 10^{\frac{P_{\mathrm{R}}^{(1,j)} - P_{\mathrm{R}}^{(1,i)}}{\bar{\hat{N}}}} - 1 \right| d^{(1,j)}. \tag{9}$$

When RSS samples from K APs are measured, (9) is applicable to all the K APs and the real distance $d_{\mathrm{Real}}^{(k,j)}$ between AP k and CP j is calculated with their coordinates, so we conclude that:

$$d^{(i,j)} \simeq \max_{k\in(1,2,\cdots,K)} \left| 10^{\frac{P_{\mathrm{R}}^{(k,j)} - P_{\mathrm{R}}^{(k,i)}}{\bar{\hat{N}}}} - 1 \right| d_{\mathrm{Real}}^{(k,j)}. \tag{10}$$

Even though the PM is optimized, due to variations of radio propagation environment, sometimes a localization result calculated by trilateration localization may deviate from its real location greatly. The estimated distance $d^{(i,j)}$ can be used as a restrictive condition to correct the localization result.

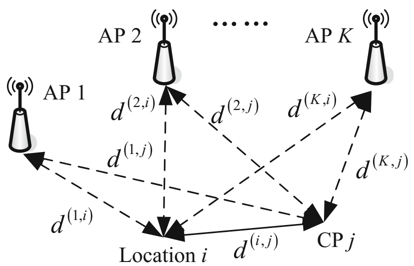

Fig. 2. Distance estimation between a user and crowdsourcing point.

3 Proposed Localization System

As shown in Fig. 3, our proposed crowdsourcing-based PM localization system is divided into three parts: data preparation, distance estimation and localization.

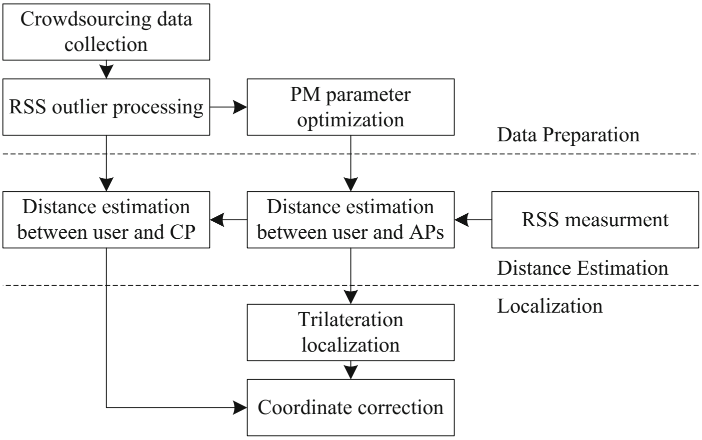

Fig. 3. Frame of the proposed localization system.

In the data preparation part, after L RSS samples from K APs are collected by a crowdsourcing participant at CP j. With the location coordinates of CP j that can be labeled on the ground, the collected RSS samples are merged with the coordinates of CP j as $\left[\boldsymbol{r}_1^{(j)}, \boldsymbol{r}_2^{(j)}, \cdots, \boldsymbol{r}_L^{(j)}, x_{\text{CP}}^{(j)}, y_{\text{CP}}^{(j)}\right]$, $j \in (1, 2, \cdots, J)$. Then L RSS data from AP k, $k \in (1, 2, \cdots, K)$ are examined with quartile method in order to eliminate RSS outliers. The recognized outliers are replaced by the mean values of normal data that have already been examined. Due to RSS data from only three APs are used by trilateration localization, we select the top three strongest crowdsourcing RSS data for optimizing the PM parameters. Then mean values $\bar{\bar{X}}$ and $\bar{\bar{N}}$ of the two optimized parameters are calculated.

Regarding distance estimation, when a user at location i measures K RSS data from all the K APs, the top three strongest ones are selected to estimate distances between the user and corresponding APs for trilateration localization. Also with the measured RSS data, the distance $d^{(i,j)}$ between the user and nearest CP j is estimated by (10) for localization coordinate correction.

In the last part of the localization system, localization coordinates $(\widehat{x}_i, \widehat{y}_i)$ are first calculated by trilateration localization with the three distances between the user and APs. Then the distance $\widehat{d}^{(i,j)}$ between the localization coordinates $(\widehat{x}_i, \widehat{y}_i)$ and CP j is calculated. If $\widehat{d}^{(i,j)}$ is greater than $d^{(i,j)}$, then $(\widehat{x}_i, \widehat{y}_i)$ are corrected to $(\widehat{x}_i', \widehat{y}_i')$ that are the final localization coordinates by:

$$\begin{cases} \hat{x}'_i = \left(\hat{x}_i - x_{\mathrm{CP}}^{(j)}\right) d^{(k,j)} \Big/ \hat{d}^{(k,j)} + x_{\mathrm{CP}}^{(j)} \\ \hat{y}'_i = \left(\hat{y}_i - y_{\mathrm{CP}}^{(j)}\right) d^{(k,j)} \Big/ \hat{d}^{(k,j)} + y_{\mathrm{CP}}^{(j)} \end{cases} . \tag{11}$$

4 Experimental Results and Analyses

4.1 Experimental Setup

Our experimental area is on an office floor with 9 Linksys WAP54G APs deployed. As shown in Fig. 4, the experimental area is a rectangular area of 24.9 m × 28.0 m and 4 APs are in the area. The area contains office rooms and a corridor that are two kinds of typical experimental environments for indoor localization. Because office rooms were not free to enter sometimes, we only selected 7 CPs in the corridor marked with blue points in Fig. 4. We selected CPs near the entrances of the floor, where it was convenient for crowdsourcing participants to collect data when they entered, and also at the corners of the corridor where radio propagation was more complicate due to multipath effect. We used a laptop to collect RSS samples for one minute at each CP with a sampling rate of 2 RSS samples per second. A total of 6500 RSS samples were collected in the shadow area for testing the proposed PM system. For performance comparison, a fingerprinting localization system was also performed in the experimental area. A total of 91 specific locations were selected and 300 RSS samples were measured at each selected location for radio-map establishment. The fingerprinting localization system was tested with the same 6500 RSS samples as PM localization.

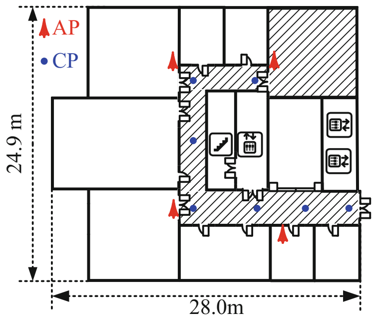

Fig. 4. Experimental area plan.

4.2 Results of RSS Outlier Processing

When we examine crowdsourcing RSS data with quartile method, the RSS outliers that are very different from the other data of the same AP are replaced by the mean values of the examined normal RSS data. Figure 5 shows the differences between original crowdsourcing RSS data and the processed ones from the same AP at one CP. The processed RSS data vary more smoothly after eliminating the RSS outliers. Meanwhile, RSS data from some APs may quite weak and these APs even cannot be sensed sometimes. Although our system exploits the top three strongest RSS data for trilateration localization and these APs have no influence on our system performance, the quartile method is also able to recognize these weak RSS data and eliminate them.

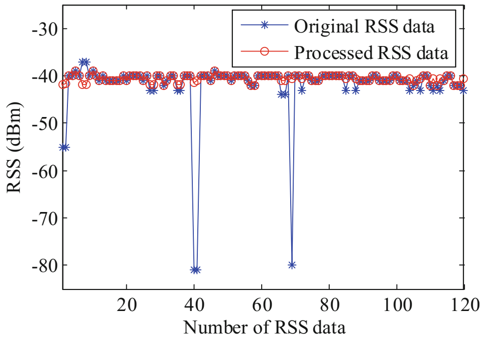

Fig. 5. Original and processed RSS data.

4.3 Localization Results with Optimized Propagation Models

We optimize PM parameters with processed crowdsourcing data and then compare performances of PM method with different optimized parameters. At first, we set parameter X to be 28 and only optimize parameter N as well as set parameter N to be 30 and only optimize parameter X both using single-variable optimization algorithm based on golden section search and parabolic interpolation. But the localization performances of PM method with the optimized parameters mentioned above are not satisfactory. Then we optimize parameters X and N at the same time with quasi-newton algorithm [11]. The starting values of the parameters X and N for quasi-newton algorithm are set equal to 28 and 30, respectively. Mean error with the two optimized parameters X and N is reduced to 5.05 m from 7.12 m. The mean errors and corresponding PM parameters X and N are listed in Table 1.

Table 1. Mean errors of propagation model localization with different parameters

X	N	Mean error (m)
28	30	7.12
28	30.46	6.58
27.67	30	6.85
25.86	30.91	5.05

4.4 Coordinate Correction with Estimated Distance Between a User and Crowdsourcing Point

To improve localization performance, we take advantage of CP data to correct localization results of optimized PM. In theory, the distance $d^{(i,j)}$ in (9) should be equal to the maximum one among all the distances $\left|d^{(k,i)} - d^{(k,j)}\right|$, $k \in (1, 2, \cdots, K)$. However, in practical application, radio propagation may vary significantly, so the maximum distance may be too large to correct localization coordinates. Thus, we eliminate distance outliers also with quartile method and take the median distance as $d^{(i,j)}$ to correct coordinates [12]. The mean error of optimized PM localization with coordinate correction are 3.78 m and its cumulative probabilities within 2 m and 3 m localization errors are 38.12% and 60.43%, respectively. As shown in Fig. 6, after correcting coordinates, the proposed system outperforms the others that also use PM method. By contrast, mean errors of fingerprinting algorithms KNN, WKNN and ANN are 2.77 mm, 2.74 mm and 2.55 mm, respectively. Although the performances of these fingerprinting algorithms are a little better than our proposed system, a total of $91 \times 300 = 27300$ RSS samples are collected to establish the radio-map, which is difficult in practical application. Our proposed system is able to achieve a comparable performance with the crowdsourcing data collected at only 7 CPs. Furthermore, 120 RSS samples at each CP can be easily collected in one minute.

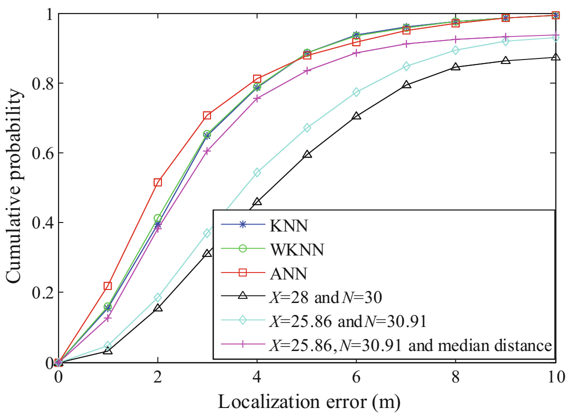

Fig. 6. Cumulative probability of localization errors.

5 Conclusion

In this paper, a crowdsourcing-based indoor PM localization system using Wi-Fi is proposed. Compared with traditional PM and fingerprinting localization systems, the proposed system has a greatly improved localization performance without intensive labor for data collection. The system makes use of RSS samples and location information of only a few CPs, which can be easily collected by crowdsourcing participants. RSS outliers in crowdsourcing data are first eliminated and the processed data are used for optimizing PM parameters. So the performance of trilateration localization for a user is improved. Then an estimated distance between the user and nearest CP is calculated and the localization results of trilateration localization can be corrected with the estimated distance. Experimental results confirm the effectiveness of our proposed crowdsourcing-based indoor PM localization system and that the system is able to achieve a comparable performance with easily collected crowdsourcing data.

Acknowledgment. The authors gratefully thank the referees for the constructive and insightful comments. This work was supported by the Natural Science Foundation of the Jiangsu Higher Education Institutions of China under Grant No. 16KJB510014 and partially by the Cooperative Innovation Foundation of Jiangsu Province Prospective Joint Research Project under Grant No. BY2014126-02.

References

1. Liu, H., Darabi, H., Banerjee, P., Liu, J.: Survey of wireless indoor positioning techniques and systems. IEEE Trans. Syst. Man Cybern. **37**(6), 1067–1080 (2007)
2. He, S.N., Chan, S.H.: Wi-Fi fingerprint-based indoor positioning: recent advances and comparisons. IEEE Commun. Surv. Tutor. **18**(1), 466–490 (2016)
3. Sun, Y.L., Xu, Y.B.: Error estimation method for matrix correlation-based Wi-Fi indoor localization. KSII Trans. Internet Inf. Syst. **7**(11), 2657–2675 (2013)
4. Makki, A., Siddig, A., Saad, M., Cavallaro, J.R., Bleakley, C.J.: Indoor localization using 802.11 time differences of arrival. IEEE Trans. Mob. Comput. **99**, 1–10 (2015)
5. Rai, A., Chintalapudi, K.K., Padmanabhan, V.N., Sen, R.: Zee: zero-effort crowdsourcing for indoor localization. In: 2012 ACM 20th Annual International Conference on Mobile Computing and Networking, pp. 1–12 (2012)
6. Park, J.G., Charrow, B., Curtis, D., Battat, J., Minkov, E., Hicks, J., Teller, S., Ledlie, J.: Growing an organic indoor location system. In: 2010 ACM 8th International Conference on Mobile Systems, Applications, and Services, pp. 271–284 (2010)
7. Mirowski, P., Tin, K.H., Saehoon, Y., Macdonald, M.: SignalSLAM: simultaneous localization and mapping with mixed WiFi, bluetooth, LTE and magnetic signals. In: 2013 International Conference on Indoor Positioning and Indoor Navigation, pp. 1–10 (2013)
8. Wu, C.S., Yang, Z., Liu, Y.H.: Smartphones based crowdsourcing for indoor localization. IEEE Trans. Mob. Comput. **14**(2), 444–457 (2015)
9. Yang, S., Dessai, P., Verma, M., Gerla, M.: FreeLoc: calibration-free crowdsourced indoor localization. In: 2013 IEEE International Conference on Computer Communications, pp. 2481–2489 (2013)
10. Rao, T.R., Balachander, D.: RF propagation investigations at 915/2400 MHz in indoor corridor environments for wireless sensor communications. Prog. Electromagnet. Res. B **47**, 359–381 (2013)
11. Aderibigbe, F.M., Adebayo, K.J., Dele-Rotimi, A.O.: On quasi-newton method for solving unconstrained optimization problems. Am. J. Appl. Math. **3**(2), 47–50 (2015)
12. Pourahmadi, V., Valaee, S.: Indoor positioning and distance-aware graph-based semi-supervised learning method. In: IEEE 2012 Global Telecommunications Conference, pp. 315–320 (2012)