IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 16, 2019, accepted July 24, 2019, date of publication July 30, 2019, date of current version August 14, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2932024

New Reconstructed Database for Cost Reduction in Indoor Fingerprinting Localization

VAHIDEH MOGHTADAIEE[1], SEYED ALI GHORASHI[1,2], (Senior Member, IEEE), AND MOHAMMAD GHAVAMI[3], (Senior Member, IEEE)

[1]Cyberspace Research Institute, Shahid Beheshti University, Tehran 19839 69411, Iran
[2]Cognitive Telecommunication Research Group, Department of Electrical Engineering, Shahid Beheshti University, Tehran 19839 69411, Iran
[3]Electrical and Electronic Engineering Department, London South Bank University, London SE1 0AA, U.K.

Corresponding author: Seyed Ali Ghorashi (a_ghorashi@sbu.ac.ir)

This work was supported in part by the Iran National Science Foundation (INSF), and in part by the Shahid Beheshti University under Grant 94027558.

ABSTRACT Location fingerprinting is a technique widely suggested for challenging indoor positioning. Despite the significant benefits of this technique, it needs a considerable amount of time and energy to measure the Received Signal Strength (RSS) at Reference Points (RPs) and build a fingerprinting database to achieve an appropriate localization accuracy. Reducing the number of RPs can reduce this cost, but it noticeably degrades the accuracy of positioning. In order to alleviate this problem, this paper takes the interior architecture of the indoor area and signal propagation effects into account and proposes two novel recovery methods for creating the reconstructed database instead of the measured one. They only need a few numbers of RPs to reconstruct the database and even are able to produce a denser database. The first method is a new zone-based path-loss propagation model which employs fingerprints of different zones separately and the second one is a new interpolation method, zone-based Weighted Ring-based (WRB). The proposed methods are compared with the conventional path-loss model and six interpolation functions. Two different test environments along with a benchmarking testbed, and various RPs configurations are also utilized to verify the proposed recovery methods, based on the reconstruction errors and the localization accuracies they provide. The results indicate that by taking only 11% of the initial RPs, the new zone-based path-loss model decreases the localization error up to 26% compared to the conventional path-loss model and the proposed zone-based WRB method outperforms all the other interpolation methods and improves the accuracy by 40%.

INDEX TERMS Cost reduction, indoor localization, interpolation, location fingerprinting, path-loss propagation model, reconstructed database.

I. INTRODUCTION

With increasing user demands on Location-based Services (LBS) and Social Networking Services (SNS), indoor positioning has become more crucial. Satellite positioning such as Global Positioning System (GPS) supports outdoor localization well, due to the appropriate power level and good visibility of satellites. However, because of the general failure of these technologies in indoor environment, non-satellite-based technologies are vastly proposed for indoor localization [1], [2]. Wireless Local Area Networks (WLAN) have widely been employed for indoor localization because of two main reasons: the wireless networks are widely available in indoor areas, and almost all mobile devices are equipped with Wi-Fi receivers, which make positioning feasible for the users [3].

Location fingerprinting technique is one of the most suggested methods for indoor positioning [4]. The main advantage of location fingerprinting is its capability of alleviating the multipath and Non-Line-of-Sight (NLOS) propagations problems in indoor environments. Furthermore, it needs no additional infrastructure hardware as Wi-Fi access points (APs) are already deployed indoors, and the Received Signal Strength (RSS) values are easily accessible from the Application Programming Interface (API) of mobile devices. It has two stages: *training* and *positioning*. It stores the location-dependent characteristics of a signal collected at Reference Points (RPs) in a database in the training stage, and

The associate editor coordinating the review of this manuscript and approving it for publication was Marina Barbiroli.

in the positioning stage, data mining methods are applied to approximate the position of the user based on the fingerprint of the user and the database. In general, the precision of location estimation depends on the density of RPs, accuracy of measured RSSs and the estimating algorithm [5], [6].

Despite the aforementioned advantages, there are some drawbacks for fingerprinting method. The initial deploying of the training stage in the area of operation and collecting many measurements at different locations to construct the database are so costly in terms of required time, equipments, and human efforts. The cost varies based on the size of the environment, the number of RPs, and the number of measurements per location. In addition, it is necessary to rebuild the database in case of modifications or changes in the environment. Furthermore, some parts of the area may be inaccessible or restricted in practical situations, preventing the surveyors to construct an appropriate fingerprint database for the entire area [7]. It is also suggested to record the fingerprints for different body orientations so that the effect of orientation remains minimal [4]. All these limitations of fingerprinting technology make the training stage costly, which result in failing the fingerprinting procedure when the rapid deployment is required.

One of the main challenges of fingerprinting localization is how to reduce the training cost [8], [9]. Various methods have been suggested for constructing the database by omitting some of the RPs to save the time/manpower consumptions. Despite the fact that reducing the number of RPs greatly decreases the surveying cost, it dramatically degrades the positioning accuracy as well. Therefore, it is of interest to find a method which could lessen the accuracy degradation along with the cost reduction.

This paper focuses on taking the interior architecture and environmental effects of the indoor area into account for the purpose of taking back the lost localization accuracy when the number of RPs is highly reduced in the low-cost indoor location fingerprinting. Therefore, a new zone-based model is proposed for RSS reconstruction, which deals with different zones separately. Due to the signal propagations and attenuations in indoor areas, the zones are defined here by the interior architecture of the buildings, such as rooms or corridors divided by walls, glasses, etc. We apply this zone-based model on the conventional path-loss model and six interpolation methods including Nearest Neighbor (NN), Linear, Natural, Cubic spline, Inverse Distance Weighting (IDW), and Radial Basis Function (RBF) to recover the RSS values at the query points. According to the propagation model around an AP, a new zone-based Weighted Ring-based (WRB) interpolation method is also suggested here to improve the accuracy. The reconstructed database can also help the fingerprinting network designers have a bright insight into the network error merely by measuring RSS at a few numbers of RPs with the minimum effort. Based on the application and a desirable network performance for that, they can choose the necessary numbers of RPs and APs as they affect the localization accuracy [5], and make a trade-off between the accuracy and the needed time/energy costs. For location estimation techniques, we have adopted a well-known deterministic algorithm, Weighted-K-Nearest-Neighbor (WKNN) method [4].

The major contributions of the paper are listed as follows:

- A new zone-based path-loss model is proposed, which outperforms the conventional path-loss model and provides more accurate recovered RSS values and less positioning error.
- A novel zone-based Weighted Ring-based (WRB) interpolation method is introduced that can decrease the indoor localization error compared to the other renowned interpolation methods.
- A comparative analysis in two groups of recovery methods is fully provided including conventional and suggested zone-based path-loss and six interpolation methods.
- Extensive experiments in three real-world indoor WLAN environments are carried out to evaluate the performance of our proposed methods in terms of reconstruction and localization accuracies.

The remainder of the paper is organized as follows. Section II highlights the related studies on reducing the fingerprinting surveying effort and the novelties of this paper. In Section III, the indoor location fingerprinting system model and notations are determined and two stages of that are described. Section IV, introduces the proposed methods for recovering the fingerprinting database in details. The experimental test environments are presented in Section V and the evaluation metrics and parameter tuning are discussed there. Section VI is devoted to the performance evaluation. The reconstruction errors and localization errors are also fully assessed and compared when various methods are used for generating the recovered database. Different configurations of RPs, computational complexities, and device diversity are also investigated and discussed there. In addition, the proposed methods are evaluated in a benchmarking testbed at the end of Section VI. Finally, the conclusions and the future work are given in Section VII.

II. RELATED WORK

There are a lot of researches have been done to reduce the training stage cost with various techniques. These studies can be mainly divided into three categories. In the first group, no one is specifically employed to survey the fingerprinting area. Authors in [10] introduced Simultaneous Localization and Mapping (SLAM) in which an equipped autonomous mobile robot was utilized to create a database of an unknown environment by collecting RSS measurements at several locations. This technique decreases the labor work, but needs additional hardwares. There might also be some accessibility limitations or security regulations for the mobile robot during the data collection process [11]. Furthermore, due to the widespread use of mobile devices, crowdsourcing methods [12] can also construct a fingerprint database using the collected crowdsourced Wi-Fi RSS data from several

users. However, designing a suitable incentive mechanism in crowdsourcing-based indoor localization systems and lack of quality control of crowdsourced data are still two main challenges of this technology [13].

The second group of techniques can generate the database analytically without any site survey. For instance, the ray tracing-based methods construct the fingerprinting database analytically, in which the knowledge about the precise building structure and material details of walls and roofs are necessary [14]. The radiosity model was originally adopted in physics to describe the heat transfer in bodies with different temperatures [15].

The last group needs to have RSS measurements at some known RPs, so that they could generate the RSS at more unknown points and reconstruct the database. There are many reconstruction methods employed for this purpose, such as compressive sensing (CS) [16] theory, randomized least absolute shrinkage and selection operator (LASSO) [17], and Gaussian process (GP) regression [11]. Path-loss model and interpolation-based methods are also adopted for database reconstruction purposes using some RPs.

The main focus of this paper is on the last group when using path-loss model and interpolation-based methods. Utilizing the path-loss model for the purpose of reducing the database construction effort is discussed in [6], [9], [18], [19]. They used a log-distance path-loss model to generate the RSS of unknown locations to build the fingerprinting database. For this purpose, the actual measurements at formerly surveyed RPs are utilized to train the parameters of the path-loss model. However, due to the complexity of indoor environments, propagation models cannot precisely predict the signal fading and multipath effect, which causes lower positioning accuracy in comparison with the fingerprinting measurements [11], [20]. In order to improve the accuracy, authors in [6] tried to decrease the positioning error by zone-based remedy algorithm which divided the area into four zones and approximated the position of a device by finding the zone with maximum overall probability. In [9] also the path-loss exponent is considered as an additional fingerprint factor along with the RSS value for each RP to improve the quality of the fingerprints in the resource-limited environments. In addition, authors in [18] recommended a multi-antenna system to provide different path-loss parameters at each RP and achieved almost the same localization accuracy when 33% of surveyed RPs were decreased. This technique, however, requires deploying additional hardware in the operation area. Also the use of more advanced path-loss model with wall attenuations were proposed in [19], in which the number of walls between APs and RPs needed to be calculated, making that complicated in many indoor areas.

Inter/extrapolation approaches also can practically decrease the number of surveyed RPs [3], [21]. They have been widely used in the image processing field as well. Adding more RPs to the fingerprint database with the interpolation methods reduces the cost and enhances the localization accuracy [22]. Several interpolation methods such as the nearest neighbor (NN), linear [3], cubic spline [8], Inverse Distance Weighting (IDW) [23], Radial Basis Function (RBF) [24], and Kriging [25] have been applied to generate the fingerprinting database. A new integrated cubic spline interpolation approach with manifold learning is proposed in [21] to enhance the accuracy. The authors in [26] used linear Delaunay Triangulation interpolation only, whereas authors in [3] implemented the NN, linear, IDW algorithms and extrapolation methods to recover the RSS at the query points. In [27], an adaptive smoothing algorithm is utilized to tackle the discontinuity of RSS due to the presence of walls and the complete fingerprint database is then produced using IDW method. Also K. Arai et al. recommended merging the measurements from two adjacent RPs and assigning that to a RP in the middle of them to create a dense fingerprinting map [28].

The difference of our proposed zone-based model is that it considers the structural and environmental effects of the indoor area for fingerprinting localization without any knowledge about the material of walls and floors. When using path-loss model, it is common to either adopt single fixed path-loss exponent for all APs to describe the behavior of all the paths from all APs to all RPs [24], [29], [30] or different path-loss exponents for different APs over the entire indoor area [6], [31]–[33]. Both approaches can produce an error in the RSS estimation and decrease the localization accuracy, particularly over a large area with several rooms. In proposed zone-based path-loss model, however, the path-loss parameters are different for each zone. Also when using the interpolation, despite the fact that it can significantly decrease the fingerprinting cost, the related works did not take the interior architecture influences of indoor areas into account and still are inefficient in terms of accuracy [34]. Therefore, we propose an zone-based interpolation model that adapts more to the characteristics of a fingerprinting network in indoor environments, which are severely affected by the walls and the distance of the RPs from the Wi-Fi APs.

III. INDOOR LOCATION FINGERPRINTING

In this section, first the notation of all elements of the fingerprinting network in this paper is defined and then two stages of fingerprinting are briefly explained.

A. SYSTEM MODEL AND NOTATIONS

It is assumed that there are A APs at locations $\{\phi_1, \phi_2, ..., \phi_i, ..., \phi_A\}$ and total N RPs with known locations, $\{\theta_1, \theta_2, ..., \theta_n, ..., \theta_N\}$, in a 2D area, where $\phi_i = [x_i^{AP}, y_i^{AP}]^T$ and $\theta_n = [x_n^{RP}, y_n^{RP}]^T$. There are T collected fingerprinting measurements of all A APs averaged and saved in a database for all RPs.

In order to reduce the surveying cost, we only keep few number of RPs for training and reconstruct the RSS of other RPs by the methods presented in the next section. We call these sets of RPs as the *main RPs (MRPs)* and *query RPs (QRPs)* respectively throughout this paper. The number of MRPs and QRPs are denoted by M and Q, where $N = M + Q$.

TABLE 1. List of notations.

Notation	Description
AP, A	Access Points, the number of APs
RP, N	Reference Points, the number of RPs
TP, P	Test Points, the number of TPs
MRP, M	Main RPs, the number of MRPs
QRP, Q	Query RPs, the number of QRPs
Y, $\hat{Y}$	Measured and reconstructed fingerprinting databases
T	Number of samples per location
S_n^{RP}, $\hat{S}_n^{RP}$	Measured and reconstructed RSS vectors of n^{th} RP
S_p^{TP}	Measured RSS vector of p^{th} TP
$s_{n,i}$, $\hat{s}_{n,i}$	Measured and reconstructed RSS from i^{th} AP at n^{th} RP
ϕ_i	Location of i^{th} AP
θ_n	Location of n^{th} RP
θ_p^{TP}, $\hat{\theta}_p^{TP}$	Real and estimated locations of p^{th} TP
Z	Number of zones
M_z	Number of MRPs in z^{th} zone
Q_z	Number of QRPs in z^{th} zone
M_{zr}	Number of MRPs in z^{th} zone and r^{th} ring
Q_{zr}	Number of QRPs in z^{th} zone and r^{th} ring
r_{max}	Number of rings for an AP

The measured fingerprint database (Y) and the recovered one ($\hat{Y}$) associated to all N RPs is shown below. Both databases have the same size of $N \times (A+2)$ and differ only in the RSS of the QRPs.

$$Y = \begin{bmatrix} \theta^{MRP}, S^{MRP} \\ \theta^{QRP}, S^{QRP} \end{bmatrix}, \tag{1a}$$

$$\hat{Y} = \begin{bmatrix} \theta^{MRP}, S^{MRP} \\ \theta^{QRP}, \hat{S}^{QRP} \end{bmatrix}, \tag{1b}$$

where θ^{MRP} and θ^{QRP} are the x,y locations of MRPs and QRPs. The matrices S^{MRP} and S^{QRP} include the measured RSS vectors of all MRPs and QRPs and $\hat{S}^{QRP}$ consists of reconstructed fingerprint vectors of all QRPs. To general refer to the fingerprint vector of n^{th} RP, we use $S_n^{RP} = [s_{n,1}, s_{n,2}, ..., s_{n,i}, ..., s_{n,A}]$ in this paper where $n \in MRP$ or $n \in QRP$.

For the localization accuracy calculation of an indoor area, there are assumed P number of Test Points (TPs) randomly distributed over the whole area, where the localization stage of fingerprinting is carried out using the created databases discussed above. The positions of TPs are assumed to be unknown and estimated in the localization process, so that we could have an approximation about the localization error of measured/reconstructed databases. The real and estimated position of the pth TP is considered as θ_p^{TP} and $\hat{\theta}_p^{TP}$, respectively and the fingerprint vector of that TP from all APs is shown by S_p^{TP}. The complete list of notations is shown in Table 1.

B. TRAINING STAGE

The first stage of fingerprinting is to establish a database for all N RPs in 2D space, which will be used as training samples in the positioning stage. The average of all measurements from each AP is calculated, and is logged along with the corresponding coordinates in the database. For the purpose of lowering the surveying cost, the surveying is carried out only on a limited numbers of RPs and the fingerprints of many other RPs are predicted and added to the database. In the real world, no measured database can cover all areas of interest even when the number of RPs is high. Therefore, utilizing RSS prediction for RPs not only saves the cost of training but also helps to cover the area more appropriately.

C. POSITIONING STAGE

In the positioning stage or online localization stage, the location of a TP is estimated employing the measured RSS fingerprints from all APs at that point. For location estimation and positioning results comparison, the deterministic approach is exploited here as it is the most commonly used localization algorithm. The location estimation is carried out by comparing the fingerprints of TP with the reconstructed database using Euclidean distance [35]. The Euclidean distance between TP and n^{th} RP is calculated by

$$D_n = \|S^{TP} - S_n^{RP}\|. \tag{2}$$

If D_n is the minimum Euclidean distance between TP and all RPs, the n^{th} RP is called the nearest point to TP. In WKNN method, the weighted average of the coordinates of these K nearest points are calculated. The weighting values are the inverse of the Euclidean distance. The estimated location of TP by WKNN algorithm is

$$\hat{\theta}_{WKNN} = \frac{\sum_{n=1}^{K} \frac{1}{D_n} \theta_n}{\sum_{n=1}^{K} \frac{1}{D_n}}. \tag{3}$$

IV. PROPOSED RECONSTRUCTED FINGERPRINTING DATABASE

In this section, two proposed methods to reconstruct the fingerprinting database are explained.

A. A ZONE-BASED PATH-LOSS MODEL

To recover the RSS values at different points in an indoor environment for completion of the fingerprinting database, a commonly used large scale model which is the log-distance path-loss model can be used [36]. To simulate the strength of a propagated signal in indoor environments more accurately, including the impact of obstacles is also necessary. The RSS values from the i^{th} AP are then obtained as follows [37]

$$\hat{s}_{n,i} = s_{d_0 i} - 10\alpha_i \log\left(\frac{d_{n,i}}{d_0}\right) - \sum_w L_w - \sum_f L_f + X_{\sigma_i}, \tag{4}$$

where $\hat{s}_{n,i}$ (dBm) is the power received from the i^{th} AP at the n^{th} RP with the $d_{n,i}$ physical distance to that AP in meter. $s_{d_0 i}$ (dBm) is the power received at d_0-meter (usually 1m) from the i^{th} AP and α_i is the path-loss exponent. There are constant attenuations if the signal passes through walls and floors. L_w and L_f are the attenuation constants in dB for the w^{th} and f^{th} wall and floor between the AP and the point at distance $d_{n,i}$. $X_{\sigma_i} \sim \mathcal{N}(0, \sigma_i)$ models the path-loss variation at one point due to the shadowing caused by obstacles in

propagation and assumed to be a zero mean Gaussian random variable with a standard deviation given by σ_i (dB). In order to employ the path-loss model, the positions of APs should be known. There are researches presented some techniques to estimate the APs' locations in case they are unknown [32].

A common way of training the path-loss parameters is to use the Least Squares (LS) technique on the RSS data collected at RPs across the whole area. However, the RSS measurements vary in different regions based on the distance from APs and the paths and environments through which the signal passes [38]. Therefore, taking all MRPs of the entire area and calculating a single path-loss exponent to estimate the RSS of all QRPs definitely degrades the quality of the predicted RSS and the localization accuracy.

The signal propagation in indoor areas is complex and depends on various environmental parameters such as multipath, refraction, blockage, and noise. In addition, according to Tobler's first law of geography [39], each value is related to all other values on a geographic surface, but near values are more related than distant ones [7]. Therefore, here a new zone-based model is proposed for the RSS reconstruction at QRPs, which supposes different zones for an area of interest regarding the interior structure and signal propagations and attenuations there. The zones are separated by walls, glasses, etc. This model considers that all the points inside one zone are more related to each other, since not only they are close but also they are placed in almost the same situation from an AP regarding LOS/NLOS radio propagations. In the proposed zone-based path-loss model, the path-loss parameters are calculated separately for each zone using only the MRPs inside that zone. As a result, each zone is assigned with its own path-loss parameters.

If we suppose Z number of zones for an indoor area, $\{\zeta_1, \zeta_2, ..., \zeta_z, ..., \zeta_Z\}$, we re-write (4) and consider the zone-based path-loss model for a RP in z^{th} zone as follows

$$\hat{s}_{n,i} = A_{i,z} + 10\alpha_{i,z} \log\left(\frac{d_{n,i}}{d_0}\right) + X_{\sigma_{i,z}}, \tag{5}$$

where $A_{i,z}$ (dBm) is the sum of three terms of (4): $s_{d_0 i}$, $\sum_w L_w$ and $\sum_f L_f$. In contrast to (4), here it is not required to have a prior knowledge on the transmitter power, walls/floors materials and attenuations, and calculations of the number of walls/floors by examining several paths from an AP to a receiver. For generating the recovered database, all $A_{i,z}$, $\alpha_{i,z}$, and $\sigma_{i,z}$, which we consider as path-loss parameters here, have to be computed for each zones based on each AP. The values of $A_{i,z}$ and $\alpha_{i,z}$ for i^{th} AP in z^{th} zone are trained by the LS method using the measured RSS of MRPs inside ζ_z. Knowing $A_{i,z}$ and $\alpha_{i,z}$, the value of $\sigma_{i,z}$ can be also obtained utilizing the measured values and the predicted ones as below

$$\sigma_{i,z} = \sqrt{\frac{1}{M_z}\sum_{n=1}^{M_z}(S_n^{RP} - \hat{S}_n^{RP})^2}, \tag{6}$$

where M_z is the total number of MRPs in the zone ζ_z. We can expect that the zone-based path-loss parameters result in a more accurate reconstructed database and can eventually decrease the positioning error. Hence, we need to apply (5) and (6) for all the zones using their own MRPs to gain the specific path-loss parameters of the zones and then calculate the RSS values at QRPs. To the best of our knowledge, this is the first time the path-loss parameters are treated this way.

The suggested zone-based path-loss model can be also used in distance-based applications in which the distance between a point and an AP is of interest based on the measured RSS at that point. If we assume that the n^{th} RP is located inside the z^{th} zone, the estimated distance to the i^{th} AP is calculated in meters based on (5) as below

$$\hat{d}_{n,i} = 10^{\frac{s_{n,i} - A_{i,z}}{10\alpha_{i,z}}}. \tag{7}$$

Here, $X_{\sigma_{i,z}}$ can be ignored since the average RSS value for RPs, $s_{n,i}$, is used. The distance estimation error in meters can be then calculated as follows

$$e_d = |\hat{d}_{n,i} - d_{n,i}|. \tag{8}$$

Note that in the conventional path-loss model, the whole indoor area is treated as one zone. Therefore, z in (5) to (8) refers to the whole environment and can be ignored.

B. ZONE-BASED INTERPOLATION METHODS

Interpolation is a mathematical technique that exploits the function values at known points to predict the function values at the query points. Using interpolation techniques in location fingerprinting produces a denser database even by measuring the RSS values at a few number of RPs [17]. If the points are not placed in a regular or uniform grid, the interpolation called scattered data interpolation. Here we consider scattered data interpolation techniques so that we could treat the fingerprinting localization in a more general way.

Various interpolation methods usually produce different values at query points. Therefore, in location fingerprinting technique, we need to apply them depending on the required accuracy and complexity [22]. Before applying interpolation methods, we need to understand their behavior for estimating the RSS values when they are employed in fingerprinting networks with various RPs and APs configurations. The parameter fitting processes should be also carried out to find the best behavior of a interpolation method with respect to the desired fingerprinting network.

Six well-known interpolation functions are used and compared here: NN, linear, natural, cubic spline, IDW, and RBF. In interpolation, a convex hull is called to the smallest convex polygon containing all the given points. For some interpolation algorithms, such as linear, natural and cubic spline, the QRPs need to be inside the created convex hull by MRPs. Hence, they provide a limit for choosing the location of MRPs. However, for other three methods it does not differ that the QRPs are inside or outside of the convex hull.

According to the mentioned Tobler's first law of geography and similar to what we have discussed for the zone-based path-loss model, here we also propose to apply zone-based

method to the interpolation methods. Some interpolation functions are intrinsically local such as the ones use a Delaunay triangulation [40] in which an interpolation is employed within each triangle. However, the proposed zone-based model is inspired by the characteristics of a fingerprinting network in indoor environments that are severely affected by the walls and the distance of the RPs from the Wi-Fi APs. Based on that, a new interpolation method is also proposed in the next sub-section.

1) PROPOSED WEIGHTED RING-BASED (WRB) INTERPOLATION

The main idea behind this proposed method is that the points at the same distance from an AP are supposed to receive the same amount of power due to the APs' isotropic propagation patterns. Therefore, this method considers some virtual rings around the AP similar to Fig. 1 and supposes the same RSS for all the points inside each ring. All the rings have the same width that should be tuned first. The WRB method includes three steps explained below.

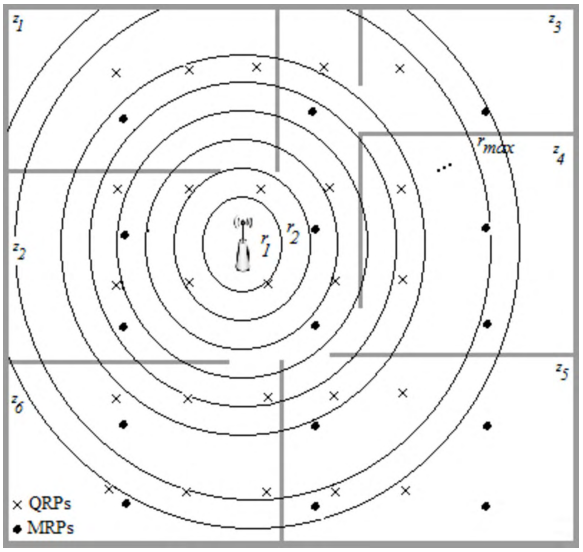


FIGURE 1. WRB interpolation concept. Showing an AP and the virtual rings around that in a typical indoor environment, includng six zones, MRPs, and QRPs.

Step 1: The algorithm starts from the first ring (r_1) near the AP to the furtherest ring (r_{max}) and finds all the MRPs and QRPs inside each ring. If there is no QRPs inside a ring, we move to the next ring.

Step 2: If there are Q_r QRPs and M_r MRPs inside ring r_i, the RSSs of all QRPs in that ring are considered the same and equal the mean RSS of those M_r MRPs as follows:

$$\hat{S}^{RP}_{q,r_i} = \frac{1}{M_r}\sum_{m=1}^{M_r} S^{RP}_{m,r_i},\ q = 1, 2, ..., Q_r, \tag{9}$$

where S^{RP}_{m,r_i} is the measured RSS of m^{th} MRP inside ring r_i.

Step 3: If there are Q_r QRPs but no MRPs inside ring r_i, we need to employ MRPs in upper and lower rings. For this purpose, we find the two upper and two lower rings such that all of which include at least one MRP. These upper and lower rings are denoted by r_{u_1}, r_{u_2} and r_{l_1}, r_{l_2}, respectively. We also assign a weight to the RSS of MRPs in each of rings, r_{u_1}, r_{u_2}, r_{l_1}, and r_{l_2}, according to the inverse of their distances to the current ring, r_i. The corresponding weights then are $w_{u_1} = \frac{1}{|r_i - r_{u_1}|}$, $w_{u_2} = \frac{1}{|r_i - r_{u_2}|}$, $w_{l_1} = \frac{1}{|r_i - r_{l_1}|}$, and $w_{l_2} = \frac{1}{|r_i - r_{l_2}|}$. Therefore, the RSS of QRPs inside r_i are calculated as below

$$\hat{S}^{RP}_{q,r_i} = \frac{\sum_{i=1}^{2} w_{u_i}\bar{S}^{RP}_{m,r_{u_i}} + \sum_{j=1}^{2} w_{l_j}\bar{S}^{RP}_{m,r_{l_j}}}{\sum_{i=1}^{2} w_{u_i} + \sum_{j=1}^{2} w_{l_j}},\quad q = 1, 2, ..., Q_r, \tag{10}$$

where $\bar{S}^{RP}_{m,r_{u_i}}$ and $\bar{S}^{RP}_{m,r_{l_j}}$ are the mean RSS of MRPs inside upper and lower rings and are computed similar to (9). Therefore, in the WRB approach, we basically deal with the virtual rings around an AP, *separately*. In other words, the QRPs inside each ring are affected only with MRPs inside that ring or with MRPs inside the two nearest rings in case no MRPs are located in that ring.

Similar to what we mentioned for the proposed zone-based model, the zone-based version of the WRB interpolation method also employs only the MRPs inside each zone for predicting the RSS values of QRPs in that zone. So the RSS values of the QRPs are calculated based on the RSS of the most related MRPs in the same ring of the same zone. That is why we expect the zone-based WRB to show the low positioning error. The whole procedure of zone-based WRB is described in Algorithm 1.

Algorithm 1 Calculate the RSS of QRPs From One AP in the Zone-Based WRB Method

Input: The RSS values and locations of MRPs in all zones
Output: The RSS values of QRPs in all zones

```
for z = 1, 2, .., Z do
  for r = r_1, r_2, .., r_max do
    if any QRPs of z^th zone are located in ring r (Q_zr ≠ 0) then
      if any MRPs of z^th zone are located in ring r (M_zr ≠ 0) then
        Run (9)
      else
        Find r_u1, r_u2, r_l1, r_l2 inside z^th zone
        Compute w_u1, w_u2, w_l1, w_l2
        Run (10)
      end if
    end if
  end for
end for
```

V. EXPERIMENTAL SETUP

The accuracy of the mentioned path-loss and interpolation methods can be evaluated only if the measured RSS data are available at QRPs. Therefore, two experimental tests are carried out to record the fingerprints of all MRPs and QRPs

so that we can analyze the RSS data that presented methods could generate and the accuracy they provide. In order to compare all the methods, MRPs are chosen such that all the QRPs could be placed inside the convex hull and the Delaunay triangulation could be done for the linear, natural and cubic spline interpolation methods. Therefore, we choose RPs at the corners of each zone as MRPs, so all QRPs are inside the convex-hull created by those MRPs and the triangulation-based interpolation techniques can be also supported. Other MRPs' configurations will be discussed in sub-section VI-C.

A. EXPERIMENTAL TESTBEDS

The RSS fingerprint samples are recorded by a Samsung Galaxy SM-J500H Smartphone with the Android version of 6.0.1. We have written and developed an Android code for data acquisition from existing APs, in which RSS values and Media Access Control (MAC) addresses of sensed APs are simultaneously sent to the server and recorded there to build the database.

Testbed 1: The first experimental testbed is located on the 2nd floor in Cyberspace Research Institute at Shahid Beheshti University. It has dimensions of 51 m by 18 m in x and y directions. Fig. 2a illustrates the layout of this experiment which is a typical indoor office environment consisting of fifteen zones, including eleven rooms and a four-section corridor. Here the rooms or room-sized sections are considered as the zones. In this experiment, there are 354 RPs and 30 TPs. The RPs are distributed as evenly as possible. There are 9 APs sensed when the Wi-Fi data are collected. At each RP the user faces south first and records 100 RSS vectors from all the sensed APs in 10 s. (Note that we can sample in the 1 or 1.5 s intervals to avoid getting RSS replicates). The orientation is then changed to the west, north, and east. The total number of the temporal samples logged per location from each AP is 400 in four directions within 40 seconds. The MRPs, QRPs, TPs, and APs distributions are also shown in Fig. 2a. Two of the APs, AP_8 and AP_9, are located downstairs. Surveying this area took approximately 15 hours and a lot of energy from the surveyor. If some of the APs are not sensed at one RP, the corresponding RSS value is assigned to -110 dBm to show its unavailability.

Testbed 1- RP removal: To decrease the surveying cost by employing the proposed methods, we deliberately remove 75.7% of RPs (268 points) from the database and consider them as QRPs, RSS of which need to be predicted employing the path-loss and interpolation methods. The remained 86 MRPs are used to estimate the RSS of the QRPs. By doing so, the 75.7% of the needful surveying time and energy are actually reduced, and the area can be surveyed in less than 4 hours instead of 15 hours. The location of TPs are the same as before so that we could have a fair comparison.

Testbed 2: The second experimental testbed shown in Fig. 2b is the former experiment in an indoor office area on the 4nd floor of Electrical Engineering and Telecommunications building at University of New South wales [5], [41]. The device used for data acquisition was a Compaq iPAQ 3970. The dimensions are 23 m by 11 m in x and y directions, including seven zones. There are 130 RPs, 28 TPs and 5 APs in this experiment. The data is collected in all four orientations with total 12 samples at each RP.

Testbed 2- RP removal: In this experiment, we remove 68% of RPs from the fingerprinting database, so the numbers of MRPs and QRPs are 42 and 88, respectively. The location of TPs are also the same as before.

Because the number of zones, the total number of RPs, and the structure of indoor environments are distinct in testbed 1 and 2, so the number of MRPs/QRPs will be also different in both testbeds, which results in different percentages of removed RPs. The same method and equipment are used for the data acquisition at TPs in the positioning stage. However, users usually have different types of equipments in the real situations than what has been formerly used in the training stage. Two different devices are employed in this paper in these two testbeds. The effect of different devices and handling device diversity is discussed later in sub-section VI-E.

B. EVALUATION METRICS

1) RECONSTRUCTION ERROR

In order to assess the reconstructed RSS values obtained by the methods presented in the previous section, the Mean Absolute Error (MAE) at the n^{th} RP is computed between the estimated and actual fingerprints in dB as follows

$$MAE_n = \frac{1}{A}\sum_{i=1}^{A} |s_{n,i} - \hat{s}_{n,i}|, \tag{11}$$

where $n \in QRP$, $s_{n,i}$ is the formerly measured RSS value of AP_i in the training stage and $\hat{s}_{n,i}$ is the corresponding reconstructed value.

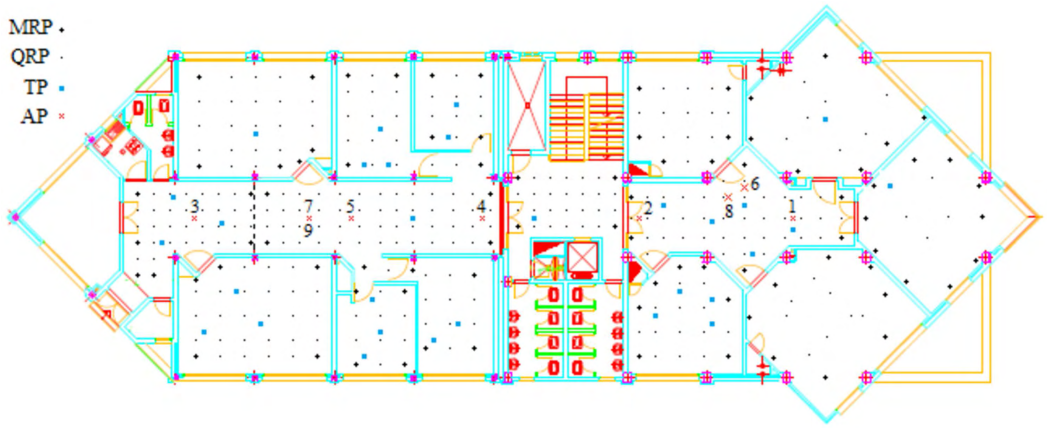
2) LOCALIZATION ERROR

For evaluating the localization accuracy with WKNN algorithms, we used the distance error between the estimated and true locations. If there are P number of TPs and the real and estimated position of the pth TP is considered as θ_p^{TP} and $\hat{\theta}_p^{TP}$, the Mean Distance Error (MDE) of fingerprinting network is defined in meter by [17]

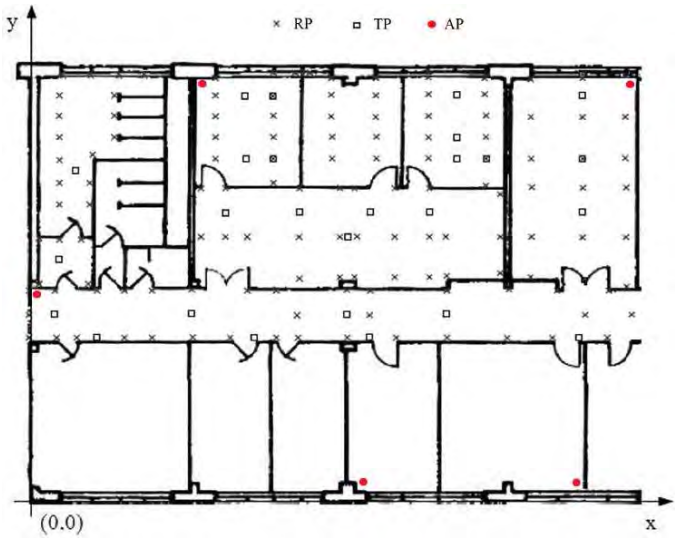
$$MDE = \frac{1}{P}\sum_{p=1}^{P} \|\hat{\theta}_p^{TP} - \theta_p^{TP}\|. \tag{12}$$

C. PARAMETERS TUNING

All the mentioned methods are sensitive to the input parameters and we choose their values based on the best results they have provided. For IDW, the power to the distance weight is set to 2, which is its optimal value for these experiments. In RBF interpolation, Multiquadric Radial Basis Function is considered as it is rated as the best RBF interpolation method [24]. The width of the rings in WRB method is also tuned to 0.2m.



(a)



(b)

FIGURE 2. The real test experiment layouts with APs, MRPs, QRPs, and TPs locations, where a) includes 15 zones and b) includes 7 zones.

VI. PERFORMANCE EVALUATION

In this section, the measured and reconstructed RSS values are analyzed and the reconstruction errors are calculated for all the presented methods. The localization results are also brought here to evaluate the accuracy of the location fingerprinting network. In addition, other MRPs' configurations are assessed. The computational complexities and the device diversity issue are also discussed. Finally the proposed methods are evaluated in a benchmarking testbed.

A. DATA SIMILARITY ANALYSIS AND COMPARISON

This part analyzes the recovered RSS data obtained by path-loss and interpolation methods and also compares them with the recorded RSS data formerly collected, when we surveyed the area. The RSS values of QRPs are reconstructed and evaluated utilizing all recovery methods based on the data retrieved using two recovery methods, the conventional and the proposed zone-based (ZB) methods. The former considers all the MRPs to recover the RSS of all QRPs and the latter examines the MRPs inside each zone to estimate the RSS of the QRPs in that zone.

In order to estimate the RSS values at QRPs using conventional path-loss model in the whole environment, all the data available for MRPs are employed to calculate the path-loss parameters in (5) and (6). Table 2 lists the path-loss parameters in this case for both test experiments. As it is expected for testbed 1, the path-loss exponents and attenuations are higher for AP_8 and AP_9 since they are located at a lower floor. These parameters are then exploited to estimate the RSS of QRPs by (5). For the ZB case, the specific path-loss parameters of each zone are computed using the MRPs data within that part and used to approximate the RSS of QRPs inside that area. Therefore, there will be 15 and 7 sets of path-loss parameters similar to Table 2 for 15 and 7 zones of testbed 1 and testbed 2, respectively as shown in Fig. 2.

The RSS values comparison between measured data and the recovered ones using conventional and zone-based path-loss models are illustrated in Fig. 3a and Fig. 3b for testbed 1

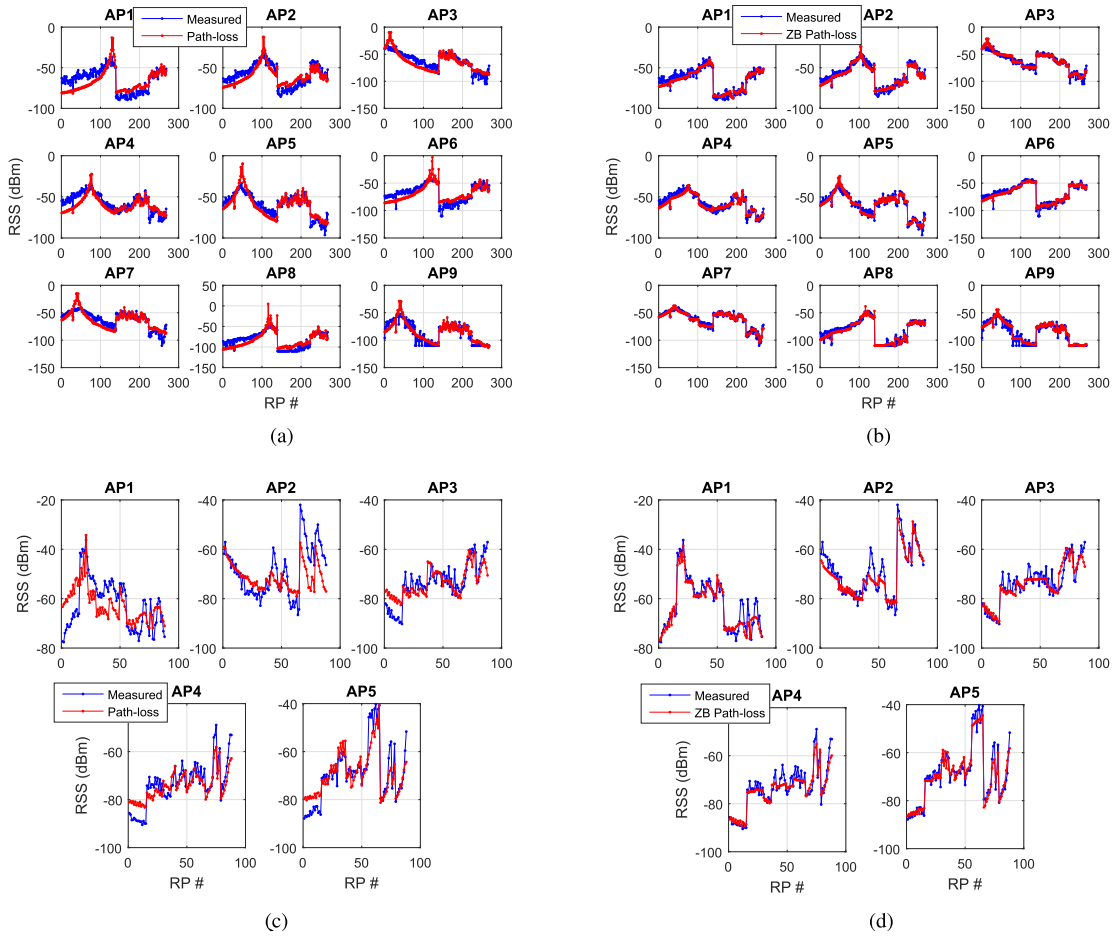


FIGURE 3. RSS values comparison between the measured and recovered databases using Conventional path-loss and proposed zone-based path-loss models.

TABLE 2. Path-loss parameters related to all APs in both testbed areas.

	testbed 1			testbed 2		
AP#	A (dBm)	α	σ (dB)	A (dBm)	α	σ (dB)
1	-26.25	3.64	6.80	-36.96	2.84	8.41
2	-24.64	3.51	5.56	-50.96	2.01	7.22
3	-19.69	4.28	7.56	-20.65	4.79	4.33
4	-24.63	3.21	5.95	-24.88	4.16	4.62
5	-20.06	4.27	6.28	-24.92	4.06	6.13
6	-22.54	4.26	8.20			
7	-23.16	4.27	8.10			
8	-26.60	5.33	9.08			
9	-38.90	4.91	5.87			

TABLE 3. Correlation values between the measured and recovered databases using path-loss model before and after considering zones.

	Correlation values in testbed 1		Correlation values in testbed 2	
AP#	Conventional	Zone-based	Conventional	Zone-based
1	0.73	0.96	0.63	0.95
2	0.83	0.97	0.74	0.95
3	0.85	0.96	0.82	0.93
4	0.77	0.95	0.90	0.97
5	0.88	0.97	0.91	0.98
6	0.80	0.97		
7	0.84	0.98		
8	0.83	0.98		
9	0.92	0.97		

and in Fig. 3c and Fig. 3d for testbed 2. The figures indicate that in both cases the estimated data have similar trend with the measured data but the proposed ZB cases in both testbeds provide much more similarity. The corresponding Pearson correlation values of measured and reconstructed data for each AP are shown in Table 3 for both testbeds. Comparing results of the conventional and ZB cases in this Table demonstrates how high the correlation values become when the zones are considered separately.

Similar to what we have done for path-loss model, the interpolation techniques are also assessed. The average correlation values between the measured and recovered databases over all APs before and after considering zones are shown in Table 4. The Table implies that the interpolation methods are also affected in zone-based scenario. In order to confirm whether the reconstructed RSS values are significantly correlated with the RSS measure-

TABLE 4. Mean correlation values between the measured and recovered databases for all APs before and after considering zones.

	Correlation values in testbed 1		Correlation values in testbed 2	
Recovery method	Conventional	Zone-based	Conventional	Zone-based
Path-loss	0.83	0.97	0.80	0.96
NN	0.92	0.96	0.84	0.94
Linear	0.93	0.97	0.90	0.96
Natural	0.94	0.97	0.91	0.96
Cubic spline	0.93	0.97	0.89	0.96
IDW	0.91	0.97	0.87	0.96
RBF	0.95	0.97	0.91	0.96
WRB	0.83	0.97	0.86	0.96

ments, the two-tailed p-values for Pearson's correlation are calculated between the measured RSS database and each of reconstructed databases by the conventional and zone-based methods. All the p-values are very small, which are expected as the correlation values are more than 0.8 in all cases and the sample sizes of testbeds 1 and 2 are almost large and medium. The p-values of the conventional and the zone-based methods are in order of 10^{-30} and 10^{-90} for testbed 1 (sample size of 268), and in order of 10^{-7} and 10^{-20} for testbed 2 (sample size of 88), respectively. Therefore, the p-values highly decreased when the zone-based models are applied, which indicates that the correlation of zone-based databases with the measured one is statistically more significant, and there is a lower probability that this strong correlation is due to chance.

For the all path-loss and interpolation techniques, we compute the MAE between the actual and reconstructed fingerprints. The average MAE and the standard deviations of the MAEs over all QRPs are illustrated in Fig. 4. It can be seen that how much recovery errors are decreased when zones considered independently. It has noticeable impact on the path-loss and WRB methods. The zone-based MAE results are so close for all methods as their correlation values are almost the same shown in Table 4. However, it is visible in the figure that the proposed WRB interpolation technique

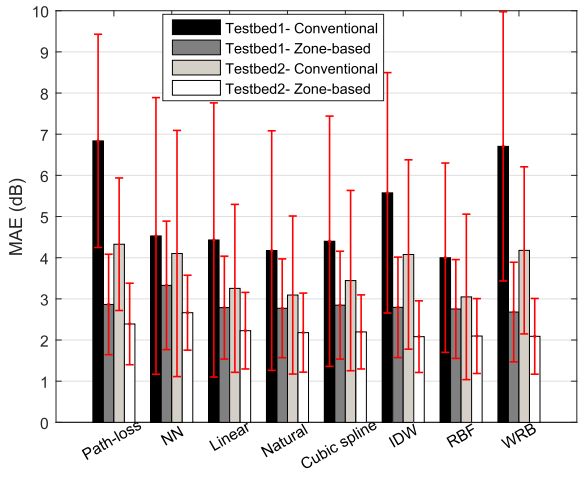


FIGURE 4. The average MAE and the related standard deviation of measured and recovered RSS values based on all APs.

expresses the minimum average MAE among all recovery methods in the zone-based case. Comparing MAE of the path-loss and all interpolation methods demonstrates that the most influenced methods using zone-based method is the path-loss model with 57% and 46% error improvement for testbeds 1 and 2, respectively. RBF also shows the least error improvement, 31.2% and 31.8% in testbed 1 and 2 areas.

When standard deviations are considered, Fig. 4 indicates the zone-based versions of all mentioned conventional methods significantly decrease the standard deviations of the MAE as well. In the zone-based methods, NN provides the largest standard deviation compared to the others, whereas all other methods show very similar standard deviations in both testbeds.

B. LOCALIZATION RESULTS

Reconstructing and recovering the fingerprint database has been discussed so far and all the eight methods were compared in terms of their RSS similarities with the measured values. In this part, we move on to the positioning stage of fingerprinting in order to evaluate and verify the localization performance of our suggested recovery methods in comparison with traditional ones. To have a comparable result with the real situations, the measured RSS of TPs are considered to compute the positioning error, as in the real situations also the users always send their measured fingerprints for localization to the server regardless of having a measured or reconstructed database.

Among the deterministic techniques explained in sub-Section III-C, the WKNN method is employed here as a positioning algorithm as it provides the highest accuracy. The WKNN method shows the lowest MDE at $K = 6$ and $K = 4$ for the testbeds 1 and 2, respectively, so we use them throughout the paper. The MDEs and their related standard deviations before and after removing RPs are listed in Table 5. As it is shown, the MDEs and their related standard deviations are higher after RP removal process than MDEs and standard deviations before removing RPs, because the localization is carried out only with 24.4% and 32% of the total RPs in testbeds 1 and 2. By applying the recovery methods, the RSS of QRPs can be reconstructed so that we reduce the error obtained by removing the RPs.

TABLE 5. MDE and the related standard deviation before and after removing RPs for both testbed areas.

	testbed 1 ($K = 6$)		testbed 2 ($K = 4$)	
Recovery method	MDE(m)	Std.	MDE(m)	Std.
Full measured	1.45	0.71	1.19	0.76
MRPs only	3.28	1.73	1.97	1.27

The achieved MDEs using all the eight recovery methods are reported in Table 6, which indicate how localization accuracy is improved and is getting closer to the actual error by using reconstructed RSS data. Considering traditional and proposed path-loss models only, the Table shows that MDE

of proposed zone-based path-loss model is decreased by 31% and 19% in both testbeds 1 and 2, respectively. Among all interpolation methods, the suggested zone-based WRB outperforms traditional methods and provides the lowest positioning error. The MDE of zone-based WRB in testbed 1 is reduced by 23.2% and 13.1% compared to highest and lowest MDEs in traditional methods (Natural and RBF), and in testbed 2 the error is decreased by 33% and 12% in comparison with highest and lowest MDEs in traditional methods (NN and RBF). Furthermore, the zone-based methods generally gives lower positioning errors except RBF. The results in Table 6 imply that despite the fact that zone-based interpolation methods show the lower MAEs according to Fig. 4, they do not necessarily provide lower MDEs as well. This outcome corresponds to the results in [3] in which similar problem is discussed and improving the positioning accuracy is highlighted rather than finding the best fit for fingerprints. For instance, although all the zone-based interpolation methods shows almost similar MAEs (except NN), they do not provide the same localization accuracy. Hence, a high similarity between the measured and recovered RSS values, does not automatically result in a high positioning accuracy.

According to Table 6, the proposed zone-based versions of all mentioned recovery methods, result in lower standard deviations in localization errors as well. The standard deviations of the proposed zone-based path-loss model and zone-based WRB are the lowest compared to the others, and also the closest to the actual standard deviations (shown in Table 5) when the full measured databases are employed.

The cumulative distribution function (CDF) of the localization errors are plotted in Fig. 5 and Fig. 6 for testbeds 1 and 2. The solid black line is the CDF of full workload when the full measured database are used and the dash black line shows the CDF of when only MRPs are used for positioning. The CDF of conventional path-loss and interpolation methods are also compared with the proposed zone-based path-loss and zone-based WRB techniques. A recovery method with a closer CDF to the full measured database demonstrates the higher efficiency. As can be seen in both testbed areas, the proposed zone-based path-loss and zone-based WRB outperforms traditional methods and their CDF lines are closer to the CDF of full workload.

Therefore, the accuracy achieved by the proposed methods and the CDF lines confirm the higher localization performance of both zone-based path-loss and WRB methods.

TABLE 6. MDE and the related standard deviation of using recovered databases before and after considering zones.

	testbed 1				testbed 2			
	Conventional		Zone-based		Conventional		Zone-based	
Recovery method	MDE(m)	Std.	MDE(m)	Std.	MDE(m)	Std.	MDE(m)	Std.
Path-loss	3.20	1.59	**2.21**	**1.37**	1.74	1.03	**1.41**	**0.83**
NN	2.71	1.58	2.61	1.53	2.14	1.27	1.90	1.03
Linear	2.78	1.74	2.52	1.60	1.68	1.00	1.60	0.97
Natural	2.84	1.86	2.45	1.55	1.68	0.92	1.67	0.89
Cubic spline	2.52	1.58	2.61	1.61	1.75	0.97	1.58	0.88
IDW	2.81	1.73	2.31	1.66	1.69	1.04	1.61	0.93
RBF	2.51	1.49	2.63	1.58	1.63	0.90	1.65	0.86
WRB	2.78	1.50	**2.18**	**1.45**	1.72	0.97	**1.44**	**0.85**

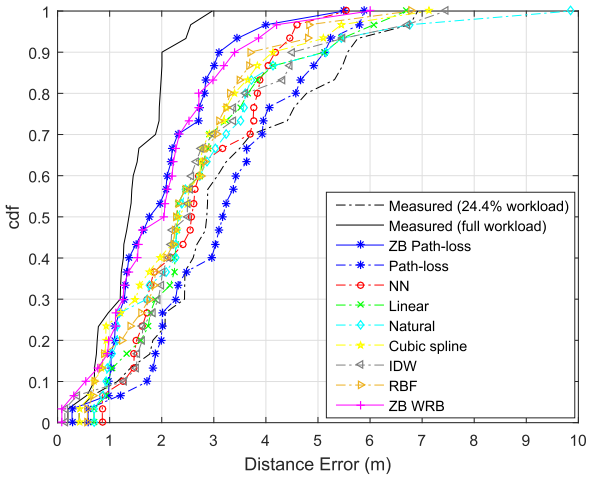


FIGURE 5. Cumulative distribution functions for all recovery methods in Testbed 1 environment.

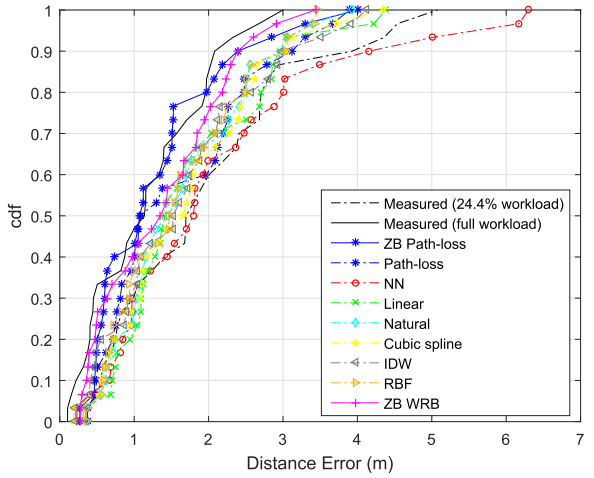


FIGURE 6. Cumulative distribution functions for all recovery methods in Testbed 2 environment.

To assess the conventional and zone-based path-loss model when they are used for ranging, (7) is utilized for all QRPs. The measured RSS values of QRPs and the zone-related path-loss parameters are used to estimate the distance between the QRPs and APs. Then (8) can produce the distance estimation error for all points regarding each AP. Fig. 7 illustrates the average distance estimation errors (m) from all QRPs to each AP using both conventional and zone-based path-loss models in testbeds 1 and 2. This figure indicates that the error is noticeably decreased when the zone-based model is employed. The average error over all APs for conventional and zone-based path-loss models are 5.92 (m) and 3.40 (m) in testbed 1 and 4.35 (m) and 2.88 (m) in testbed 2, respectively. Since in distance-based positioning, reducing the distance calculation error results in a higher localization and tracking accuracies [42], we can expect to achieve better positioning accuracies when the proposed zone-based path-loss model is used.

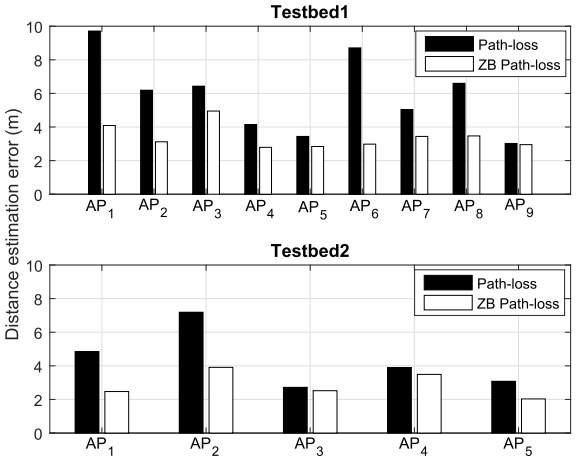


FIGURE 7. Average distance estimation error (m) from all QRPs to APs using conventional and zone-based path-loss models.

In order to statistically evaluate the significance of the localization accuracy improvement, we adopted the two-tailed paired-sample T-test similar to [43]. In this paper, the two proposed zone-based path-loss model and the zone-based WRB method were evaluated as follows:

i) The positioning distance errors before and after applying proposed zone-based method on path-loss model are statistically compared. The degree of freedom for both testbeds is 29. The p-values obtained are 0.002 and 0.028 for testbeds 1 and 2, which are lower than the significance level of 0.05. Therefore, there is a statistically significant difference in the localization accuracies achieved by the conventional and zone-based path-loss models.

ii) The positioning distance errors of the zone-based WRB and any of conventional NN, Linear, Natural, Cubic spline, IDW, and RBP interpolation methods were statistically compared. The degree of freedom for both testbeds is 29. The p-values results are reported in Table 7, which indicates that the localization accuracy of the zone-based WRB model is significantly different from accuracies of other interpolation methods when the significance level is set to 0.1.

TABLE 7. The p-values of two-tailed paired-sample T-test on the positioning distance errors of the zone-based WRB method and other evaluated interpolation methods.

Recovery method	NN	Linear	Natural	Cubic spline	IDW	RBF
p-value (Testbed 1)	0.06	0.006	0.01	0.09	0.01	0.10
p-value (Testbed 1)	0.002	0.03	0.09	0.03	0.07	0.09

C. VARIOUS RPS RATIOS

Here we investigate the conventional and zone-based recovery methods in various ratios of the selected MRPs. In subsection VI-B, we were limited for choosing the MRPs because the QRPs had to be inside the convex hull created by MRPs for triangulation-based interpolation techniques. Here, however, we only consider path-loss model and the non-triangulation-based interpolation methods, so the location of MRPs/QRPs can be anywhere. We are also free to remove more of those MRPs, since there is no need to create the convex hull. We try to keep at least two and three MRPs inside small and large zones, respectively. If one MRP is only considered, an appropriate interpolation cannot be supported. Furthermore, as the aim of this paper is reducing the surveying cost of the fingerprinting approach, only the low percentages of MRPs to all RPs are assessed here. The NN method is also excluded as it provides very poor results when MRPs are very sparse and far from each other.

Fig. 8 and Fig. 9 illustrate the MDEs for various ratios of MRPs' number to the total number of RPs in percent for testbeds 1 and 2, respectively. In testbed 1, the numbers of MRPs are 38, 61, and 86 out of 354 RPs, which are 11%, 17% and 24% of all RPs. The numbers of MRPs in testebed 2 are 16, 27, and 42 out of 130 RPs, which are 12%, 21% and 32% of all RPs. The figures demonstrate that employing recovery methods to reconstruct the RSSs of

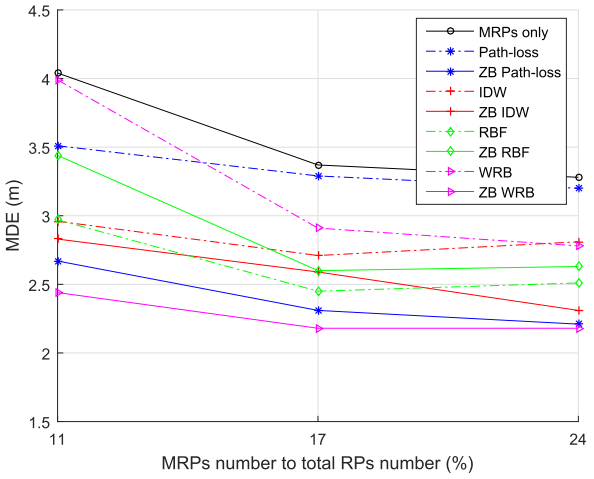


FIGURE 8. MDE of various percentages of MRPs in the testbed 1 environment.

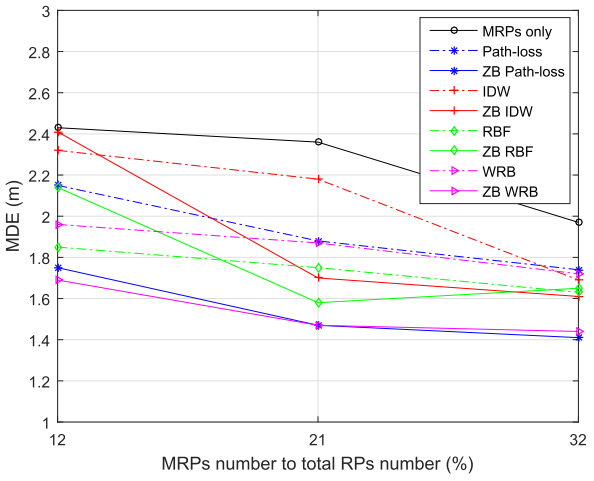


FIGURE 9. MDE of various percentages of MRPs in the testbed 2 environment.

QRPs can provide higher localization accuracies compared to using MRPs only which is shown by the black line. It indicates the influence of the number of RPs on positioning error even when the reconstructed version of RSSs are used for RPs. Moreover, the figures imply that MDE is improved by increasing the number of MRPs, since higher number of measured RPs contributes to carry out the localization. In addition, similar to what we achieved in sub-section VI-B, the MDE of zone-based methods is less than the conventional ones even in lower percentages of MRPs' number (except for RBF as seen before). The figures also show that both the proposed zone-based path-loss method and the zone-based WRB interpolation method outperform the other recovery methods when different numbers of MRPs are considered. The suggested zone-based WRB provides the highest accuracy even when we have a very low number of MRPs.

According to Fig. 8, when the 89% of surveying cost is reduced, utilizing the zone-based WRB helps improve the accuracy by 40%. In testbed 2 also, the zone-based WRB enhances 30% of the localization error when the surveying cost is decreased by 88%. Comparing conventional and proposed zone-based path-loss models only in the lowest MRPs numbers of testbeds 1 and 2, the zone-based version enhances the positioning accuracy by 26% and 19%, respectively.

This procedure can help fingerprinting network designers understand the behavior of the system and its localization accuracy merely by measuring RSS at a few numbers of RPs with minimum effort. Based on the application and a desirable network performance for that, they can choose the necessary numbers of RPs and APs and make a trade-off between the accuracy and the essential time/energy costs.

D. COMPUTATIONAL COSTS COMPARISON

In this part, we discuss the computational costs of all the evaluated methods before and after applying zone-based model on them. The significant point here is that we are dealing with the training stage only, which is done offline. Therefore, the computational costs here are only related to the training stage and do not affect the online localization process, in which the used memory and running time are much more critical. We assess the computational costs in terms of the required space (memory) and running time complexity. The needful space in all mentioned recovery methods is the same because they all require only as much memory as the size of output database ($\hat{Y}$). Therefore, by knowing the $N \times (A+2)$ as the dimensions of the database and considering B as the number of required units (e.g. bytes) to save fingerprints at each RP, the required memory is obtained by $BN(A+2)$.

The running time complexities for all evaluated methods are compared in Table 8. According to this Table, we have a summation of running times over all zones for zone-based cases. As the numbers of MRPs and QRPs in each zone, M_z and Q_z, are very much lower than the total numbers of MRPs and QRPs (M and Q), the number of comparisons is reduced for the proposed zone-based model. The zone-based WRB method also shows the lowest time complexity since the number of MRPs and QRPs at each ring (M_{zr} and Q_{zr}) are very low or even zero.

TABLE 8. Time complexity for all recovery methods.

	Time complexity	
Recovery method	Conventional	Zone-based
Path-loss	$\mathcal{O}(AQM^2)$	$\sum_{z=1}^{Z} \mathcal{O}(AQ_zM_z^2)$
NN	$\mathcal{O}(AQM)$ [46]	$\sum_{z=1}^{Z} \mathcal{O}(AQ_zM_z)$
Linear	$\mathcal{O}(AQM \log M)$ [47]	$\sum_{z=1}^{Z} \mathcal{O}(AQ_zM_z \log M_z)$
Natural	$\mathcal{O}(AQM \log M)$ [47]	$\sum_{z=1}^{Z} \mathcal{O}(AQ_zM_z \log M_z)$
Cubic spline	$\mathcal{O}(AQM \log M)$ [47]	$\sum_{z=1}^{Z} \mathcal{O}(AQ_zM_z \log M_z)$
IDW	$\mathcal{O}(AQM^2)$ [48]	$\sum_{z=1}^{Z} \mathcal{O}(AQ_zM_z^2)$
RBF	$\mathcal{O}(AQM^2)$ [49]	$\sum_{z=1}^{Z} \mathcal{O}(AQ_zM_z^2)$
WRB	$\sum_{r=1}^{r_{max}} \mathcal{O}(AQ_rM_r)$	$\sum_{z=1}^{Z} \sum_{r=1}^{r_{max}} \mathcal{O}(AQ_{zr}M_{zr})$

E. DEVICE DIVERSITY

The influence of device diversity on the proposed methods in the fingerprinting technique can be divided into two parts. The first one is whether using different devices in the training stage affects the associated reconstructed databases and the localization results they provide. Note that we are not talking about the mixed RSS fingerprints of all devices and we deal with creating database using the fingerprints of each device separately. This issue is evaluated in this paper by investigating three experiments with three different devices. The results demonstrated that all the reconstructed databases using the proposed zone-based path-loss model and zone-based WRB interpolation resulted in the lower reconstruction and localization errors in all three test experiments. Therefore, the diversity of devices used to collect the RSS fingerprints, does not affect the results of the proposed methods.

The second part is that the diversity of the users' devices in the localization stage definitely affects the localization results, since different devices may receive significantly distinct signal strengths from the same AP due to the difference in hardware. There have been many researches to tackle the RSS fluctuation and device diversity problems [31], [44], [45]. Although the focus of this paper is to create a low-cost database with proposed models, once the database is constructed, many of the suggested methods in [31], [44], [45] can be applied on that in order to lessen the influence of the device diversity on the localization accuracy.

F. TEST ON A BENCHMARKING TESTBED

We evaluated our proposed methods employing another Wi-Fi fingerprinting database available in the public domain. It is an important source of common data under the EVARILOS project. The experiment was carried out by authors of [30], [50] in the EVARILOS benchmarking Suite (EBS) in Berlin, which is developed to assess the localization solutions. The small size office TKN testbed environment (the second floor) is used here since it includes several zones. The dimensions of the testbed area is 30 m $\times$ 15 m. The MacBook Pro notebook is used as a client device. They assigned 41 locations to RPs and 20 fingerprints were collected per

location in 13 zones. There are also 20 TPs as the evaluation points. They reported an average error of 2.21 m in this area [30].

In order to assess our approach, we only considered 7 zones of this testbed, where at least three RPs are measured. Since we need at least two MRPs per zone, the number of RPs and TPs are reduced to 29 and 14 in 7 zones. For zone-based model evaluation, we removed 48% of RPs (14 RPs). The Mean Distance Error (MDE) with full measured 29 RPs is 3.25 m while it increases to 3.77 m after removing 14 RPs. The location of MRPs, QRPs, TPs, and APs are shown in Fig. 10. There are 15 MRPs, 14 QRPs, 14 TPs, and 4 APs in this test area. As the number of measured RPs were limited in each zone, we only compared the non-triangulation-based algorithms so that we could interpolate with at least two MRPs in each zone. The localization results are listed in Table 9. Similar to what we achieved for testbeds 1 and 2, the proposed zone-based path-loss model provides the highest localization accuracy. The proposed zone-based WRB method also shows the lowest MDE among other interpolation methods.

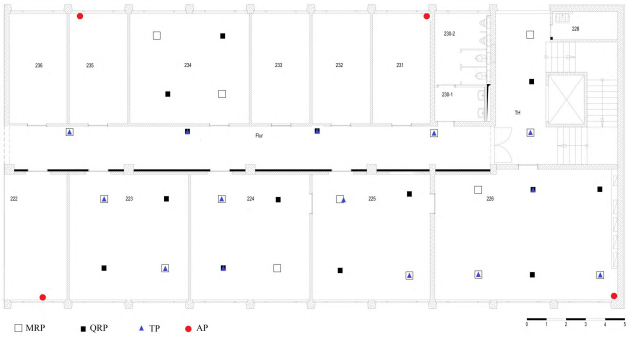


FIGURE 10. The EVARILOS benchmarking layout with APs, MRPs, QRPs, and TPs.

TABLE 9. MDE of using recovered databases before and after considering zones in EVARILOS layout.

	MDE (m) in EVARILOS test area	
Recovery method	Conventional	Zone-based
Path-loss	3.76	**3.34**
IDW	3.60	3.55
RBF	3.47	3.49
WRB	3.52	**3.37**

VII. CONCLUSION

In indoor localization, constructing the measured database in the training stage of location fingerprinting technique is so costly in terms of the required time and human efforts. Removing RPs considerably decreases this cost but it highly degrades the localization accuracy at the same time. Therefore, in order to enhance the accuracy, we propose two novel cost-effective recovery methods that predict the RSS of omitted RPs and build a reconstructed version of the database. A novel zone-based method is suggested here which is based on the environmental effects of the indoor area and the signal propagation behavior inside the indoor area. It is applied in the path-loss model and six interpolation methods used in the indoor fingerprinting technique, in which the function parameters are computed separately for different zones. The zone-based path-loss model is fully compared with the conventional path-loss and interpolation methods. Moreover, a new zone-based WRB interpolation method is introduced, and its performance is compared with the six other interpolation methods. The experiments conducted in two different indoor environments and in a benchmarking testbed with three distinct devices and various MRPs ratios are utilized to validate the proposed recovery methods.

The reconstruction error results demonstrate that the zone-based version of all mentioned methods highly decreases the error (up to 57%) between the measured and reconstructed RSS values. The proposed zone-based WRB interpolation technique expresses the minimum reconstruction error among all recovery methods. In addition, the localization results indicate that the suggested methods not only can produce a more dense database with only a very few numbers of RPs for the database construction, but also enhance the accuracy of indoor fingerprinting localization compared to the conventional methods. By reducing the workload by 89% of the initial RPs, the new zone-based path-loss model decreases the localization error up to 26% in comparison with the conventional path-loss model and the new zone-based WRB method outperforms all the other interpolation methods and improves the accuracy up to 40%. Hence, the database can be updated with a very low cost if there are any environmental changes in the target area. The computational costs of the zone-based methods also are lower than the conventional methods. Moreover, it was shown that the lowest reconstruction error, does not necessarily result in the highest localization accuracy, but regarding localizing purposes, the lowest positioning error is highly desired. The benchmarking results also confirmed that zone-based path-loss and WRB methods achieved the highest positioning results.

In our ongoing work, we are trying to investigate these approaches to a multi-floor indoor environment. Also we plan to apply the proposed method on an indoor Bluetooth Low Energy (BLE) network to have a low-cost location fingerprinting and compare the achieved localization accuracy with the MDE of current WLAN network, eventually. Finally, the impact of the RPs with possible unusual RSS values in the interpolation field needs to be investigated.

ACKNOWLEDGMENT

The author, Vahideh Moghatdaiee, would like to thank Amir Mahdi Sazdar for his help in data collection procedure and Dr. Binghao Li for his Wi-Fi data.

REFERENCES

[1] G. Han, J. Jiang, C. Zhang, T. Q. Duong, M. Guizani, and G. Karagiannidis, "A survey on mobile anchor node assisted localization in wireless sensor networks," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 2220–2243, 3rd Quart., 2016.

[2] V. Moghtadaiee, S. Lim, and A. G. Dempster, "System-level considerations for signal-of-opportunity positioning," in *Proc. Int. Symp. GPS/GNSS*, Taipei, Taiwan, 2010, pp. 1–7.
[3] J. Talvitie, M. Renfors, and E. S. Lohan, "Distance-based interpolation and extrapolation methods for RSS-based localization with indoor wireless signals," *IEEE Trans. Veh. Technol.*, vol. 64, no. 4, pp. 1340–1353, Apr. 2015.
[4] P. Bahl and V. Padmanabhan, "RADAR: An in-building RF-based user location and tracking system," in *Proc. IEEE 19th Annu. Joint Conf. IEEE Comput. Commun. Soc. (INFOCOM)*, vol. 2, Mar. 2000, pp. 775–784.
[5] V. Moghtadaiee and A. G. Dempstera, "Indoor location fingerprinting using FM radio signals," *IEEE Trans. Broadcast.*, vol. 60, no. 2, pp. 336–346, Jun. 2014.
[6] Y.-H. Wu, Y.-L. Chen, and S.-T. Sheu, "Indoor location estimation using virtual fingerprint construction and zone-based remedy algorithm," in *Proc. Int. Conf. Commun. Problem-Solving (ICCP)*, Taipei, Taiwan, Sep. 2016, pp. 1–3.
[7] J. Zuo, S. Liu, H. Xia, and Y. Qiao, "Multi-phase fingerprint map based on interpolation for indoor localization using iBeacons," *IEEE Sensors J.*, vol. 18, no. 8, pp. 3351–3359, Jan. 2018.
[8] M. Zhou, Y. Tang, Z. Tian, and X. Geng, "Semi-supervised learning for indoor hybrid fingerprint database calibration with low effort," *IEEE Access*, vol. 5, pp. 4388–4400, 2017.
[9] J. Zhang, G. Han, N. Sun, and L. Shu, "Path-loss-based fingerprint localization approach for location-based services in indoor environments," *IEEE Access*, vol. 5, pp. 13756–13769, 2017.
[10] T. Bailey and H. Durrant-Whyte, "Simultaneous localization and mapping (SLAM): Part II," *IEEE Robot. Autom. Mag.*, vol. 13, no. 3, pp. 108–117, Sep. 2006.
[11] M. M. Atia, A. Noureldin, and M. J. Korenberg, "Dynamic online-calibrated radio maps for indoor positioning in wireless local area networks," *IEEE Trans. Mobile Comput.*, vol. 12, no. 9, pp. 1774–1787, Sep. 2013.
[12] M. Zhou, Q. Zhang, Y. Wang, and Z. Tian, "Hotspot ranking based indoor mapping and mobility analysis using crowdsourced Wi-Fi signal," *IEEE Access*, vol. 5, pp. 3594–3602, 2017.
[13] X. Zhou, T. Chen, D. Guo, X. Teng, and B. Yuan, "From one to crowd: A survey on crowdsourcing-based wireless indoor localization," *Frontiers Comput. Sci.*, vol. 12, no. 3, pp. 423–450, 2018.
[14] F. Fuschini, E. M. Vitucci, M. Barbiroli, G. Falciasecca, and V. Degli-Esposti, "Ray tracing propagation modeling for future small-cell and indoor applications: A review of current techniques," *Radio Sci.*, vol. 50, no. 6, pp. 469–485, Jun. 2015.
[15] O. Belmonte-Fernández, R. Montoliu, J. Torres-Sospedra, E. Sansano-Sansano, and D. Chia-Aguilar, "A radiosity-based method to avoid calibration for indoor positioning systems," *Expert Syst. Appl.*, vol. 105, pp. 89–101, Sep. 2018.
[16] C. Feng, W. S. A. Au, S. Valaee, and Z. Tan, "Received-signal-strength-based indoor positioning using compressive sensing," *IEEE Trans. Mobile Comput.*, vol. 11, no. 2, pp. 1983–1993, Dec. 2012.
[17] A. Khalajmehrabadi, N. Gatsis, and D. Akopian, "Structured group sparsity: A novel indoor WLAN localization, outlier detection, and radio map interpolation scheme," *IEEE Trans. Veh. Technol*, vol. 66, no. 7, pp. 6498–6510, Jul. 2017.
[18] S.-T. Sheu, Y.-M. Hsu, and H.-Y. Chen, "Indoor location estimation using smart antenna system with virtual fingerprint construction scheme," in *Proc. Int. Conf. Mobile Ubiquitous Comput., Syst., Services Technol. (UBICOMM)*, Rome, Italy, 2014, pp. 1–6.
[19] R. Kubota, S. Tagashira, Y. Arakawa, T. Kitasuka, and A. Fukuda, "Efficient survey database construction using location fingerprinting interpolation," in *Proc. IEEE 27th Int. Conf. Adv. Inf. Netw. Appl.*, Barcelona, Spain, Mar. 2013, pp. 469–476.
[20] L. Ma, W. Zhao, Y. Xu, and C. Li, " Radio map efficient building method using tensor completion for WLAN indoor positioning system," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kansas City, MO, USA, May 2018, pp. 1–6.
[21] M. Zhou, Y. Tang, Z. Tian, and F. Qiu, "Reducing calibration effort for indoor WLAN localization using hybrid fingerprint database," in *Machine Learning and Intelligent Communications* (Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering), vol. 183. Cham, Switzerland: Springer, 2017, pp. 159–168.
[22] S. Ezpeleta, J. M. Claver, J. J. Pérez-Solano, and J. V. Mart, "RF-based location using interpolation functions to reduce fingerprint mapping," *Sensors*, vol. 15, no. 10, pp. 27322–27340, 2015.
[23] A. H. Ismail, H. Kitagawa, R. Tasaki, and K. Terashima, "WiFi RSS fingerprint database construction for mobile robot indoor positioning system," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Oct. 2016, pp. 1561–1566.
[24] S. Pino-Povedano, C. Bousoño-Calzón, and F. J. González-Serrano, "Radial basis function interpolation for signal-model-independent localization," *IEEE Sensors J.*, vol. 16, no. 7, pp. 2028–2035, Apr. 2016.
[25] H. Zhao, B. Huang, and B. Jia, "Applying kriging interpolation for WiFi fingerprinting based indoor positioning systems," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Apr. 2016, pp. 1–6.
[26] B. J. Dil and P. J. M. Havinga, "On the calibration and performance of RSS-based localization methods," in *Proc. Internet Things (IOT)*, Nov./Dec. 2010, pp. 1–8.
[27] W. Bong and Y. C. Kim, "Fingerprint Wi-Fi radio map interpolated by discontinuity preserving smoothing," in *Proc. 6th Int. Conf. Hybrid Inf. Technol.*, Daejeon, South Korea, 2012, pp. 138–145.
[28] K. Arai and H. Tolle, "Color radiomap interpolation for efficient fingerprint WiFi-based indoor location estimation," *Int. J. Adv. Res. Artif. Intell.*, vol. 2, no. 3, pp. 10–15, 2013.
[29] A. Arya, P. Godlewski, M. Campedel, and G. du Chene, "Radio database compression for accurate energy-efficient localization in fingerprinting systems," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 6, pp. 1368–1379, Jun. 2013.
[30] F. Lemic, A. Behboodi, V. Handziski, and A. Wolisz, "Experimental decomposition of the performance of fingerprinting-based localization algorithms," in *Proc. Int. Conf. Indoor Positioning Indoor Navigat. (IPIN)*, Oct. 2014, pp. 355–364.
[31] L. Li, G. Shen, C. Zhao, T. Moscibroda, J.-H. Lin, and F. Zhao, "Experiencing and handling the diversity in data density and environmental locality in an indoor positioning service," in *Proc. Int. Conf. Mobile Comput. Netw.*, 2014, pp. 459–470.
[32] Y. Zhuang, Z. Syed, Y. Li, and N. El-Sheimy, "Evaluation of two WiFi positioning systems based on autonomous crowdsourcing of handheld devices for indoor navigation," *IEEE Trans. Mobile Comput.*, vol. 15, no. 8, pp. 1982–1995, Aug. 2016.
[33] J. Talvitie, M. Renfors, M. Valkama, and E. S. Lohan, "Method and analysis of spectrally compressed radio images for mobile-centric indoor localization," *IEEE Trans. Mobile Comput.*, vol. 17, no. 4, pp. 845–858, Apr. 2018.
[34] B. Jang and H. Kim, "Indoor positioning technologies without offline fingerprinting map: A survey," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 1, pp. 508–525, 1st Quart., 2019.
[35] V. Moghtadaiee and A. G. Dempster, "Determining the best vector distance measure for use in location fingerprinting," *Pervasive Mobile Comput.*, vol. 23, pp. 59–79, Oct. 2015.
[36] Y. Chen and A. Terzis, "On the implications of the log-normal path loss model: An efficient method to deploy and move sensor motes," in *Proc. 9th ACM Conf. Embedded Netw. Sensor Syst.*, 2011, pp. 26–39.
[37] S. Xia, Y. Liu, G. Yuan, M. Zhu, and Z. Wang, "Indoor fingerprint positioning based on Wi-Fi: An overview," *ISPRS Int. J. Geo-Inf.*, vol. 6, no. 5, p. 135, Apr. 2017.
[38] Y. Du, D. Yang, and C. Xiu, "A novel method for constructing a WiFi positioning system with efficient manpower," *Sensors*, vol. 15, no. 4, pp. 8358–8381, 2015.
[39] J. J.-J. Li, S. Sen, and B. Hecht, "Leveraging advances in natural language processing to better understand Tobler's first law of geography," in *Proc. ACM SIGSPATIAL Int. Conf. Adv. Geograph. Inf. Syst.*, 2014, pp. 513–516.
[40] A. Okabe, B. Boots, and K. Sugihara, *Spatial Tessellations*. Hoboken, NJ, USA: Wiley, 1992, p. 532.
[41] B. Li, Y. Wang, H. K. Lee, A. Dempster, and C. Rizos, "Method for yielding a database of location fingerprints in WLAN," *IEE Proc.-Commun.*, vol. 152, no. 5, pp. 580–586, 2005.
[42] J. Li, G. Feng, W. Wei, C. Luo, L. Cheng, H. Wang, H. Song, and Z. Ming, "PSOTrack: A RFID-based system for random moving objects tracking in unconstrained indoor environment," *IEEE Internet Things J.*, vol. 5, no. 6, pp. 4632–4641, Dec. 2018.
[43] I. H. Alshami, N. A. Ahmad, S. Sahibuddin, and F. Firdaus, "Adaptive indoor positioning model based on WLAN-fingerprinting for dynamic and multi-floor environments," *Sensors*, vol. 17, no. 8, p. 1789, 2017.
[44] Y. Kim, H. Shin, Y. Chon, and H. Cha, "Smartphone-based Wi-Fi tracking system exploiting the RSS peak to overcome the RSS variance problem," *Pervasive Mobile Comput.*, vol. 9, no. 3, pp. 406–420, Jun. 2013.
[45] X. Du, K. Yang, and D. Zhou, "Mapsense: Mitigating inconsistent WiFi signals using signal patterns and pathway map for indoor positioning," *IEEE Internet Things J.*, vol. 5, no. 6, pp. 4652–4662, Dec. 2018.

[46] S. Raschka, "Stat 479—Machine learning," Dept. Statist., Univ. Wisconsin–Madison, Madison, WI, USA, Lect. Notes, Tech. Rep., 2018, pp. 1–21. [Online]. Available: https://sebastianraschka.com/pdf/lecture-notes/stat479fs18/01_ml-overview_notes.pdf
[47] C. He, S. Guo, Y. Wu, and Y. Yang, "A novel radio map construction method to reduce collection effort for indoor localization," *Measurement*, vol. 94, pp. 423–431, Dec. 2016.
[48] J. Wua, C. Zhenga, and C. C. Chien, "Cost-effective sampling network design for contaminant plume monitoring under general hydrogeological conditions," *J. Contaminant Hydrol.*, vol. 77, pp. 41–65, Mar. 2005.
[49] V. Skala, "Fast interpolation and approximation of scattered multidimensional and dynamic data using radial basis functions," *WSEAS Trans. Math.*, vol. 12, no. 5, pp. 501–511, 2013.
[50] T. Van Haute, E. De Poorter, F. Lemic, V. Handziski, N. Wirström, T. Voigt, A. Wolisz, and I. Moerman, "Platform for benchmarking of RF-based indoor localization solutions," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 126–133, Sep. 2015.

VAHIDEH MOGHTADAIEE received the B.Sc. and M.Sc. degrees in electrical engineering, in 2004 and 2007, respectively, and the Ph.D. degree from the School of Electrical Engineering and Telecommunications, University of New South Wales, Sydney, Australia, in 2013. She is currently with the Cyberspace Research Institute, Shahid Beheshti University. Her research interests include indoor positioning, signal processing, wireless communications and networks, and the IoT

SEYED ALI GHORASHI received the B.Sc. and M.Sc. degrees in electrical engineering from the University of Tehran, Iran, and the Ph.D. degree from the King's College London, in 2003, where he has been a Research Associate on projects sponsored by Mobile VCE. In 2006, he joined Samsung Electronics (UK) Ltd., and then, he served at the Department of Telecommunications, Faculty of Electrical Engineering, Shahid Beheshti University, Tehran, Iran, as an Assistant and Associate Professor. He holds three US patents, and has published over 100 technical papers mainly related to the applications of optimization, AI, and ML in wireless communications.

MOHAMMAD GHAVAMI is currently a Professor of telecommunications with the London South Bank University. Prior to this appointment, he was with King's College London, from 2002 to 2010, and with the Sony Computer Science Laboratories, Tokyo, from 2000 to 2002. He has authored the books, namely the *Ultra Wideband Signals and Systems*, and the *Adaptive Antenna Systems*, and has published over 140 technical papers mainly related to UWB and its medical applications. He holds three US and one European patents. He won the esteemed European Information Society Technologies Prize, in 2005, and two invention awards from Sony. He has been the Guest Editor of the *IET Proceedings Communications*, the *Special Issue on Ultra Wideband Systems*, and the Associate Editor of the *Special Issue of the IEICE Journal on UWB Communications*.