Accuracy Enhancement with Integrated Database Construction for Indoor WLAN Localization

Qiao Zhang(✉), Mu Zhou, and Zengshan Tian

Chongqing Key Lab of Mobile Communications Technology,
Chongqing University of Posts and Telecommunications,
Chongqing 400065, China
18716322725@139.com, {zhoumu, tianzs}@cqupt.edu.cn

Abstract. In this paper, we rely on the neighborhood relations of the physically adjacent Reference Points (RPs) to construct a physical neighborhood database with the purpose of enhancing the accuracy of the Receive Signal Strength (RSS) fingerprint based localization algorithms in Wireless Local Area Network (WLAN) environment. First of all, based on the Most Adjacent Points (MAPs) and their corresponding Physically Adjacent Points (PAPs), we construct the Feature Groups (FGs), and then calculate the New Reference Point (NRP) with respect to each FG. Second, the RSS at each NRP is estimated by using the least square method based surface interpolation algorithm. Finally, we apply the K Nearest Neighbor (KNN), Weighted KNN (WKNN), and Bayesian inference algorithms to locate the target. The experimental results show that the proposed integrated database construction helps a lot in improving the localization accuracy of the widely-used KNN, WKNN, and Bayesian inference algorithms.

Keywords: WLAN localization · Location fingerprinting · Physical neighborhood · Reference Points · Received Signal Strength

1 Introduction

With the significant development of Wireless Local Area Network (WLAN) technique and the wide deployment of WLAN Access Points (APs) in public environments, it is particularly valuable and cost-efficient to rely on the WLAN infrastructures and the off-the-shelf smartphones to conduct the people's location tracking [1]. By employing the WLAN Received Signal Strength (RSS), the WLAN fingerprint based localization techniques have been carefully studied in recent decade due to the advantages of the free ISM band and high enough accuracy performance [2–4]. And as far as we know, most of the existing RSS fingerprint based localization techniques do not pay much attention to the physical adjacency relations of Reference Points (RPs), while in fact, these relations can help a lot in improving the localization accuracy [5, 6]. On this basis, we construct an integrated physical neighborhood and location fingerprinting database in off-line phase. Then, in on-line phase, we first select the k RPs with the RSSs having the smallest distances from the newly recorded RSS by the target as the k Most Adjacent Points (MAPs). Second, based on the physical neighborhood

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2017
X.-L. Huang (Ed.): MLICOM 2016, LNICST 183, pp. 169–177, 2017.
DOI: 10.1007/978-3-319-52730-7_17

database, we search the n Physically Adjacent Points (PAPs) corresponding to the k MAPs. Third, we use every c MAPs and PAPs (also named as Feature Points (FPs)) to construct a Feature Group (FG). Obviously, the number of FGs equals to C_{k+n}^{c}. Fourth, in each FG, we calculate a New Reference Point (NRP), as well as estimate the RSS at the NRP by using the least square method based surface interpolation algorithm. Finally, based on the C_{k+n}^{c} NRPs, we apply the RSS fingerprint based localization algorithm (e.g., K Nearest Neighbor (KNN), Weighted KNN (WKNN), and Bayesian inference) to locate the target.

The rest of this paper is structrued as follows. In Sect. 2, we show the steps of the integrated physical neighborhood and location fingerprinting database construction. In Sect. 3, the performance of the integrated database for indoor WLAN localization is examined. The experimental results with the WLAN RSSs recorded in an actual indoor WLAN environment are provided in Sect. 4. Finally, Sect. 5 concludes the paper and provides some future directions.

2 Integrated Database Construction

The physical neighborhood database is constructed based on the physical layout of the target environment. In concrete terms, for the physical layout of an actual indoor WLAN environment [9–17], we represent each office room or each segment of straight corridors as a representative node. Every two adjacent representative nodes are connected by an edge. Then we obtain a physical graph describing the physical layout of the target environment. To construct the physical neighborhood database, we first label each RP with a unique Reference Point Identifier (RPID). Second, based on the geographic relations of RPs described in physical graph, we construct a set of r adjacent PAPs with respect to each RP. Finally, we construct the physical neighborhood database consisting of the sets of PAPs for all the RPs. And the construction of location fingerprinting database consists of two main steps as follow. First of all, we record a sequence of RSS measurements at each RP, notated as $\{S_1 = (\mathrm{rss}_{11}, \mathrm{rss}_{12,} \ldots, \mathrm{rss}_{1w}), S_2 = (\mathrm{rss}_{21}, \mathrm{rss}_{22}, \ldots, \mathrm{rss}_{2w}), \ldots\}$, where w is the number of APs and rss_{ij} is the RSS value from the j-th AP in S_i. Second, we calculate the mean and standard deviation of RSS at each RP to form a RSS fingerprint. Finally, the location fingerprinting database is constructed to describe the relationship between the RSS fingerprints and the locations of RPs.

3 Accuracy Enhancement for Indoor WLAN Localization

To enhance the localization accuracy of the KNN, WKNN, and Bayesian inference algorithms, we first select the k MAPs with respect to each newly recorded RSS. Second, using the physical neighborhood database, we search the n PAPs corresponding to the k MAPs, and then calculate the coordinates of the C_{k+n}^{c} NRPs. Finally, the localization algorithm (e.g., KNN, WKNN, and Bayesian inference) is applied to estimate the locations of the target.

In the results that follow, we mainly focus on the three typical combinational localization algorithms: (i) KNN based WKNN (i.e., KNN is applied to calculate the NRPs, and then WKNN is used to estimate the locates of the target), named as KbW; (ii) WKNN based Bayesian inference (i.e., WKNN is applied to calculate the NRPs, and then Bayesian inference is used to estimate the locates of the target), named as WbB; and (iii) Bayesian inference based KNN (i.e., Bayesian inference is applied to calculate the NRPs, and then KNN is used to estimate the locates of the target), named as BbK.

3.1 Combinational Localization Algorithms

Steps of KbW. After the k MAPs are selected, we search the n PAPs corresponding to the k MAPs by using the physical neighborhood database. After that, we construct C_{k+n}^{c} FGs, and the NRP in each FG, P_{ref}, is calculated by

$$P_{ref} = \frac{1}{c}\sum\nolimits_{i=1}^{c} (x_{zi}, y_{zi}) \tag{1}$$

where (x_{zi}, y_{zi}) is the 2-dimensional (2-D) coordinates of the i-th FP in the z-th FG. To estimate the RSS at P_{ref}, we assume that in each FG, the relationship between the coordinates of FPs and their corresponding RSSs satisfies

$$s_{zi-j} = ax_{zi} + by_{zi} + d + \delta_{zi} \tag{2}$$

where s_{zi-j} is the RSS from the j-th AP at the i-th FP in the z-th FG. δ_{zi} is the RSS distance between s_{zi-j} and the estimated RSS at the i-th FP. By using the least square method based surface interpolation algorithm, the coefficients a, b, and d in (2) are calculate by

$$\begin{cases} a = \frac{c(sy)c(xy) - c(sx)d(y)}{(c(xy))^2 - d(x)d(y)}, \\ b = \frac{c(sx)c(xy) - c(sy)d(x)}{(c(xy))^2 - d(x)d(y)}, \\ d = \overline{s} - a\overline{x} - b\overline{y}. \end{cases} \tag{3}$$

where

$$\begin{cases} c(sy) = \sum\nolimits_{i=1}^{c} \left(s_{zi-j} - \overline{s}\right)\left(y_{zi} - \overline{y}\right), \\ c(xy) = \sum\nolimits_{i=1}^{c} \left(x_{zi} - \overline{x}\right)\left(y_{zi} - \overline{y}\right), \\ c(sx) = \sum\nolimits_{i=1}^{c} \left(s_{zi-j} - \overline{s}\right)\left(x_{zi} - \overline{x}\right), \\ d(x) = \sum\nolimits_{i=1}^{c} \left(x_{zi} - \overline{x}\right)^2, \\ d(y) = \sum\nolimits_{i=1}^{c} \left(y_{zi} - \overline{y}\right)^2, \\ \overline{s} = \frac{1}{c}\sum\nolimits_{i=1}^{c} s_{zi-j}, \overline{x} = \frac{1}{c}\sum\nolimits_{i=1}^{c} x_{zi}, \overline{y} = \frac{1}{c}\sum\nolimits_{i=1}^{c} y_{zi}. \end{cases} \tag{4}$$

After the coefficients in (2) are calculated, we estimate the RSS at P_{ref} based on the fitted surface function [8], $z = ax + by + d$. Finally, WKNN is applied to estimate the locations of the target P_{user}, as described in (5).

$$P_{user} = \frac{\sum_{i=1}^{q} (1/d_i)(x_i, y_i)}{\sum_{i=1}^{q} (1/d_i)} \tag{5}$$

where d_i is the distance between the estimated RSS at the i-th selected NRP, (x_i, y_i), and the newly recorded RSS. In KbW, the selected NRPs are the NRPs with the RSSs having the smallest distances from the newly recorded RSS.

Steps of WbB. In WbB, we apply WKNN to calculate the NRP in each FG, as shown in (6).

$$P_{ref} = \frac{\sum_{i=1}^{c} (1/d_{zi})(x_{zi}, y_{zi})}{\sum_{i=1}^{c} (1/d_{zi})} \tag{6}$$

where d_{zi} is the distance between the estimated RSS at the i-th FP in the z-th FG, (x_{zi}, y_{zi}), and the newly recorded RSS. The estimation of RSS at each NRP in WbB follows the same steps involved in KbW. After that, we use Bayesian inference to calculate the posterior probability of each NRP with respect to each newly recorded RSSs. We take the NRP L_f as an example. By using the Bayesian inference, the posterior probability of L_f with respect to s, $P\left(L_f|s\right)$, is equivalent to the product of the prior probabilities, as shown in (7).

$$P\left(s|L_f\right) = P\left(s_1|L_f\right)P\left(s_2|L_f\right)\ldots P\left(s_w|L_f\right) \tag{7}$$

where s_j is the newly recorded RSS from the j-th AP. By assuming that the RSS distribution at each NRP obeys a Gaussian distribution, we have

$$P(s_j|L_f) = \frac{1}{\sqrt{2\pi}\delta}\exp[\frac{-(s_j-\mu)^2}{2\delta^2}] \tag{8}$$

where μ and δ are the mean and standard deviation of the RSS distribution from the j-th AP at L_f respectively. Then, we rely on the q selected NRPs with the largest posterior probabilities with respect to the newly recorded RSS to estimate the locations of the target, as shown in (9).

$$P_{user} = \frac{\sum_{i=1}^{q} \mathrm{pro}_i (x_i, y_i)}{\sum_{i=1}^{q} \mathrm{pro}_i} \tag{9}$$

where pro_i and (x_i, y_i) are the posterior probability and the 2-D coordinates of the i-th selected NRP respectively.

Steps of BbK. In BbK, we calculate the NRP P_{ref} in each FG by

$$P_{ref} = \frac{\sum_{i=1}^{c} \text{pro}_{zi}(x_{zi}, y_{zi})}{\sum_{i=1}^{c} \text{pro}_{zi}} \tag{10}$$

where pro_{zi} and (x_{zi}, y_{zi}) are the posterior probability and the 2-D coordinates of the i-th FP in the z-th FG. The steps of the estimation of RSS at each NRP in BbK are the same to the ones in KbW and WbB. After the estimated RSS at each NRP is obtained, we can estimate the locations of the target by

$$P_{user} = \frac{1}{q} \sum_{i=1}^{q} (x_i, y_i) \tag{11}$$

where the 2-D coordinates of the q selected NRPs are denoted as $(x_i, y_i)(i = 1, , q)$. In BbK, the selected NRPs are the NRPs with the RSSs having the smallest distances from the newly recorded RSS.

3.2 Modified Bayesian Inference

In WbB, the RSS distribution at each location is assumed to obey a Gaussian distribution, while in fact, the Gaussian distribution of RSS cannot always be approximately obeyed especially in the Non-Line-of-Sight (NLOS) scenario. To solve this problem, we use (12) to calculate the similarity between the RSS distributions at each RP and the distributions of the newly recorded RSSs.

$$S_i = \frac{1}{\sum_{j=1}^{w} \left(\sum_{x=\text{RSS}_{\text{lower}}}^{\text{RSS}_{\text{upper}}} P_{on-j}(x) \ln \frac{P_{on-j}(x)}{Q_{ij}(x)} \right)} \tag{12}$$

where $Q_{ij}(x)$ and $P_{on-j}(x)$ are the RSS distribution from AP j at the i-th RP and the distribution of the newly recorded RSSs from AP j respectively. The RSS value x is in the range of $[\text{RSS}_{\text{lower}}, \text{RSS}_{\text{upper}}]$.

4 Experimental Results

4.1 Accuracy Discussion

We conduct the experiments in an actual indoor WLAN environment with the dimensions of 66 m × 22 m. The target environment is covered by 9 Cisco WRT54G APs which are placed on the same floor in a building [7], as shown in Fig. 1. The 182 RPs are uniformly calibrated with the same interval of 1 m and the 81 test points (TPs) are randomly selected in five straight corridors for the testing.

Figure 2 compare the Cumulative Distribution Functions (CDFs) of errors between the proposed Combinational Localization Algorithms (i.e., KbW, WbB, and BbK) and the conventional WKNN, Bayesian inference, and KNN, named as C-W, C-B, and C-K

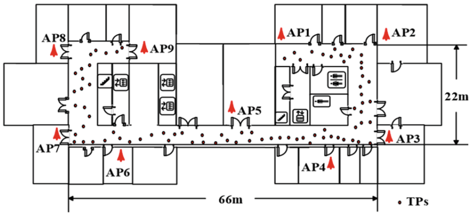

Fig. 1. Experimental layout.

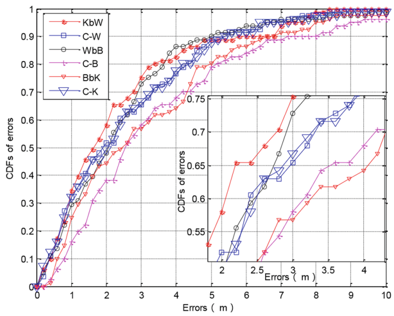

Fig. 2. CDFs of errors by KbW, WbB, BbK, C-W, C-B, and C-K.

respectively. As can be seen from Fig. 2, the proposed algorithms generally perform better than the conventional WKNN, Bayesian inference, and KNN in localization accuracy. We take KbW as an example. By using KbW, the probabilities of errors within 3 m and 2.5 m are about 10% and 5% more than the ones achieved by C-W respectively.

4.2 Parameter Discussion

To examine the performance of the proposed integrated database construction for indoor WLAN localization, we use the control variable approach to investigate the relationship between the localization errors and the four parameters as follows: (i) number of MAPs, k; (ii) number of the adjacent PAPs for each RP, r; (iii) number of FPs in each FG, c; and (iv) number of NRPs, q.

The optimal parameters which are corresponding to the smallest mean of errors for all the combinational localization algorithms and the conventional WKNN, Bayesian inference, and KNN are shown in Table 1.

From Table 1, we can observe that: (i) by using the integrated database, most of the combinational localization algorithms achieves lower mean of errors compared to C-W, C-B, and C-K; and (ii) the lowest mean of errors, 2.2946 m, is obtained by KbK. On this basis, the integrated physical neighborhood and location fingerprinting database is

Table 1. Parameters vs. Errors

Algorithms	Optimal parameters	Mean of errors (m)
KbW	$k = 9,\ r = 3,\ c = 5,\ q = 17$	2.2962
WbW	$k = 5,\ r = 5,\ c = 4,\ q = 17$	2.3416
BbW	$k = 17,\ r = 3,\ c = 2,\ q = 5$	3.1147
C-W	$k = 9$	2.5213
KbB	$k = 9,\ r = 3,\ c = 5,\ q = 1$	2.5333
WbB	$k = 5,\ r = 5,\ c = 4,\ q = 17$	2.4483
BbB	$k = 17,\ r = 3,\ c = 2,\ q = 5$	3.3674
C-B	$k = 7$	3.4197
KbK	$k = 9,\ r = 3,\ c = 5,\ q = 17$	2.2946
WbK	$k = 5,\ r = 5,\ c = 4,\ q = 17$	2.3418
BbK	$k = 17,\ r = 3,\ c = 2,\ q = 5$	3.1453
C-K	$k = 9$	2.5346

proved to be able to enhance the accuracy of the conventional indoor WLAN RSS fingerprint based localization algorithms.

Finally, to verify the efficiency of the modified Bayesian inference for indoor WLAN RSS fingerprint based localization, we compare the CDFs of errors by C-B, WbB, and the WbB using the modified Bayesian inference. In Fig. 3, we find that: (i) the C-B performs poorest in localization accuracy with the probabilities of errors within 3 m lower than 60%; (ii) compared to the C-B, the higher localization accuracy is achieved by using the proposed WbB with the probabilities of errors within 3 m more than 70%; and (iii) there is a further improvement in localization accuracy when the WbB using the modified Bayesian inference is adopted.

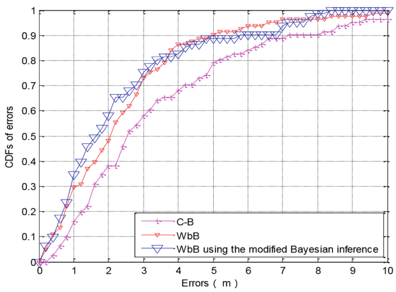

Fig. 3. CDFs of errors by C-B, WbB, and the WbB using the modified Bayesian inference.

5 Conclusion

A novel approach to improve the accuracy of the indoor WLAN RSS fingerprint based localization by using the integrated physical neighborhood and location fingerprinting database is proposed in this paper. We not only construct location fingerprinting database, but also utilize the physical adjacency relations of RPs to construct the physical neighborhood database. The extensive experiments demonstrate that the integration of the physical neighborhood database and location fingerprinting database can help a lot in improving the accuracy of the widely-used KNN, WKNN, and Bayesian inference algorithms. For the future work, the more accurate and efficient estimation of the RSS at each NRP forms an interesting topic.

Acknowledgment. This work was supported in part by the Program for Changjiang Scholars and Innovative Research Team in University (IRT1299), National Natural Science Foundation of China (61301126), Special Fund of Chongqing Key Laboratory (CSTC), and Fundamental and Frontier Research Project of Chongqing (cstc2013jcyjA40041 and cstc2013jcyjA40032) and Postgraduate Scientific Research and Innovation Project of Chongqing (CYS16157).

References

1. Zhou, M., Wong, A.K., Tian, Z., Zhang, V.Y., Yu, X., Luo, X.: Adaptive mobility mapping for people tracking using unlabelled Wi-Fi shotgun reads. IEEE Commun. Lett. **17**(1), 87–90 (2013)
2. Bahl, P., Padmanabhan, V.N.: RADAR: an in-building RF-based user locationand tracking system. In: IEEE INFOCOM, vol. 2, pp. 775–784 (2000)
3. Zhou, M., Wong, A.K., Tian, Z., Luo, X., Xu, K., Shi, R.: Personal mobility mapconstruction for crowd-sourced Wi-Fi based indoor mapping. IEEE Commun. Lett. **18**(8), 1427–1430 (2014)
4. Sun, Y., Xu, Y., Ma, L., Deng, Z.: KNN-FCM hybrid algorithm for indoor location on WLAN. In: The 2nd International Conference on Power Electronics and Intelligent Transportation System, pp. 251–254 (2009)
5. Tang, D., Liu, D., Xu, Z., Jang, P.: Research on indoor localization technology based on nearest neighbor points database. Transducer Microsyst. Technol. **32**(9), 69–71 (2013)
6. Alasti, H., Xu, K., Dang, Z.: Efficient experimental path loss exponent measurement for uniformly attenuated indoor radio channels. In: The 10th IEEE Southeast Conference, pp. 255–260 (2009)
7. Ma, L., Xu, Y., Wu, D.: A novel two-step WLAN indoor positioning method. J. Comput. Inf. Syst. **6**(14), 4627–4636 (2010)
8. Tian, L., Liu, Z.: Least-squares method piecewise linear fitting. Comput. Sci. **39**(6A), 482–484 (2012)
9. Zhou, M., Zhang, Q., Tian, Z., Qiu, F., Wu, Q.: Integrated location fingerprinting and physical neighborhood for WLAN probabilistic localization. In: International Conference on Computing, Communication and Networking Technologies, pp. 1–5 (2014)
10. Zhou, M., Qiu, F., Kunjie, X., Tian, Z., Haibo, W.: Error bound analysis of indoor Wi-Fi location fingerprint based positioning for intelligent access point optimization via fisher information. Comput. Commun. **86**, 57–74 (2016)
11. Zhou, M., Zhang, Q., Tian, Z., Xu, K., Qiu, F., Wu, H.: IMLours: indoor mapping and localization using time-stamped WLAN received signal strength. In: IEEE Wireless Communications and Networking Conference, pp. 1817–1822 (2015)
12. Jiang, Q., Li, K., Zhou, M., Tian, Z., Xiang, M.: Competitive agglomeration based KNN in indoor WLAN localization environment. In: 10th International Conference on Communications and Networking in China, pp. 338–342 (2015)

13. Zhou, M., Zhang, Q., Tian, Z., Qiu, F., Wu, Q.: Correlated received signal strength correction for radio-map based indoor Wi-Fi localization. In: International Conference on Computing, Communication and Networking Technologies, pp. 1–6 (2014)
14. Zhou, M., Qiu, F., Tian, Z., Haibo, W., Zhang, Q., He, W.: An information-based approach to precision analysis of indoor WLAN localization using location fingerprint. Entropy **17** (12), 8031–8055 (2015)
15. Tian, Z., Liu, X., Zhou, M., Xu, K.: Mobility tracking by fingerprint-based KNN/PF approach in cellular networks. In: IEEE Wireless Communications and Networking Conference, pp. 4570–4575 (2013)
16. Zhou, M., Tian, Z., Kunjie, X., Xiang, Yu., Haibo, W.: Theoretical entropy assessment of fingerprint-based Wi-Fi localization accuracy. Expert Syst. Appl. **40**(15), 6136–6149 (2013)
17. Zhou, M., Zhang, Q., Kunjie, X., Tian, Z., Wang, Y., He, W.: PRIMAL: page rank-based indoor mapping and localization using gene-sequenced unlabeled WLAN received signal strength. Sensors **15**(10), 24791–24817 (2015)