

Reducing Calibration Effort for Indoor WLAN Localization Using Hybrid Fingerprint Database

Mu Zhou, Yunxia Tang^(✉), Zengshan Tian, and Feng Qiu

Chongqing Key Lab of Mobile Communications Technology,
Chongqing University of Posts and Telecommunications, Chongqing 400065, China
{zhoumu,tianzs}@cqupt.edu.cn, 13629735505@139.com,
qiufeng245@outlook.com

Abstract. Due to the implementation ease and cost-efficiency, the indoor Wireless Local Area Network (WLAN) fingerprint based localization approach is preferred compared with the conventional trilateration localization approaches. In this paper, we propose a new semi-supervised learning algorithm based on manifold alignment with cubic spline interpolation to reduce the offline calibration effort for indoor WLAN localization using hybrid fingerprint database. The proposed approach significantly reduces the number of labeled training samples collected at each survey location by constructing the hybrid database via interpolation and semi-supervised manifold learning. We carry out extensive experiments in a ground-truth indoor environment to examine the localization accuracy of the proposed approach. The experimental results demonstrate that our approach can effectively reduce the calibration effort, as well as achieve high localization accuracy.

Keywords: WLAN · Location fingerprint · Interpolation · Semi-supervised learning · Manifold alignment

1 Introduction

With the development of light-weighted mobile devices, Location-based Services (LBSs) have gained considerable attention over the last decade due to the potential in the technology and the significant challenges facing this area of research [1]. The popular Global Positioning System (GPS) has been recognized as a success for outdoor localization, but it is generally not applicable for the indoor environment. The conventional indoor localization systems based on the infrared ray [2], ultrasound [3], video [4], and Radio Frequency (RF) techniques [5–10] have been widely studied. The RF technique has the advantage of ubiquitous coverage by using the inexpensive Wireless Local Area Network (WLAN). Due to the considerations of cost overhead and localization accuracy, the fingerprint database based indoor WLAN localization has been widely studied. Two phases are involved in the fingerprint database based indoor WLAN localization, namely offline phase and online phase. In off-line phase, we calibrate a series of Reference

Points (RPs) in target area, and then collect the Received Signal Strength (RSS) measurements from hearable Access Points (APs) at each RP to construct the fingerprint database, namely radio-map. In on-line phase, we match the newly collected RSS measurements against the radio map to estimate target location.

Since the point-by-point RP calibration is time-consuming, we aim to reduce the offline calibration effort, as well as maintain the high-enough localization accuracy. The proposed approach reduces both the number of labeled samples collected at survey points and number of survey points. However, reducing the number of the labeled samples and survey points may result in the inaccurate radio-map and deteriorate the localization accuracy. To solve this problem, the proposed approach relies on the cubic spline interpolation algorithm to obtain the predicted location fingerprints, and employs the manifold alignment (MA) algorithm [10] to label the locations at which we collect the sequences of RSS measurements according to the users motion traces. Since the users motion traces can be recorded easily without the labeling process, our approach is able to reduce the labor cost, as well as improve the accuracy of the reconstructed fingerprint database.

The remainder of the paper is organized as follows. Section reviews some related work. In Section, we introduce the proposed approach in detail. Section conducts the performance evaluation under different parameters and shows the experimental results. Finally, the conclusion is provided in Section.

2 Related Work

There are bathes of studies focusing on the indoor WLAN localization, such as the RADAR [5] and Horus [6]. Although the approaches in [5] and [6] can achieve high localization accuracy, a large number of RSS measurements are required to be collected and manually labeled at survey points. Since the labeled RSS measurements collection is time consuming and labor intensive, the existing literatures mainly focused on using the unlabeled data to reduce the time overhead involved in offline phase. In [7], the authors addressed a label propagation algorithm based semi-supervised learning approach to construct a hybrid database of labeled and unlabeled data using the concept that the similar data are corresponding to the similar labels. In [8], the authors proposed a hybrid generative and discriminative semi-supervised learning algorithm by predicting a large amount of unlabeled data to replenish the sparse labeled database, and meanwhile the online test data are selected as the offline unlabeled data and the labels of unlabeled data are learned from the labeled data. The authors in [9] exploited a new approach in which a manifold-based model is built from a batch of labeled and unlabeled data in offline phase, and then the weighted K Nearest Neighbor (KNN) algorithm is used to estimate the target locations in online phase. However, since the aforementioned approaches significantly depend on the RSS measurements, the performance could be seriously degraded when the RSS changes abruptly among the neighboring RPs. The main contribution of this paper is that we build a more accurate and reliable radio map by using

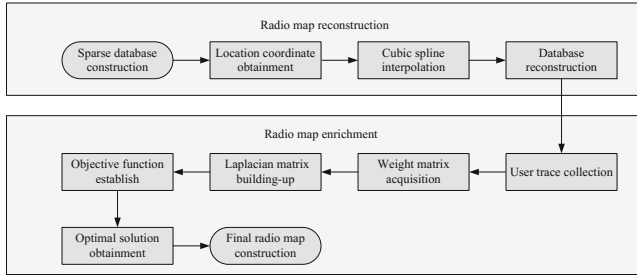


Fig. 1. Flow chart of the proposed approach.

the cubic spline interpolation and manifold learning algorithms to supplement the sparse fingerprint database in indoor WLAN localization. In addition, we propose to use the timestamp information during the process of intra-manifold graph construction in manifold learning, which can avoid the sharp deterioration of localization accuracy when the RSS changes abruptly.

3 Algorithm Description

3.1 Algorithm Overview

In this paper, we propose to use the cubic spline interpolation and manifold learning algorithms to construct the hybrid fingerprint database. In concrete terms, we rely on the cubic spline interpolation algorithm to enrich the sparse fingerprint database. After that, we apply the manifold learning algorithm to label the unlabeled trace locations based on the known RPs and the corresponding RSS sequences. The flow chart of the proposed approach is shown in Fig. 1.

3.2 Radio Map Reconstruction

To study the performance of the proposed approach, we carry out the experiments in a real indoor WLAN environment under different interpolation algorithms. Figure 2 shows the mean of errors with different radii of RPs used for the Radial Basis Functions (RBF), Linear, and Cubic Spline Interpolation (CUBIC) interpolation algorithms respectively. From this result, we can find that the CUBIC interpolation algorithm achieves the highest localization accuracy.

3.3 Radio Map Enrichment

Since the labeled and unlabeled data are collected in the same environment, the RSS measurements share the similar properties of the low-dimensional manifold [10]. Based on this, we label the unlabeled data by aligning their corresponding manifolds in the physical location space. In concrete terms, we build two intra-manifold graphs with respect to the labeled and unlabeled data respectively, as well as one inter-manifold graph between them. Then, we construct the weighted

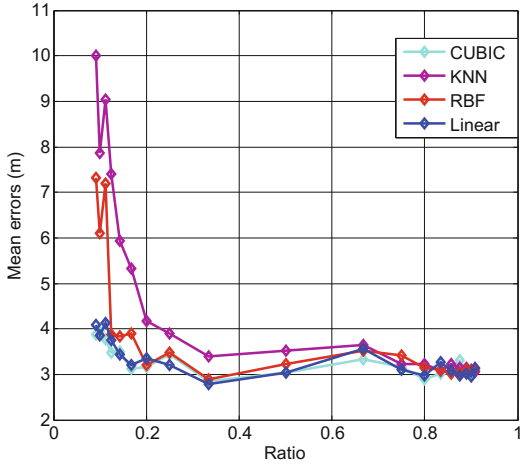


Fig. 2. Mean of errors under different interpolation algorithms.

graph matrices to describe the relations of RSS measurements, timestamps, and physical locations respectively, i.e., W_r , W_t , W_{loc} .

To build the graph matrix W_r , we simply connect each RSS vector with its K nearest neighbors, and then assign a value to each pair of vectors by using the heat Kernel [10], such that

$$W_r(i, j) = \begin{cases} e^{-\frac{\|x_i - x_j\|^2}{\theta_r}}, & \text{if } i \text{ and } j \text{ are connected} \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

where θ_r is the heat kernel of radio space.

The graph matrix W_t is built for the unlabeled traces by using the timestamps. Based on the assumption that the samples collected within short time duration are corresponding to the physically adjacent locations, we construct this matrix according to the time difference between every two samples. For the k -th trace, the W_t is defined as

$$W_t^{u_k}(i, j) = \begin{cases} e^{-\frac{|t_i - t_j|^2}{\theta_t}}, & |t_i - t_j| \leq T_{thr} \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

where θ_t and T_{thr} are the heat kernel and threshold of time space respectively.

We build W_{loc} only for the labeled data with known location information. We define that the two samples are connected when they are physically closest. Thus, W_{loc} is defined as

$$W_{loc}^l(i, j) = \begin{cases} e^{-\frac{dist(l_i, l_j)^2}{\theta_{dist}}}, & \text{if } i \text{ and } j \text{ are connected} \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

where θ_{dist} is the heat kernel of physical location space.

After that, we construct two intra-manifold graphs for the labeled and unlabeled data by using a relative weight $\alpha \in [0, 1]$, as shown in (4) and (5) respectively.

$$W^l = \alpha W_r^l + (1 - \alpha) W_{loc}^l \quad (4)$$

$$W^{u_k} = \alpha W_r^{u_k} + (1 - \alpha) W_t^{u_k} \quad (5)$$

Since the labeled and unlabeled data have the common feature space with RSS measurements in the inter-manifold graph, we link the labeled and unlabeled data by using the properties of RSS measurements.

Then, we can obtain the graph Laplacian matrices for the labeled data, unlabeled data, and the corresponding inter-dataset respectively as follows,

$$L^l_{N_l \times N_l} = D^l - W^l \quad (6)$$

$$L^{u_k}_{T_k \times T_k} = D^{u_k} - W^{u_k} \quad (7)$$

$$L^{lu_k}_{(N_l + T_k) \times (N_l + T_k)} = \begin{bmatrix} D^{lu_k} & -W^{lu_k} \\ -W^{lu_k T} & D^{u_k l} \end{bmatrix} \quad (8)$$

where D^l , D^{u_k} , D^{lu_k} , and $D^{u_k l}$ are the diagonal matrixes with the diagonal elements where D^l is a diagonal matrix with diagonal elements $D^l(i, i) = \sum_{j=1}^{N_l} W^l(i, j)$ ($i = 1, \dots, N_l$), D^{u_k} is a diagonal matrix with diagonal elements $D^{u_k}(i, i) = \sum_{j=1}^{T_k} W^{u_k}(i, j)$ ($i = 1, \dots, T_k$), D^{lu_k} is a diagonal matrix with diagonal elements $D^{lu_k}(i, i) = \sum_{j=1}^{T_k} W^{lu_k}(i, j)$ ($i = 1, \dots, N_l$), and $D^{u_k l}$ is a diagonal matrix with diagonal elements $D^{u_k l}(i, i) = \sum_{j=1}^{N_l} W^{lu_k}(j, i)$ ($i = 1, \dots, T_k$).

We continue to combine the intra-manifold graphs with inter-manifold graph by using a relative weight μ for manifold alignment. The composite graph Laplacian is described as

$$L_k = \begin{bmatrix} L^l & 0 \\ 0 & L^{u_k} \end{bmatrix} + \mu L^{lu_k} \quad (9)$$

We choose the two-dimensional physical space as the common low-dimensional latent space. By denoting the coordinate matrix for the new space of the labeled data and k-th trace as $q_k \in R^{(N_l + T_k) \times 2}$, we formulate the optional manifold alignment problem as

$$\hat{q}_k^{(h)} = \arg \min_{q_k^{(h)}} \left(q_k^{(h)} - Y_k^{(h)} \right)^T J_q \left(q_k^{(h)} - Y_k^{(h)} \right) + \gamma q_k^{(h)T} L_k q_k^{(h)}, h = 1, 2 \quad (10)$$

where

$$J_q = \begin{bmatrix} I_{N_l \times N_l} & 0 \\ 0 & 0 \end{bmatrix} \quad (11)$$

$A^{(h)}$ is the h-th column of matrix A, and $I_{N_l \times N_l}$ is the $N_l \times N_l$ identity matrix. In $Y_k \in R^{(N_l + T_k) \times 2}$, the previous N_l rows are the coordinates of the labeled data, i.e., $Y^l = [l_1, l_2, \dots, l_{N_l}]^T$, while the latter T_k rows are the arbitrary values. Based

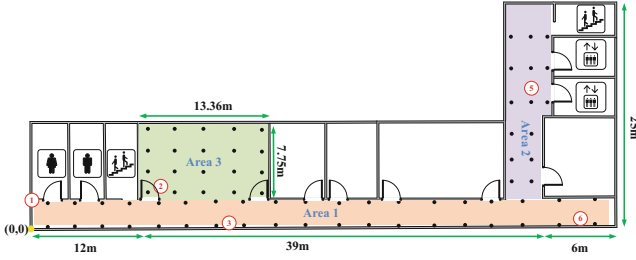


Fig. 3. Physical layout of target environment.

on the objective function in (10), the first term stands for the fitting error of the labeled data, while the second term enforces the smoothness of the manifold. The parameter γ controls the relative strength of location constraint.

Then, we calculate the objective function $q_k^{(h)}$ as

$$q_k^{(h)} = (J_q + \gamma L_k)^{-1} J_q Y_k^{(h)}, h = 1, 2 \quad (12)$$

Based on (12), we can assign location coordinates to the unlabeled data in the low-dimensional space. Finally, after the process of labeling the unlabeled data, we use the KNN to estimate the target locations.

4 Performance Evaluation

4.1 Experimental Setup

To investigate the performance of the proposed approach, we conduct the experiments in a real indoor WLAN environment with the size of 57 m by 25 m on the fifth floor of an office building, as shown in Fig. 3. The target area is covered by five APs. A Samsung S7568 mobile phone is selected as the receiver installed with our developed Wi-Fi localization software. The data are stored as the TXT files. We calibrate 73 RPs with the same interval of 3 m in three subareas, namely Area 1, 2, and 3. In addition, we record 30 traces without location information for the testing.

4.2 Parameters in Manifold Learning

In our approach, the parameters in manifold learning are significantly important and require to be carefully studied. Figure 4 shows the impact of different parameters in manifold learning on localization errors. In Fig. 4, we can find that as the values of θ_r , θ_t , α , and γ increase, the localization error decreases first and then slightly increases or approximately maintains the same. For the parameter μ , the localization error reaches the lowest when the value μ is greater than 3. Our approach achieves the best performance when the parameters $\theta_r = 50$, $\theta_t = 0.1$, $\alpha = 0.1$, $\gamma = 0.1$, and $\mu = 3$, and the number of neighbors used in manifold learning equals 4, i.e., $k = 4$. Since the variation of value θ_{dist} has no impact on localization error, we set $\theta_{dist} = 1$ in the results that follow.

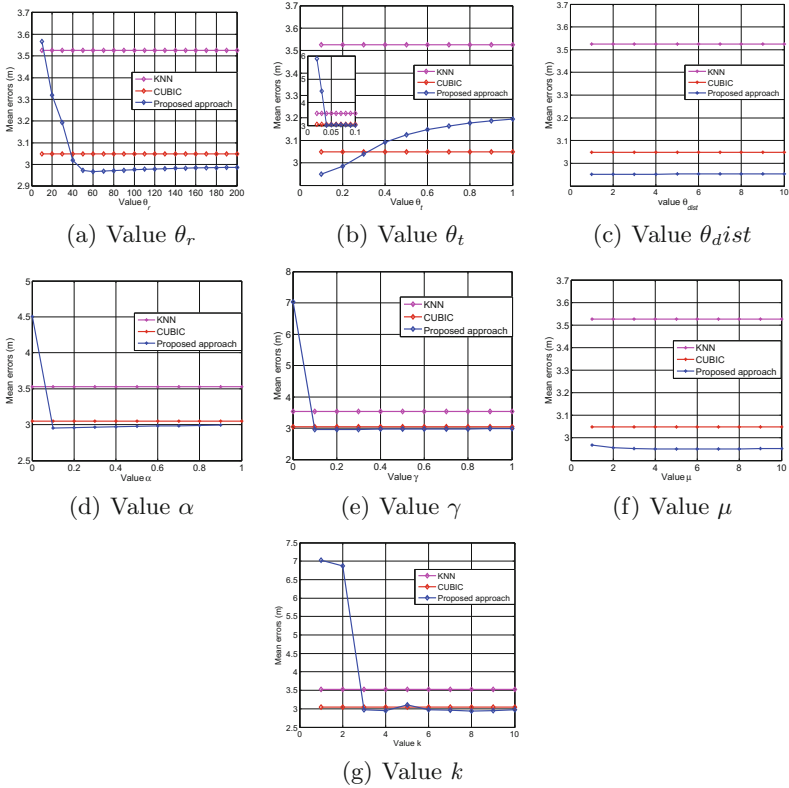


Fig. 4. Mean of errors under different parameters in manifold learning.

4.3 Localization Algorithms

We compare the Cumulative Distribution Functions (CDFs) of errors by the proposed approach and the conventional KNN, MA, and CUBIC approaches in Fig. 5. From this figure, we can find that our approach performs best in localization accuracy, and meanwhile it reduces the mean of errors by 16.4% compared with the result without our approach.

4.4 Number of RPs

Figure 6 compares the mean of errors under different ratios of the number of RPs by the proposed, KNN, MA, and CUBIC approaches. From this figure, we can find that in the small ratios condition, the increase of the number of RPs significantly reduces the localization error, whereas when the ratio is over 0.5, the variation of the number of RPs generally has slight impact on localization error.

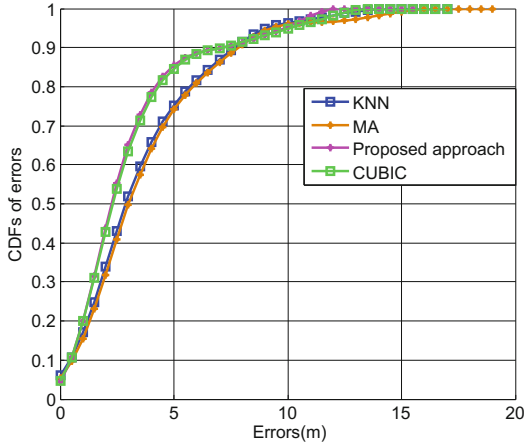


Fig. 5. CDFs of errors under different localization algorithms.

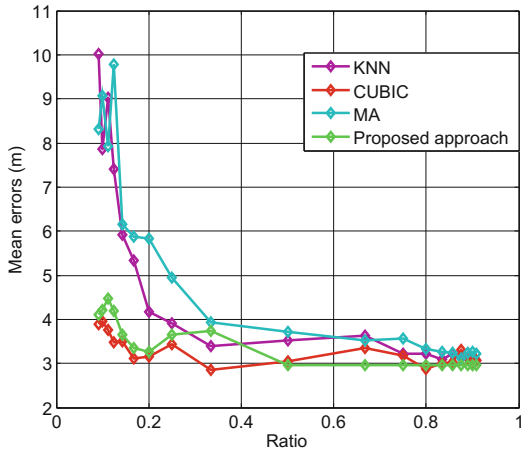


Fig. 6. CDFs of errors under different localization algorithms.

4.5 Number of Unlabeled Traces

Figure 7 compares the mean of errors under different number of unlabeled traces by the proposed, KNN, and CUBIC approaches. From this figure, we can find that the increase of the number of traces reduces the localization error by the proposed approach, and meanwhile when the number of traces is over 12, the number of trace has slight impact on localization error.

4.6 Impact of Timestamps

Figure 8 compares the CDFs of errors with and without timestamps by the proposed approach. From this figure, we can find that the localization performance

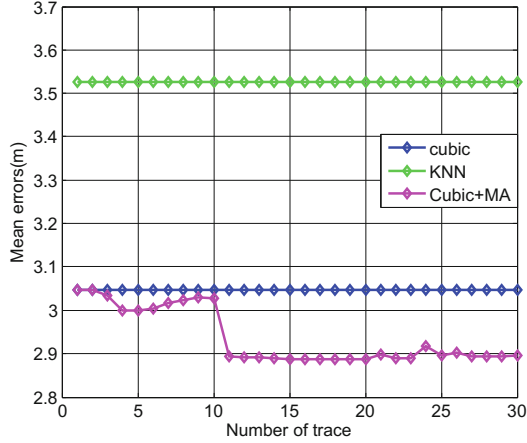


Fig. 7. Mean of errors under different number of traces.

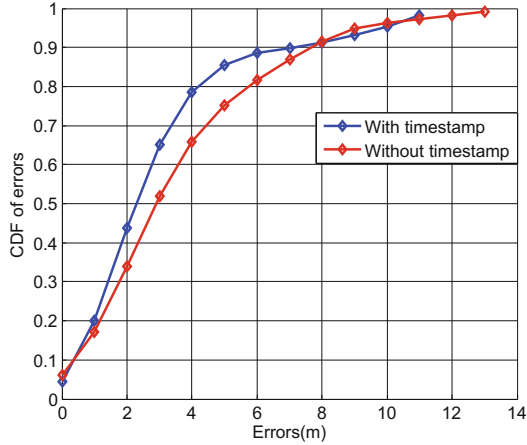


Fig. 8. CDFs of errors with and without timestamps.

is enhanced when the timestamp is considered. This result can be interpreted by the fact that the timestamp of traces is able to strengthen the correlation between the successively collected RSS measurements.

5 Conclusion

To reduce the labor effort involved in indoor WLAN fingerprint based localization, we propose a new integrated cubic spline interpolation approach with manifold learning from the low-overhead unlabeled traces of users in target environment. The experiments conducted in a real indoor WLAN environment demonstrate that our approach can not only reduce the density of RPs used for radio

map construction, but also improve the accuracy of indoor WLAN fingerprint based localization. In future, the application of this approach to a larger-scale or multi-floor indoor WLAN environment forms an interesting work.

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, cstc2015jcyjBX0065).

References

1. Axel, K.: Location-Based Services: Fundamentals and Operation, pp. 185–245. Wiley, New York (2005)
2. Want, R., Hopper, A., Falcao, V., Gibbons, J.: The active badge location system. *ACM Trans. Inf. Syst.* **10**(1), 91–102 (1992)
3. Hazas, M., Ward, A.: A novel broadband ultrasonic location system. In: Borriello, G., Holmquist, L.E. (eds.) *UbiComp 2002. LNCS*, vol. 2498, pp. 264–280. Springer, Heidelberg (2002). doi:[10.1007/3-540-45809-3_21](https://doi.org/10.1007/3-540-45809-3_21)
4. Darrell, T., Gordon, G., Harville, M., Woodfill, J.: Integrated person tracking using stereo, color, and pattern detection. *Int. J. Comput. Vis.* **37**(2), 175–185 (2000)
5. Bahl, P., Padmanabhan, V.N.: RADAR: an in-building RF-based user location and tracking system. In: *IEEE INFOCOM*, pp. 775–784 (2002)
6. Youssef, M., Agrawala, A.: The Horus WLAN location determination system. *Wirel. Netw.* **13**(3), 357–374 (2005)
7. Liu, S., Luo, H., Zou, S.: A low-cost and accurate indoor localization algorithm using label propagation based semi-supervised learning. In: *International Conference on Mobile Ad-hoc and Sensor Networks*, pp. 108–111 (2009)
8. Ouyang, R., Wong, A., Lea, C., Chiang, M.: Indoor location estimation with reduced calibration exploiting unlabeled data via hybrid generative/discriminative learning. *IEEE Trans. Mob. Comput.* **11**(11), 1613–1626 (2011)
9. Pan, J.J., Pan, S.J., Yin, J., Ni, L.M., Yang, Q.: Tracking mobile users in wireless networks via semi-supervised colocalization. *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(3), 587–600 (2012)
10. Mikhail, B., Partha, N.: Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.* **15**(6), 1373–1396 (2003)