WinIPS: WiFi-based Non-intrusive Indoor Positioning System with Online Radio Map Construction and Adaptation

Han Zou, Ming Jin, Hao Jiang, Lihua Xie, *Fellow, IEEE* and Costas J. Spanos, *Fellow, IEEE*

***Abstract*—WiFi fingerprinting-based Indoor Positioning System (IPS) has become the most promising solution for indoor localization. However, there are two major drawbacks that hamper its large-scale implementation. Firstly, an offline site survey process is required which is extremely time-consuming and labor-intensive. Secondly, the RSS fingerprint database built offline is vulnerable to environmental dynamics. To address these issues comprehensively, in this paper, we propose WinIPS, a WiFi-based non-intrusive IPS that enables automatic online radio map construction and adaptation, aiming for calibration-free indoor localization. WinIPS can capture data packets transmitted in existing WiFi traffic and extract the RSS and MAC addresses of both WiFi Access Points (APs) and mobile devices in a non-intrusive manner. APs can be used as online reference points for radio map construction. A novel Gaussian process regression model is proposed to approximate the non-uniform RSS distribution of an indoor environment. Extensive experiments were conducted, which demonstrated that WinIPS outperforms existing solutions in terms of both RSS estimation accuracy and localization accuracy.**

***Index Terms*—Indoor Positioning System (IPS), radio map construction and adaptation, WiFi, Gaussian process regression.**

I. Introduction

LOCATION Based Service (LBS) has become an indispensable part of our daily lives due to its widespread applications, e.g., navigation, advertisement, shopping, etc., in smart buildings. The quality of LBS largely depends on the localization accuracy [1]. Global Positioning System (GPS) can provide satisfactory localization accuracy for most outdoor LBS. However, it is incapable of providing sufficient localization accuracy in indoor environments due to the lack of line of sight (LoS) propagation channel. Therefore, a lot of efforts have been devoted to developing Indoor Positioning Systems (IPSs) in the past two decades [1]–[3]. Among the proposed techniques, WiFi has been acknowledged as the most promising alternative to GPS for indoor localization because commercial off-the-shelf (COTS) WiFi devices and infrastructures are widely available in indoor environments and most of mobile devices (MDs) are equipped with WiFi modules.

Fingerprinting-based localization algorithm is the most widely adopted algorithm for WiFi-based IPS due to its ability to capture signal variances in complex indoor environments more accurately than other algorithms [4]–[6]. However, there are two major drawbacks that restrain them for large-scale implementation. One is that the offline site survey process is extremely time-consuming and labor-intensive. Multiple RSS samples need to be measured at numerous calibration points to ensure localization accuracy. The other is that the offline calibrated database is vulnerable to environmental dynamics [7], as the real-time RSS readings collected during the online localization phase can deviate from those stored in the offline radio map due to variation in temperature, humidity, occupancy distribution and multipath effects. Serious localization errors may be introduced if the radio map is not updated adaptively. Previous works have tried to use an indoor radio propagation model for online radio map construction to replace the laborious offline site survey process [8], [9]. However, the simple log-distance path loss model fails to capture the non-uniform RSS distribution in complex indoor environments. Some works deploy fixed reference anchors to obtain real-time RSS readings for radio map adaptation [10], [11]. Nevertheless, the requirement of extra hardware implementation is the bottleneck of these methods. Learning-based approaches are also introduced to reduce the number of reference anchors to be deployed [12], [13]. However, these methods still need to conduct an offline initialization phase to collect RSS fingerprints as label data for learning purposes. Although certain crowdsourcing methods have been introduced in [14], [15] recently to tackle the issues mentioned, extra user intervention is required. Therefore, an efficient, easily implementable and non-intrusive scheme for online radio map construction and adaptation is urgently needed.

In this paper, we propose, WinIPS, a WiFi-based non-intrusive indoor positioning system that enables automatic online radio map construction and adaptation for calibration-free indoor localization to overcome the aforementioned issues of WiFi fingerprinting-based IPS. For RSS data acquisition, we develop WinSMS, a novel intelligent wireless system that can capture data packets transmitted in existing WiFi traffic

This work is funded by the Republic of Singapore National Research Foundation (NRF) through a grant to the Berkeley Education Alliance for Research in Singapore (BEARS) for the Singapore-Berkeley Building Efficiency and Sustainability in the Tropics (SinBerBEST) Program. BEARS has been established by the University of California, Berkeley as a center for intellectual excellence in research and education in Singapore. The work is also partially funded by Republic of Singapore NRF under grant NRF2013EWT-EIRP04-012 and NRF2011NRF-CRP001-090.

H. Zou, M. Jin, and C. J. Spanos are with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720 USA (e-mail: hanzou, jinming, spanos@berkeley.edu).

L. Xie is with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798 (e-mail: elhxie@ntu.edu.sg).

H. Jiang is with the College of Electrical Engineering and Automation, Fuzhou University, Fuzhou, China (e-mail: jiangh@fzu.edu.cn).

and extract the RSS and MAC addresses of both APs and MDs in a non-intrusive manner without introducing any extra hardware. Since we can obtain the real-time RSS measurements of APs, they become natural online reference points for online radio map construction and adaptation. Therefore, we can completely avoid the tedious offline site survey process. Furthermore, in order to build up a more fine-grained radio map, we propose the Gaussian Process Regression (GPR) with Polynomial Surface Fitting Mean (PSFM-GPR), a reliable regression technique dedicated to predict RSS on virtual reference points (VRPs). It can well capture non-uniform RSS distributions over complex indoor environments. PSFM-GPR models the RSS distribution with a two-dimensional surface which is closer to practical scenarios. Moreover, this online radio map better adapts and is robust to environmental dynamics than the traditional offline calibrated RSS database since it is up-to-date all the time. Since the online radio map is based on AP generated RSS values, it is not suitable for localizing MDs directly due to the device heterogeneity issue. Instead of raw RSS values, we leverage Signal Tendency Index (STI) [16], which compares the shapes of RSS vectors between RSS readings of MD and online RSS fingerprint database. Then, we propose Signal Tendency Index - Weighted K Nearest Neighbor (STI-WKNN), that adopts the similarity index STI as a novel weighting scheme for WKNN, to improve the localization accuracy of WinIPS across heterogeneous devices. Extensive experiments were carried out over a duration of six months to validate the effectiveness of WinIPS in a real-world multi-functional office. The experimental results demonstrate that PSFM-GPR achieves a 4.8 dBm average RSS estimation accuracy and a 1.718m average localization accuracy, which outperforms the existing approaches, such as GPR with Log-Distance Mean (LDM-GPR) [9] and Geography Weighted Regression (GWR) [17]. Furthermore, STI-WKNN improves the localization accuracy by 23.95% over traditional algorithms across heterogeneous MDs.

In summary, we make the following contributions:

- We develop a WiFi-based non-intrusive IPS, WinIPS, that is able to estimate locations of mobile devices without app installation on the user's side.
- For online RSS data acquisition, we design WinSMS to overhear WiFi traffic and extract RSS values and MAC addresses of mobile devices and APs from the data packets in a non-intrusive manner. WinSMS can be directly implemented on COTS WiFi routers, making them natural reference points without introducing any extra hardware infrastructure.
- For online radio map construction and adaptation, we propose PSFM-GPR, which is able to build up and update fine-grained radio map automatically over environmental dynamics and discard the impractical laborious offline site survey process.
- We introduce STI-WKNN that allows WinIPS to provide a high localization accuracy consistently across heterogeneous mobile devices.
- We prototype WinIPS and test it in real complex indoor environment. Promising results indicate that WinIPS makes substantial progress towards fortifying WiFi fingerprint-based IPS for feasible large-scale commercialization.

The rest of the paper is organized as follows. The related work is briefly reviewed in Section II. Section III introduces the detailed system design of WinIPS, as well as the methodologies of WinSMS, PSFM-GPR and STI-WKNN. In Section IV, our experimental testbed and data collection procedure are described first, and experimental results and performance evaluation of WinIPS are then reported. We conclude this paper with Section V.

II. Related Work

In this section, we first present a brief overview on fingerprinting-based localization algorithms and their limitations, and then introduce existing approaches that try to tackle the problems.

A. Limitations of Fingerprinting-based Localization Algorithms

Fingerprinting-based localization algorithms can be classified into two categories: deterministic approaches [4], [11] and probabilistic approaches [18], [19]. Pioneered by RADAR [4], deterministic approaches measure the difference between real-time RSS samples and the mean of RSS fingerprints, calculating the most matched fingerprints. They can provide meter-level localization accuracy with a dense radio map. On the other hand, probabilistic approaches calculate the likelihood between the real-time RSS samples and RSS distributions of fingerprints stored in the database. Statistical techniques such as maximum likelihood estimation [18], maximum a posteriori estimation [19] and Gaussian process [20] are employed to estimate the user location.

Several published results have shown that the fingerprinting-based localization algorithms outperform other methods, such as the time-of-arrival, angle-of-arrival and model-based approaches [1]. Some detailed performance analysis of RSS fingerprinting-based localization algorithms, such as Cramr-Rao lower bound, are elaborated in [21], [22]. There are two major drawbacks of the existing fingerprinting-based algorithms. One is that the offline site survey process is time consuming, labor exhaustive and expensive. In order to achieve sufficient localization accuracy, the WiFi RSS fingerprints from different access points (APs) need to be measured at a huge number of calibration points, which is impractical for large indoor environments such as shopping malls, stadiums and airports. The other drawback is that the offline calibrated RSS fingerprint database is vulnerable to environmental dynamics [7]. RSS is known to be susceptible to various environmental changes including instant interference, such as the opening and closing of doors and moving metal objects, as well as continuous interference, such as variations in temperature, humidity and occupancy distribution. Another source of interference is multipath effects, which include reflection, diffusion and diffraction in indoor environments. As a consequence, the real-time RSS samples collected during the online localization phase can severely deviate from those stored in the offline radio map, leading to serious location errors. In summary, the laborious and prolonged offline site survey process and the vulnerability to environmental dynamics of fingerprinting-based approaches hinder its further commercialization.

B. Radio Map Construction and Adaptation

Several schemes have been proposed to reduce the manual efforts for offline site survey and update the radio map online, including fixed reference anchor methods [10], [11], [17], [23], calibration-free methods [8], [9], learning-based methods [12], [13] and crowdsourcing methods [14], [15].

Specifically, LANDMARC [10] and LEASE [11] developed an adaptive offset of the RSS variations by employing reference anchors deployed at known fixed locations with real-time RSS observations. Nonetheless, these approaches require a very dense deployment of reference anchors to construct the radio map accurately. In [17], self-made WiFi anchors are introduced to obtain real-time RSS observations and Geography Weighted Regression (GWR) is adopted for online radio map construction to reduce the workload for offline site survey. It is noted that all these methods still require extra hardware to be deployed and are infeasible for large-scale implementation. A calibration-free method which uses an indoor radio propagation model for online radio map construction to remove the offline site survey process is presented in [8]. Nevertheless, the simple log-distance path loss model cannot describe the complex RSS distribution precisely. In [9], the idea of employing RSS data among APs to establish a radio map and using the GPR with Log-Distance path loss model for RSS modeling is introduced. However, they fail to modify the AP firmware due to its technical difficulty and instead put wireless monitors beside each AP. As a result, an extra device is still needed.

Several learning-based approaches are also introduced to reduce the number of reference anchors to be deployed [12], [13], [24]. LEMT [12] performed radio map adaptation by training the functional relationship between each location and its neighboring locations based on nonlinear regression analysis and the model tree method, since neighboring locations have highly correlated RSS characteristics in general. The drawback of LEMT is that the process of building huge numbers of trees in each RSS sniffing period is time-consuming, which makes it difficult for real-time applications. Other learning techniques such as multi-view learning [24] and manifold alignment [13] are also utilized to transfer RSS information across different times and devices. Nevertheless, they still need to collect certain numbers of offline RSS fingerprints as label data for learning purpose.

Crowdsourcing methods, which employ the full sensing capabilities of MDs, are introduced to reduce the efforts for radio map construction as well [14], [15]. Zee [14] utilized inertial measurement unit (IMU), comprised of accelerometers, gyroscopes and magnetometers, and RSS reading from the MDs to build up a radio map. Walkie-Markie [15] used landmarks, such as turns, escalators and elevators, to enhance crowdsourcing performance. Nevertheless, extra user intervention is needed for these approaches and continuous IMU monitoring will consume a lot of MDs' batteries, which is an impractical solution.

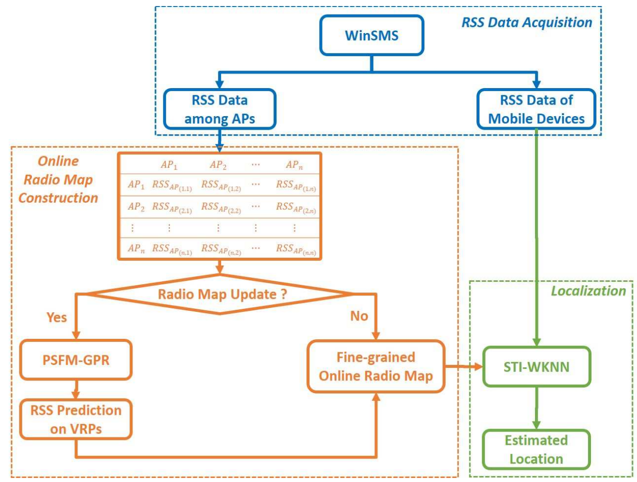


Fig. 1. WinIPS system architecture, illustrating modules of RSS data acquisition, online radio map construction and localization.

III. System Design

A. System Overview

The objective of WinIPS is to realize automatic online radio map construction and adaptation for calibration-free indoor localization. The system architecture of WinIPS is illustrated in Fig. 1. It consists of three main parts: RSS data acquisition, online radio map construction and online localization. For RSS data acquisition, we develop the WiFi-based non-intrusive Sensing and Monitoring System (WinSMS), which enables COTS WiFi APs to intercept the data packets transmitted in the existing WiFi traffic and extract RSS values in a non-intrusive manner without extra hardware infrastructure. All the data will be forwarded to a back-end server for radio map construction and localization. We propose PSFM-GPR, a reliable regression technique dedicated for RSS predictions on each VRP to construct and update a fine-grained online RSS radio map over various environmental dynamics. For online localization, STI-WKNN is adopted to estimate the locations of heterogeneous MDs with consistent high localization accuracy. The users can use any browser on their MDs to obtain the estimated location through the WinIPS Web server without the need of installing an app. The following sections will introduce the methodologies of WinSMS, PSFM-GPR and STI-WKNN, respectively.

B. WinSMS for RSS Data Acquisition

The main drawbacks of fingerprinting-based approaches, the laborious offline site survey process and the vulnerability to environmental dynamics have been elaborated in Section II. In addition, Apple Inc. has not provided any RSS API for third-party developers. Due to these reasons, active WiFi scanning via MD is not a practical method for establishing radio maps anymore. Therefore, it is urgent and indispensable to design a scheme for online RSS radio map construction and adaptation in an accurate, reliable, efficient, practical and non-intrusive manner.

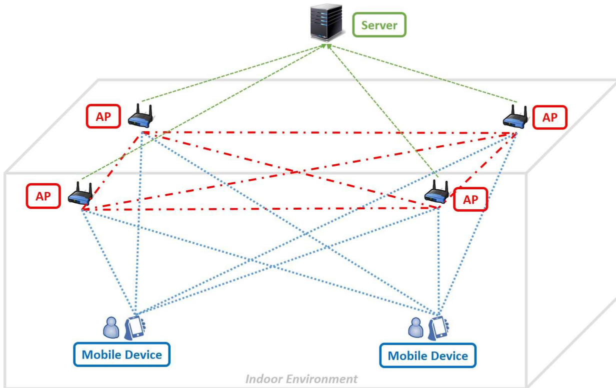


Fig. 2. WinSMS system to collect RSS data from communication among APs and mobile devices.

To overcome this bottleneck, we develop, WinSMS, an intelligent wireless system that enables COTS WiFi APs to overhear the data packets transmitted in the existing WiFi traffic in real-time without any intrusion on the user side. It can be implemented on most of the COTS WiFi routers that support the OpenWrt [25] operating system. WinSMS can create a WiFi LAN to provide basic Internet services for users in its wireless network coverage. More importantly, it has the ability to overhear the data packets transmitted between each MD and WiFi routers, and accurately retrieve the RSS values and corresponding MAC addresses as identifiers. Then, all the information will be sent to a back-end server without requiring user to install any dedicated app for data acquisition.

Fig. 2 presents the system architecture of WinSMS. The main components of WinSMS includes the COTS WiFi APs, a back-end server, as well as users and their MDs. All the APs in WinSMS perform the following major tasks: capture the $802.11n$ data packets in the network, extract relevant information from the packets, arrange them in a particular format and forward them to the back-end server. We upgrade the firmware of APs with OpenWrt and add a designed software based on Libpcap [26] to sniff existing WiFi traffic, and capture as well as analyze the data packets. Unlike traditional active RSS scanning via a MD which has a limited sampling rate, APs are able to overhear sustainable amount of data packets generated by various existing applications on MDs, such as data stream from watching videos, push notification services and periodic email fetching, at the maximum rate around 100 packets per second in a non-intrusive manner. Furthermore, since WinSMS opportunistically captures the data packets from existing WiFi traffic, it poses no additional burden on the battery life of MD. Noticing that usually a person cannot move a significant distance in a second and the RSS value cannot change dramatically in such a short time, the RSS values received within 1 second are averaged out as a pre-filtering step. In this way, the RSS values collected by WinSMS are smoother than those by the active scanning method. The weakest signal strength is set to be -95 dBm. If a particular data packet is received by only one AP, we set the value received by the others as -95 dBm which effectively means that the device is outside the range of that AP. After that, the retrieved RSS values of MDs with their corresponding

TABLE I
ONLINE RSS OBSERVATIONS AMONG APS CAPTURED BY WINSMS.

	AP_1	AP_2	$\cdots$	AP_n
AP_1	$RSS_{AP_{(1,1)}}$	$RSS_{AP_{(1,2)}}$	$\cdots$	$RSS_{AP_{(1,n)}}$
AP_2	$RSS_{AP_{(2,1)}}$	$RSS_{AP_{(2,2)}}$	$\cdots$	$RSS_{AP_{(2,n)}}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
AP_n	$RSS_{AP_{(n,1)}}$	$RSS_{AP_{(n,2)}}$	$\cdots$	$RSS_{AP_{(n,n)}}$

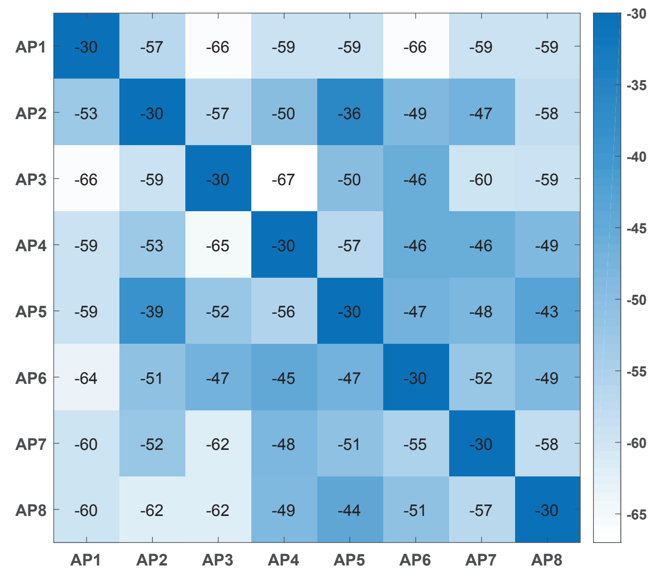


Fig. 3. Visualization of pairwise RSS matrix among 8 APs (dBm). For instance, the RSS measurement between AP5 and AP6 is -47 dBm. As discussed in the main text, we assume the self-sensed RSS of each AP is -30 dBm.

MAC addresses will be sent to the back-end server through the UDP protocol. The server is responsible for parsing the data and building up the online RSS fingerprint database for localization.

For each AP, in addition to capturing the data packets sent and received by each MD, it can overhear packets of other APs as well. Therefore, the RSS measurements at these APs can be leveraged for online map construction. As summarized in Table I, all the APs can be used as natural online reference points for RSS radio map construction and adaptation since we have their physical coordinates and real-time RSS readings. Fig. 3 demonstrates the visualized pairwise RSS of 8 APs. In principle, each AP cannot sense the signal strength of itself. Therefore, we calibrate the average RSS of two APs placed side-by-side and assign -30 dBm as the self-sensed RSS to complete the pairwise RSS matrix of APs as shown in Fig. 3.

As shown in Fig. 4, the RSS values on the limited numbers of APs may not be good enough to describe a fine-grained RSS distribution for each AP. In order to obtain a more fine-grained radio map, we introduce VRPs and propose the PSFM-GPR, a suitable RSS modeling scheme that is able to accurately estimate the RSS values of each AP at predefined VRPs for fine-grained online radio map construction. The methodology of the PSFM-GPR is introduced in the following section.

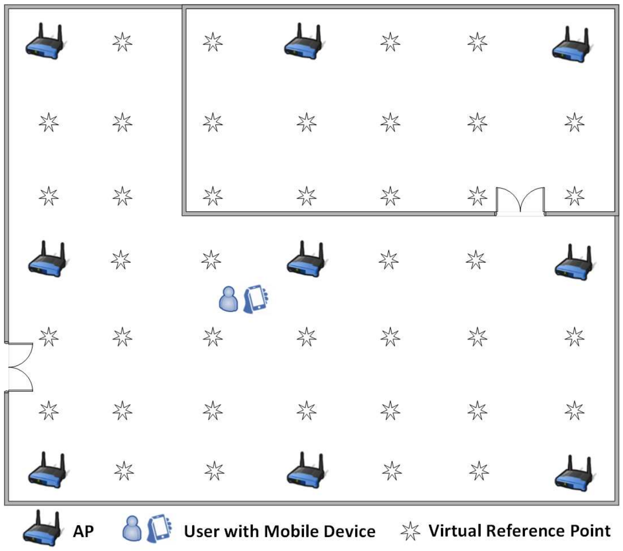


Fig. 4. Scenario of WinIPS implementation in a complex indoor environment. Following the algorithm outlined in Fig. 1, a fine-grained radio map at virtual reference points is updated with WinSMS data to localize users in real-time.

C. PSFM-GPR for Online Radio Map Construction and Adaptation

1) Gaussian Process Regression Model for RSS Modeling: Admittedly, the RSS transmitted from a WiFi AP in a free space is a log linear delay function of the distance. Nevertheless, this property does not hold in practice due to the multipath effects caused by furniture, walls and moving occupants in complex indoor environments. Therefore, the ideal log-distance path loss model is not able to predict the RSS distribution precisely anymore. An efficient and powerful nonlinear approach is required to model the anomalous distribution of RSS values. As a nonparametric nonlinear regression approach, GPR is an appropriate method for capturing the noisy nature of RSS, and predicting RSS values for online dynamic radio map construction and adaptation. In fact, GPR has been widely employed in numerous areas, including geostatistics, spatial smoothing, robotic applications and machine learning for probabilistic modeling, inference and prediction [27]. Moreover, previous works [20], [28] employed GPR for RSS interpolation to reduce the number of reference points during the offline calibration phase.

A Gaussian Process (GP) generates data located at any point of a finite set of random variables $\boldsymbol{Z}$ which follows a joint multivariate Gaussian distribution. It is characterized by its mean function $m(z) = \mathbb{E}[f(z)]$ and the covariance function $k(z, z') = \mathbb{E}[(f(z) - m(z))(f(z') - m(z'))]$, where $z \in \boldsymbol{Z}$. The marginalization property of GP [27] allows us to predict the posterior probability with an unknown input z^* according to some given inputs z and their corresponding observations.

Since the online radio map construction process of each AP is similar, we will explain how to use GPR to predict RSS values of AP_i as an example, where AP_i is one of the n APs. WinSMS enables each AP to scan not only the RSS of MDs, but also the RSS of other APs. Therefore, all the APs are natural online reference points (training points) for radio map construction and adaptation. The corresponding dataset for each AP consist of pairs of $(\boldsymbol{l_i}, s_i)_{i=1}^n, \boldsymbol{l_i} \in \boldsymbol{L}, s_i \in \boldsymbol{S}$, where $\boldsymbol{l_i} = (x_i, y_i)$ is the two-dimensional coordinates of an AP, and s_i is an RSS value of the AP at location $\boldsymbol{l_i}$. The relationship between the two-dimensional space $\boldsymbol{L}$ and RSS $\boldsymbol{S}$ can be modeled as a GP:

$$s_i = f(\boldsymbol{l_i}) + \epsilon_i$$

where ϵ_i is independent and identically distributed (i.i.d.) additive zero-mean Gaussian noise with variance σ_ϵ^2. Assume the RSS observations at each AP can be drawn from the GP:

$$s \sim \mathcal{GP}\big(m(\boldsymbol{l}), k(\boldsymbol{l}, \boldsymbol{l'})\big)$$

where $m(\cdot)$ and $k(\cdot, \cdot)$ represent the mean and covariance function of GP respectively. GP learns the covariance of the training dataset through the kernel covariance function. In our case, the input data are the two-dimensional coordinates. The value of the kernel covariance function is higher when two points are near to each other and lower when two points are far away. We utilize the most popular squared exponential kernel covariance function:

$$k(\boldsymbol{l}, \boldsymbol{l'}) = \sigma_f^2 \exp\left[\frac{-\|\boldsymbol{l} - \boldsymbol{l'}\|^2}{2r^2}\right] + \sigma_\epsilon^2 \delta(\boldsymbol{l}, \boldsymbol{l'}), \tag{1}$$

where σ_f^2 and r are the hyperparameters of GP and $\delta(\cdot, \cdot)$ denotes the Kronecker delta function. Since we have n APs in the space, we can calculate the covariance of each pair of APs according to Equation (1) and obtain the $n \times n$ covariance matrix $K(\boldsymbol{L}, \boldsymbol{L})$ for all pairs of training data. Suppose that we would like to predict RSS values $\{s_j\}_{j=1}^m \in \boldsymbol{S}^*$ of AP_i at m VRPs $\{\boldsymbol{l_j}\}_{j=1}^m \in \boldsymbol{L}^*$ to build up a fine-grained radio map. The multivariate Gaussian distribution of training data and predicted RSSs with a zero-mean distribution can be described as follows:

$$\begin{bmatrix} \boldsymbol{S} \\ \boldsymbol{S}^* \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} K(\boldsymbol{L}, \boldsymbol{L}) + \sigma_\epsilon^2 \boldsymbol{I} & K(\boldsymbol{L}, \boldsymbol{L}^*) \\ K(\boldsymbol{L}^*, \boldsymbol{L}) & K(\boldsymbol{L}^*, \boldsymbol{L}^*) \end{bmatrix}\right),$$

where $K(\boldsymbol{L}, \boldsymbol{L}^*)$ is an $n \times m$ covariance matrix between $\boldsymbol{S}$ and $\boldsymbol{S}^*$, and $\boldsymbol{I}$ is the identical matrix. The RSS value of this AP at an interested point $\boldsymbol{l_j}$ can be predicted according to the posterior mean and variance of GP:

$$\begin{aligned} \bar{s}_j &= K(\boldsymbol{l_j}, \boldsymbol{L})[K(\boldsymbol{L}, \boldsymbol{L}) + \sigma_\epsilon^2 \boldsymbol{I}]^{-1} \boldsymbol{S}, \\ cov(s_j) &= K(\boldsymbol{l_j}, \boldsymbol{l_j}) - K(\boldsymbol{l_j}, \boldsymbol{L})[K(\boldsymbol{L}, \boldsymbol{L}) + \sigma_n^2 \boldsymbol{I}]^{-1} K(\boldsymbol{L}, \boldsymbol{l_j}), \end{aligned} \tag{2}$$

where $\bar{s}_j$ is the estimated mean RSS at this location, $cov(s_j)$ denotes the posterior variance as an estimation confidence indicator, and $K(\boldsymbol{l_j}, \boldsymbol{L})$ and $K(\boldsymbol{L}, \boldsymbol{l_j})$ are $1 \times n$ and $n \times 1$ matrices of the covariance between this point and all training points.

As shown in Equation (2), the GP model usually adopts the zero-mean function (ZeroM-GPR) as the default settings, which means that the estimated RSS values will tend to zero at locations that are far from any training points (APs). This is obviously impractical for RSS modeling. Previous works [9] used the Log-Distance path loss model to obtain a general mean of RSS and then made use of GPR to estimate the

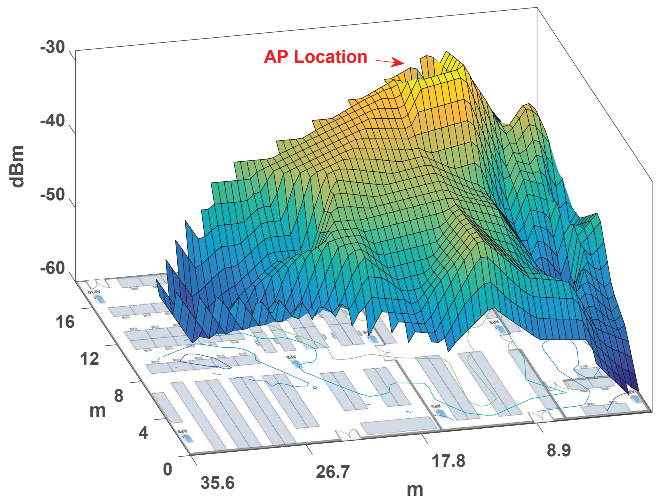


Fig. 5. Illustration of RSS distribution of an AP in a complex indoor environment, where the AP location is marked.

residual RSS errors. The estimated RSS at an arbitrary location $\boldsymbol{l_j}$ is calculated by

$$\bar{s}_j = m(\boldsymbol{l_j}) + K(\boldsymbol{l_j}, \boldsymbol{L})[K(\boldsymbol{L}, \boldsymbol{L}) + \sigma_\epsilon^2 I]^{-1}(\boldsymbol{S} - m(\boldsymbol{L})),$$
$$m(\boldsymbol{l_j}) = PL_0 + 10\alpha \log(\|\boldsymbol{l_j} - \boldsymbol{l_{AP_i}}\|/d_0) + \mathcal{X}, \quad (3)$$

where $\|\boldsymbol{l_j} - \boldsymbol{l_{AP_i}}\|$ indicates the distance between AP_i and location $\boldsymbol{l_j}$, PL_0 is the path loss coefficient of the RSS value at initial distance d_0, and α is the path loss gradient, and $\mathcal{X}$ represents the lognormal shadow fading with zero mean noise with standard deviation $\sigma_{\mathcal{X}}$ [29]. These three parameters of the Log-Distance path loss model in Equation (3) are calculated by curve fitting with the training points. The Log-Distance Mean GPR (LDM-GPR) can be used to estimate the RSS distribution in an open space because it describes the relationship between RSS and distance. However, in practice, as shown in Fig. 5, the RSS distribution is much more complicated. The RSS values at the same distance from the AP are usually distinct due to multi path effect and shadow fading results from the obstacles that attenuate signal power through absorption, reection, scattering, and diffraction in complex indoor environment. Hence, the LDM-GPR is no longer suitable since it does not consider the orientation or the surrounding environmental property on each VRP.

2) Online Radio Map Construction with PSFM-GPR: In order to address this issue, we propose the PSFM-GPR, which utilizes a two-dimensional polynomial surface fitting model to estimate the general mean of the RSS, and then utilizes GP to estimate the residual RSS errors. First of all, we assume the RSS distribution of AP_i to be a two-dimensional polynomial function as follows:

$$m(\boldsymbol{l}) = \beta_0 + \beta_1 x + \beta_2 y + \beta_3 x^2 + \beta_4 y^2 + \beta_5 xy \quad (4)$$

where $\boldsymbol{l} = (x, y)$ denotes the coordinates of other APs. Since the WinSMS can obtain the RSS values of AP_i at all other APs' locations, all the parameters $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ in Equation (4) can be estimated and updated online using two degree polynomial surface fitting. According to our data analysis regarding the fitting accuracy and the computational overhead, we found that two-degree polynomial surface fitting is good enough to capture the non-uniform RSS distribution. With this proper mean of RSS, the predicted RSS by the PSFM-GPR at any arbitrary location $\boldsymbol{l_j}$ is calculated by

$$s_j = m(\boldsymbol{l_j}) + K(\boldsymbol{l_j}, \boldsymbol{L})[K(\boldsymbol{L}, \boldsymbol{L}) + \sigma_\epsilon^2 I]^{-1}(\boldsymbol{S} - m(\boldsymbol{L})), \quad (5)$$

$$m(\boldsymbol{l_j}) = \beta_0 + \beta_1 x_j + \beta_2 y_j + \beta_3 x_j^2 + \beta_4 y_j^2 + \beta_5 x_j y_j, \quad (6)$$

where (x_j, y_j) are the coordinates of location $\boldsymbol{l_j}$.

After estimating the RSS values of all the n APs by PSFM-GPR at the m VRPs, we can obtain a RSS vector $\mathbf{s}_j = [s_j^1, s_j^2, ..., s_j^n]$, where $1 \leq j \leq m$ and s_i^j $(1 \leq i \leq n)$ denotes the RSS values from AP_i at each VRP $\boldsymbol{l_j}$. Therefore, a $m \times n$ RSS fingerprint database can be effectively built up online to avoid the cumbersome offline site survey process.

3) Online Radio Map Adaptation with PSFM-GPR: The radio map adaptation is another crucial process of WinIPS system because it keeps the radio map up-to-date automatically over various contextual dynamics including time and space. Since the WinIPS system can obtain the RSS values of all APs in real time as presented in Table I, each column in $n \times n$ RSS matrix can be used as a trigger to determine whether the system should initiate the radio map adaptation process for each AP.

Algorithm 1 Online radio map adaptation algorithm

Initialization:
Input: n - The total number of APs
m - The total number of VRPs
$\mathbf{s}^{t-1}$ - $n \times n$ RSS matrix of AP as shown in Table 1
s_i^{t-1} - The RSS vector of AP_i stored in the fingerprint database
s_i^t - The RSS vector of AP_i at the time t
θ_{th} - The RSS threshold for AP RSSI differences
Output: $\mathbf{s}_f^t$ - $m \times n$ Up-to-date RSS fingerprint database at time t
Check RSS profile of each AP:
for $i = 1, \cdots, n$ **do**
    if $\|s_i^t - s_i^{t-1}\| > \theta_{th}$ **then**
        RSS profile of AP_i is required to update
        $AP_i \in AP_Q$
    else
        RSS profile of AP_i is up-to-date
    end if
end for
Update RSS fingerprint database:
for $q = 1, \cdots, Q$ **do**
    $AP_q \rightarrow PSFM - GPR$ to predict RSS on all VRPs
    for $j = 1, \cdots, m$ **do**
        $s^j = s^j_{AP_q}$
    end for
end for
return $\mathbf{s}_f^t$

The detailed procedure of radio map adaptation is presented in Algorithm 1. First of all, we will compare the differences

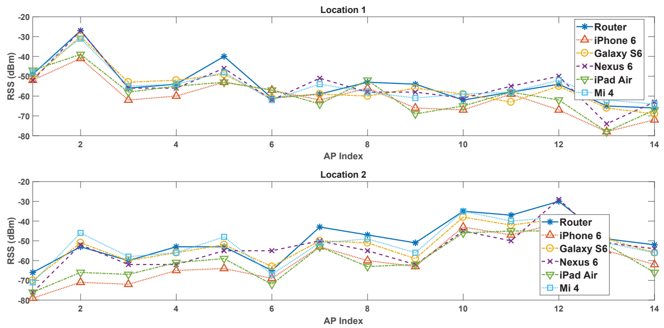


Fig. 6. Demonstration of RSS for a router and mobile devices, located at Location 1 (top plot) and Location 2 (bottom plot), measured by 14 APs (indexed in x-axis). The shapes of router and mobile device RSS curves display similar patterns.

between the real-time RSS values s_i^t and the RSS profile s_i^{t-1} stored in the database for all the APs. If the RSS distance between these two RSS vectors $\|s_i^t - s_i^{t-1}\|$ is larger than a RSS threshold θ_{th}, it implies that the RSS profile of AP_i is outdated due to some indoor environmental dynamics, and the radio map update procedure will be initiated for this AP. According to our empirical study, we set the threshold θ_{th} to 10dBm. The RSS values from this AP at each VRP will be updated by the PSFM-GPR scheme as introduced in Section III-C2. In this way, the $m \times n$ online RSS fingerprint database will be up-to-date and be more robust to various contextual dynamics compared to traditional offline fingerprint database.

D. STI-WKNN Localization Algorithm for Heterogeneous Mobile Device

As introduced in the aforementioned sections, an up-to-date and fine-grained online RSS fingerprint database is obtained using WinSMS and PSFM-GPR. However, the RSS values stored in this database are collected at APs. This database cannot be applied directly for localization of MDs because the RSS signatures of AP and MDs are usually different due to various heterogeneous factors such as distinct WiFi chipsets, WiFi antennas, hardware driver, and even operating systems [30], [31]. To illustrate this issue, we conducted an experiment that collected 500 RSS samples from a TP-Link TL-WR703N portable router, as well as five different MDs: iPhone 6, Galaxy S6, Nexus 6, iPad Air and Mi 4 at two identical locations with respect to 14 commodity WiFi APs in a complex indoor environment. As observed in Fig. 6, each curve connects the average RSSs between one device and 14 APs. The RSSs associated with router and MDs are significantly different, which verifies the effect of device heterogeneity. Therefore, the localization accuracy will be severely jeopardized if we employ the RSS fingerprint database of a router (AP) to estimate the location of an MD directly.

Meanwhile, another noteworthy observation from Fig. 6 is that the shapes of the curves display certain similarities. In other words, one curve can be roughly recovered from another one via translation and scale operations. Thus, to accommodate the device heterogeneity issue, we leverage the Signal Tendency Index (STI) [16], which compares the similarities of the RSS curve shapes by using the ordinary Procrustes Analysis (PA) method [32] instead of using the raw RSS for fingerprint matching. To be specific, given a real-time RSS vector from a MD, s_d, the translation step of the ordinary PA method will produce

$$s_d^1 - \overline{s}_d, s_d^2 - \overline{s}_d, ..., s_d^n - \overline{s}_d \tag{7}$$

where

$$\overline{s}_d = \frac{1}{n}\sum_{i=1}^{n} s_d^i.$$

Then, in the uniform scaling step, we have

$$\widehat{s}_d = [s_d^1 - \overline{s}_d, s_d^2 - \overline{s}_d, ..., s_d^n - \overline{s}_d]/\hat{\sigma}, \tag{8}$$

where

$$\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(s_d^i - \overline{s}_d)^2}.$$

The $\widehat{s}_d$ is the transformed object of ordinary PA method. Similarly, all the AP-based RSS vectors stored in the fingerprint database will be transformed as well. All the transformed RSS fingerprints $\{\widehat{s}^j\}_{j=1}^m$ will be compared with $\widehat{s}_d$ in terms of their shape similarity. We define the Procrustes distance between the two vectors $\widehat{s}_d$ and $\widehat{s}^j$, termed signal tendency index (STI), which is computed by

$$STI^j = \|\widehat{s}_d - \widehat{s}^j\| \tag{9}$$

where $\|\cdot\|$ denotes the Euclidean norm. After that, we introduce a new weighting scheme which involves STI and integrate it with the classical localization algorithm, Weighted K Nearest Neighbor (WKNN), namely STI-WKNN, instead of using the distance of RSS vectors as the weights. Since we have calculated the STI value STI^j between s_d and each s^j, a smaller STI^j indicates that s^j is similar to s_d. We further define a weight value w^j for each s^j, which is calculated as follows:

$$w^j = \frac{\frac{1}{STI^j}}{\sum_{j=1}^{m}\frac{1}{STI^j}} \tag{10}$$

Then, the m VRPs are sorted according to their w^j in a descending order. Only top K VRPs and their corresponding physical coordinates are adopted to estimate the location of MD (x_d, y_d), which is calculated by:

$$(x_d, y_d) = \frac{1}{c}\sum_{k=1}^{K}(x_k, y_k) \cdot w^k \tag{11}$$

where (x_k, y_k) denotes the coordinates of ith VRP and $c = \sum_{k=1}^{K} w^k$ is the normalization constant.

In summary, the STI-WKNN localization scheme first compares the similarities of the RSS curve shapes between real-time RSS vector of a MD and those stored in the fingerprint database by the ordinary PA method. Then, the similarity index STI is adopted as a novel weighting scheme for WKNN to estimate the location of heterogeneous MDs.

(a) Layout of the testbed at T_1.

(b) Layout of the testbed at T_2.

Fig. 7. Layout of the testbed (a) at the beginning of the experiment (T_1), and (b) six months later after renovation (T_2), showing significant indoor structure changes (shaded region).

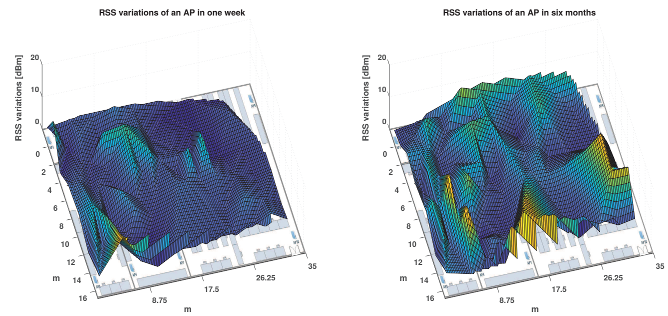


Fig. 8. Comparison of RSS variations (z-axis) in short-term (left) and long-term (right), showing much larger long-term variation.

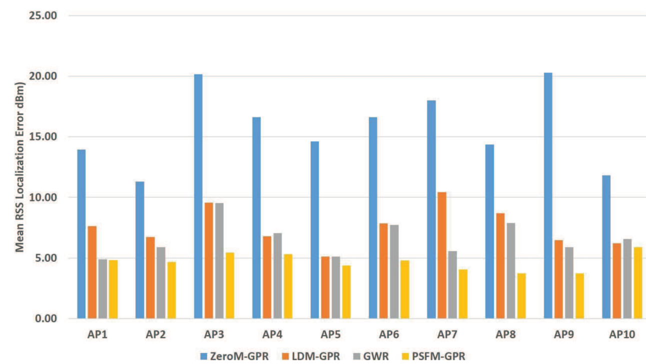


Fig. 9. Comparison of RSS estimation errors for different APs by ZeroM-GPR [9], LDM-GPR [9], GWR [17] and PSFM-GPR.

IV. Experimental Results and Discussions

A. Experimental Setup

To validate the performance of WinIPS comprehensively, extensive experiments were conducted in a 35.6×16.6 m multi-functional lab for six months. The layout of the testbed at the beginning of the experiment (labeled as T_1) is depicted in Fig. 7(a), while the layout after renovation six months later (labeled as T_2) is presented in Fig. 7(b). As shown in Fig. 7, there are several obvious layout differences during the experiment which definitely affect the RSS distribution in the area. We leveraged these changes to verify the radio map adaptation and localization performance of WinIPS under environmental dynamics. This is different from the traditional evaluation methods [9], [17] which usually adopt corridors or open spaces as testbeds, that are favorable for distance-related RSS modeling. As demonstrated in Fig. 7, our testbed includes workspaces, cubical offices, an open space for Unmanned Aerial Vehicle (UAV) testing and a discussion room. This complex indoor environment is much more suitable than an ideal environment for performance evaluation of WinIPS.

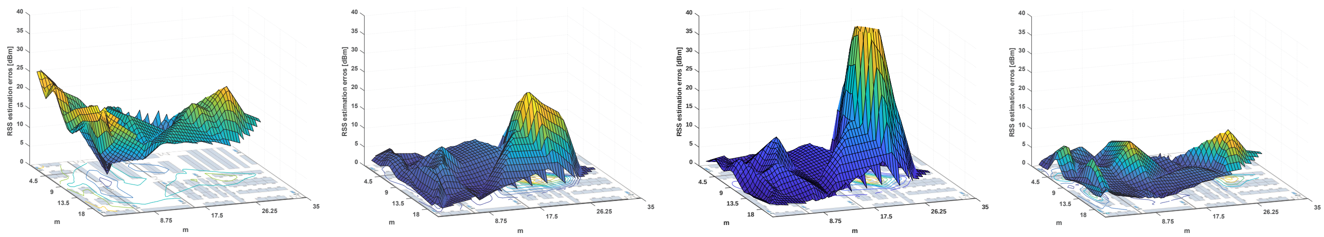
In our experiments, 10 TP-LINK TL-WR703N router, were adopted as APs for WinSMS in our experiment. TLWR703N has a 400 MHz Atheros AR7240 CPU with 4 MB flash memory and 32 MB RAM. The Atheros AR9331 chipset is used in its platform working on 2.4 GHz. To implement WinSMS, we upgraded the firmware to OpenWrt and added our designed software. As shown in Fig. 7, the TL-WR703N nano router is small in size and extremely easy to be deployed. We chose this router to demonstrate that commercial routers are becoming portable and easier for installation nowadays. Moreover, with the booming development of Internet of Things (IoT), billions of IoT devices will be densely deployed in indoor environments for various purposes in the near future. Equipped with WiFi modules, they can be easily upgraded to serve as online reference points for dynamic radio map construction and adaptation. The locations of these 10 APs are depicted in Fig. 7 and they were fixed on 1.9-meter-high tripods to keep them on the same height level. One server is employed to process the RSS data sent by APs, construct and update the RSS radio map and fingerprint database by PSFM-GPR, and adopts STI-WKNN to estimate the location of each MD. 50 testing points (small red circles in Fig. 7) were randomly selected to evaluate the performance of WinIPS. To validate the RSS estimation accuracy of PSFM-GPR, we collected the real RSS values of a TL-WR703N router at these points as the ground truth. Furthermore, we also collected the RSS measurements of five MDs, including iPhone 6, Galaxy S6, Nexus 6, iPad Air and Mi 4, at all the testing points to evaluate the localization accuracy of STI-WKNN across heterogeneous devices.

B. RSS Estimation Accuracy

Firstly, we conducted an experiment to continuously monitor the distribution of RSS variations of an AP (AP10) to understand the fluctuations of RSS caused by various environmental dynamics in six months. Fig. 8 demonstrates the distribution of RSS variation of the AP over one week and six months. As shown in Fig. 8, the RSS variation over six

TABLE II
COMPARISON OF RSS ESTIMATION ERRORS FOR DIFFERENT METHODS, SHOWING THE MEAN ($\bar{e}_{RSS}$) AND THE STANDARD DEVIATION (σ_{RSS}) (DBM).

Method	AP1	AP2	AP3	AP4	AP5	AP6	AP7	AP8	AP9	AP10	$\bar{e}_{RSS}$	σ_{RSS}
ZeroM-GPR	13.94	11.31	20.15	16.62	14.62	16.62	17.99	14.36	20.30	11.80	15.77	8.00
LDM-GPR	7.64	6.71	9.57	6.78	5.12	7.86	10.44	8.69	6.46	6.22	7.55	6.59
GWR	4.90	5.89	9.52	7.05	5.10	7.72	5.57	7.89	5.89	6.56	6.61	5.85
PSFM-GPR	4.82	4.65	5.43	5.30	4.39	4.81	4.07	3.72	3.74	5.88	**4.68**	**3.51**



(a) ZeroM-GPR (b) LDM-GPR (c) GWR (d) PSFM-GPR

Fig. 10. Comparison of RSS estimation errors for AP8 by different methods, showing the spatial distribution.

months (long-term) is much larger than over one week (short-term). Thus, it indicates that the static radio map calibrated at a particular time is definitely unable to serve as the reference for consistent location estimation at all times, since the real time RSS values can vary significantly, especially in long-term deployments. Radio map adaptation strategies such as WinSMS is urgently desired to make the IPS resilient to environmental dynamics.

For WinSMS, the real-time RSS measurements among the 10 APs collected by it can be summarized as a 10×10 RSS matrix, which is similar to Table I. By employing these data, we predicted RSS values from APs by using PSFM-GPR at the 50 testing points and compared it with the observed RSS (ground truth). Fig. 9 and Table II compared the RSS estimation of PSFM-GPR with ZeroM-GPR, LDM-GPR [9] and GWR [17] in terms of mean ($\bar{e}_{RSS}$) and standard deviation (σ_{RSS}) of the RSS estimation error. The average estimated RSS error of PSFM-GPR is 4.68 dBm which is the smallest among the four methods. It is able to reduce the average RSS error by 74.24%, 28.60%, and 29.58% compared to ZeroM-GPR, LDM-GPR and GWR respectively. Moreover, the standard deviation of RSS error of PSFM-GPR is also the smallest among the four methods, indicating that RSS predicted by PSFM-GPR are more stable than existing approaches.

Furthermore, we evaluated the RSS estimation accuracy of PSFM-GPR in two-dimensional space. To illustrate, Fig. 10 describes the estimated RSS error distribution of AP8 from 4 different RSS modeling methods. As illustrated in Fig. 10(d), most of RSS errors of PSFM-GPR are smaller than 10 dBm and are distributed evenly in a low RSS error level. The reason for such an outstanding performance is that PSFM-GPR performs two-dimensional surface fitting for RSS predictions, which well captures non-uniform RSS distributions in a different orientation. In contrast, the RSS errors of ZeroM-GPR is highest especially at the locations far away from any AP (online reference points). On the other hand, LDM-GPR and GWR, failed to capture the non-uniform RSS distribution in complex indoor environments because only the relationship between RSS and distance in RSS modeling is considered.

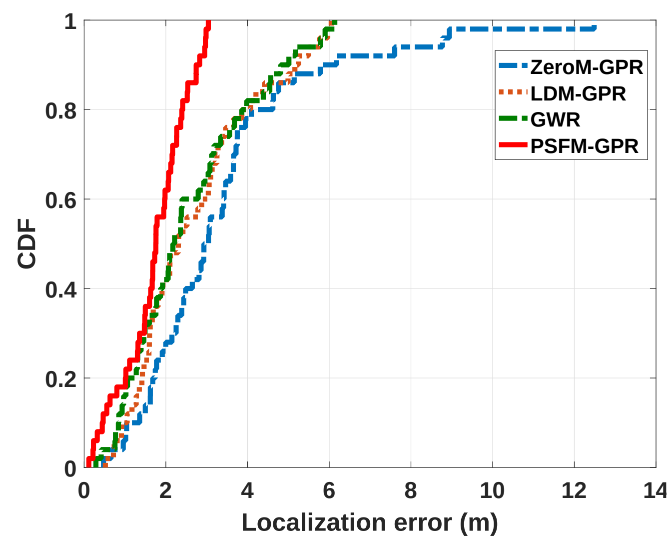
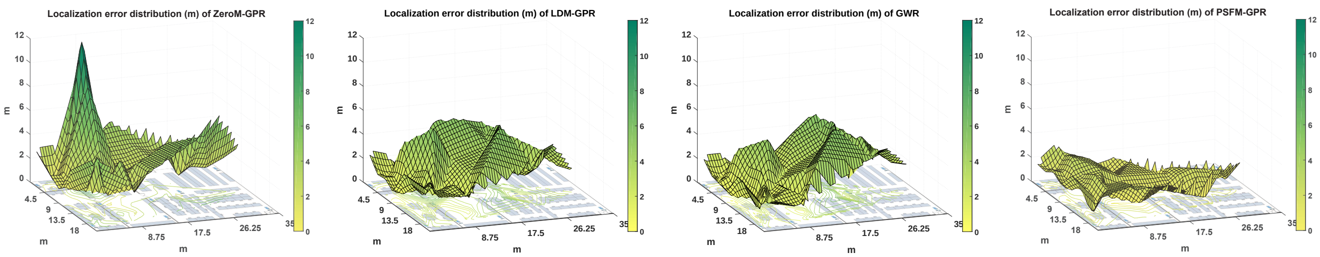


Fig. 11. Cumulative distributions of localization error between different methods.

C. *Localization Estimation Accuracy*

The aforementioned section illustrates the RSS estimation evaluation of WinIPS. We present the localization accuracy evaluation of WinIPS in this section. To prepare a fine-grained online RSS fingerprint database, the back-end server virtually divided our testbed into a 1.48×1.38 m grid and adopted the PSFM-GPR to predict RSS values from all APs at the 288 grid points (VRPs). The grid spacing between two adjacent VRPs was chosen to be around 1.5m according to the analysis in [33]. For evaluation, we collected 500 RSS samples of each MD at each testing point, and used the average location estimated by STI-WKNN to compare with the physical location of each testing point (ground truth).

1) Comparison of Localization Accuracy Between Different Online RSS Prediction Methods: First of all, we evaluate the impacts of different online RSS prediction methods on localization accuracy. In the back-end server, we established three



(a) ZeroM-GPR (b) LDM-GPR (c) GWR (d) PSFM-GPR

Fig. 12. Spatial distributions of localization errors (z-axis) by different methods.

TABLE III
COMPARISON OF LOCALIZATION ERRORS FOR DIFFERENT METHODS IN TERMS OF MEAN ($\bar{e}_{LA}$) AND STANDARD DEVIATION (σ_{LA}). THE IMPROVEMENT PERCENTAGE IS BASE-LINED AGAINST ZEROM-GPR PERFORMANCE.

Method	$\bar{e}_{LA}$ (m)	Improve (%)	σ_{LA} (m)	Improve (%)
ZeroM-GPR	3.389	-	2.264	-
LDM-GPR	2.689	20.66	1.509	33.35
GWR	2.560	24.46	1.539	32.02
PSFM-GPR	**1.718**	**49.31**	**0.803**	**64.53**

online RSS fingerprint databases using ZeroM-GPR, LDM-GPR and GWR similarly to that of PSFM-GPR. STI-WKNN was used as the localization algorithm for all the schemes in this evaluation to make a fair comparison. The statistical attributes (i.e., the mean ($\bar{e}_{LA}$) and standard deviation (σ_{LA}) of localization accuracy) via PSFM-GPR is compared with three other existing approaches. The overall performance is summarized in Table III and Fig. 11. It is evident from Table III that the localization accuracy of WinIPS is much higher when PSFM-GPR is adopted for RSS prediction on VRPs. Fig. 12 depicts the distance error distribution in 2D over the floor plan of the four approaches. Similar to the results shown in Table III, PSFM-GPR has the best performance among the four approaches. PSFM-GPR + STI-WKNN can provide a 1.718m average localization accuracy with the smallest $\sigma_{LA} = 0.803$. It enhances the precision of indoor positioning by 45.52% over ZeroM-GPR, 33.16% over LDM-GPR and 35.23% over GWR respectively. Furthermore, the smallest σ_{LA} indicates that the online RSS fingerprint database generated by the PSFM-GPR can provide more useful information for reliable localization service than the other approaches.

We also explored potential correlations between the RSS estimation accuracy and the localization accuracy using PSFM-GPR. Fig. 13 and Table IV compares the RSS estimation accuracy in terms of mean ($\bar{e}_{RSS}$) and the standard deviation (σ_{RSS}) and the localization accuracy in terms of mean ($\bar{e}_{LA}$) and standard deviation (σ_{LA}) when different number of APs are utilized. According to the analysis presented in [29], the RSS variation is proportional to the square of the distance between routers (router density). It can be seen from Fig. 13 that, RSS estimation errors become smaller when more APs are leveraged. Due to multi path effect and shadow fading results from the obstacles that attenuate signal power through absorption, reection, scattering, and diffraction in complex indoor environment, the empirical results as shown in Fig. 13 and Table IVmay not perfectly match with the theoretical

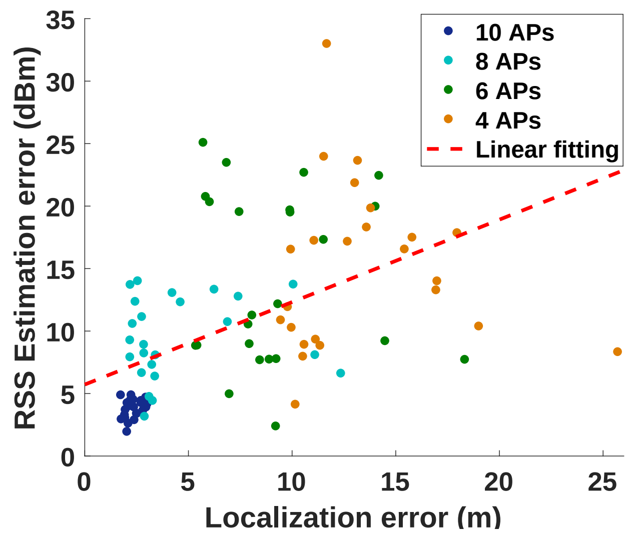


Fig. 13. Plots of the RSS estimation error vs. the localization error, color coded by the number of APs in use, showing higher RSS estimation accuracy leads to more accurate localization.

TABLE IV
COMPARISON OF THE RSS ESTIMATION ACCURACY IN TERMS OF MEAN ($\bar{e}_{RSS}$) AND THE STANDARD DEVIATION (σ_{RSS}) AND THE LOCALIZATION ACCURACY IN TERMS OF MEAN ($\bar{e}_{LA}$) AND STANDARD DEVIATION (σ_{LA}) USING DIFFERENT NUMBER OF APS.

No. of AP	$\bar{e}_{RSS}$ (dBm)	σ_{RSS} (dBm)	$\bar{e}_{LA}$ (m)	σ_{LA} (m)
4	16.16	13.04	9.245	5.187
6	13.24	10.23	5.953	4.010
8	8.58	7.22	2.929	2.575
10	4.68	3.51	1.718	0.803

analysis as presented in [29]. However, as the result of the linear fitting, the general traces of the empirical results are similar to the theoretical results, which validates that higher router density lead to smaller variances of localization error. Thus, we conclude that there is a positive correlation between RSS estimation accuracy and localization accuracy for the PSFM-GPR. Another noteworthy point is that the results in Table III are comparable to those reported in [20] which rely on a cumbersome offline calibrated RSS fingerprint database.

2) Comparison between Traditional Offline Site Survey and PSFM-GPR: In this section, we compare the localization performance of the PSFM-GPR to the traditional offline site survey method. We collected real RSS measurements of the MDs on the physical coordinates of each VRP to build up the offline

TABLE V
COMPARISON OF LOCALIZATION ERROR IN TERMS OF MEAN ($\bar{e}_{LA}$) AND STANDARD DEVIATION (σ_{LA}) BETWEEN OFFLINE SITE SURVEY AND PSFM-GPR.

Method	$\bar{e}_{LA}$ **(m)**	σ_{LA} **(m)**
Offline Site Survey (T_1)	2.327	1.471
Offline Site Survey (T_2)	1.569	1.097
PSFM-GPR (T_2)	1.718	0.803

RSS fingerprint database. Although WinSMS is able to collect RSS values at a fast speed (0.5 seconds/sample), we still spent 5 hours to complete the offline site survey process which is truly time-consuming and labor-intensive. We performed two offline site surveys and constructed the corresponding fingerprint databases at the beginning of the experiment (T_1) and six months later (T_2). The testing data was collected on T_2. The overall performance is presented in Table V. When the up-to-date online RSS fingerprint database generated by PSFM-GPR is compared with the offline database constructed on the same day (T_2), the average localization accuracy of it is only a little worse by 8.67% than the offline calibrated RSS fingerprint database. However, it is impractical to build up an offline radio map every day for localization purposes. To illustrate the vulnerability of the transitional offline site survey method to environmental dynamics, we compared the performance of an out-of-date offline RSS fingerprint database (T_1) to PSFM-GPR. Under this situation, PSFM-GPR reduces the localization error by 26.17% compared to the outdated offline RSS fingerprint database. In summary, PSFM-GPR can construct and update the RSS fingerprint database automatically that enables WinIPS to provide consistent high localization accuracy over various environmental dynamics. Furthermore, it avoids the cumbersome offline site survey process which is the major bottleneck for the large-scale commercialization of WiFi-based IPS.

3) Impact of Device Heterogeneity: To validate the effectiveness of WinIPS under the impact of device heterogeneity, we collected RSS measurements of five MDs at 50 testing points for this evaluation. The overall results are summarized in Table VI. As observed from Table VI, WinIPS can provide a high localization accuracy (within 2m on average) consistently across heterogeneous MDs using STI-WKNN. Although the online RSS fingerprint database established by the PSFM-GPR is based on data among APs, the device heterogeneity issue can be largely alleviated by comparing the similarities of RSS curve shapes (STI) rather than the raw RSS values for WKNN fingerprint matching. Fig. 14 depicts the distance error distribution of the original WKNN and STI-WKNN. STI-WKNN has a much better performance in terms of localization accuracy compared to the original WKNN. It improves localization accuracy by 23.95% over the original WKNN across heterogeneous MDs. In summary, the merit of STI-WKNN enhances the robustness of WinIPS to device heterogeneity issues for indoor localization.

4) Impact of Occupancy Density: We also analyze the impact of occupancy density on the localization accuracy of WinIPS. Fig. 15 demonstrates the functionality of each zone in the testbed. Our testbed is a real multi-functional lab, that includes one undergraduate student office (for 20 occupants),

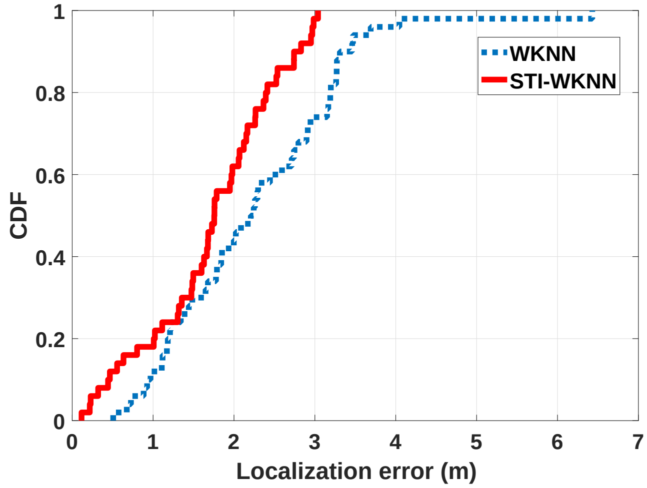


Fig. 14. Cumulative distribution of localization error with (STI-WKNN) and without (WKNN) to compensate device heterogeneity.

TABLE VI
COMPARISON OF LOCALIZATION ERROR IN TERMS OF MEAN ($\bar{e}_{LA}$) AND STANDARD DEVIATION (σ_{LA}) BETWEEN DIFFERENT MOBILE DEVICES.

Mobile device	$\bar{e}_{LA}$ **(m)**	σ_{LA} **(m)**
iPhone 6	1.749	0.733
Galaxy S6	1.683	0.794
Nexus 6	1.785	0.729
iPad Air	1.615	0.912
Mi 4	1.758	0.847

one workplace for graduate students (for 7 occupants), one UAV testbed (open space), one It includes one undergraduate student office (for 20 occupants), one workplace for graduate students (for 7 occupants), one Unmanned Aerial Vehicle (UAV) testbed (open space), one workplace for undergraduate students (for 15 occupants), and one graduate student office (for 45 occupants). As shown in the Fig. 15, the graduate student office is the most crowded area in the lab, where the occupancy density is 0.278 p/m^2. The functionality of the UAV testbed is testing the performance of UAV so it is usually empty, which has the lowest occupancy density within the lab.

Table VII elaborates the zone size, number of common occupants, occupancy density and the mean localization error in each zone. As presented in Table VII, the localization accuracy in low occupancy density area is slightly better in the high occupancy density area. The mean localization error in the graduate student office is the largest (1.945m) and it is

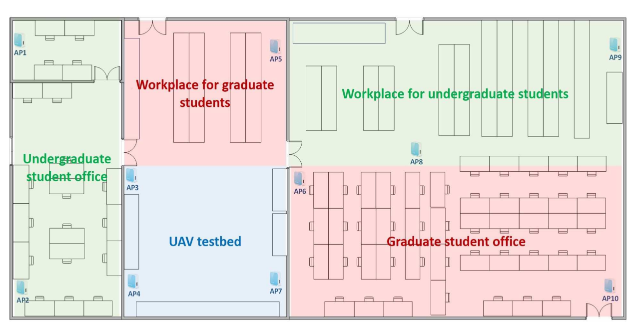


Fig. 15. The functionality of each zone in the testbed.

TABLE VII
COMPARISON OF LOCALIZATION ERROR IN EACH ZONE WITH DIFFERENT OCCUPANCY DENSITY.

Zone ID	Zone area (m^2)	No. of occupants	Occupancy density (p/m^2)	Mean localization error (m)
UAV testbed	80.10	0	0	1.227
Workplace for graduate students	77.36	7	0.09	1.691
Workplace for undergraduate students	159.86	15	0.09	1.464
Undergraduate student office	105.58	20	0.19	1.778
Graduate student office	161.57	45	0.28	1.945

the most crowded area in the lab, while the mean localization error in the UAV testbed is the minimum (1.227m) which has the lowest occupancy density. Thus, the experimental results indicate that higher occupancy density may affect the localization accuracy because the movement of occupants interfere the signal propagation paths and the multipath components, which contribute to higher uctuations of received signals. Potential solution to overcome this issue is to add additional WiFi routers in the crowded area to provide more RSSI measurements and features so that the localization accuracy can still be guaranteed. Furthermore, the localization accuracy in this can be further improved by optimizing the placement of APs in our previous work [34].

V. CONCLUSION

In this paper, we proposed, WinIPS, a WiFi-based non-intrusive IPS that enables automatic online radio map construction and adaptation for calibration-free indoor localization. For RSS data acquisition, we developed WinSMS, a novel intelligent wireless system that can capture data packets transmitted in the existing WiFi traffic and extract the RSS and MAC addresses of both APs and MDs in a non-intrusive manner. We leverage APs as natural online reference points for online radio map construction and adaptation. To construct a more fine-grained radio map, we further proposed the PSFM-GPR, a reliable regression technique dedicated to predicting RSS on VRPs which can well capture the non-uniform RSS distribution over complex indoor environments. The online radio map adapts better and is more robust to environmental dynamics than traditional offline calibrated RSS database since it keeps updated with new measurements. To alleviate the device heterogeneity issue between AP and MD, we introduced STI-WKNN, which compares the shapes of RSS vectors between RSS readings of MDs to online RSS fingerprint database rather than to raw RSS values. Extensive experiments have been carried out over six months to validate the effectiveness of WinIPS in a real-world multi-functional office. The experimental results show that the PSFM-GPR achieves a 4.8 dBm average RSS estimation error and a 1.718 m average localization accuracy, which outperforms existing approaches. In summary, WinIPS overcomes the bottlenecks of WiFi-based IPS, making it promising for large-scale practical implementations.

REFERENCES

[1] D. Lymberopoulos, J. Liu, X. Yang, R. R. Choudhury, V. Handziski, and S. Sen, "A realistic evaluation and comparison of indoor location technologies: experiences and lessons learned," in *Proceedings of the ACM International Conference on Information Processing in Sensor Networks*, pp. 178–189, 2015.

[2] Z. Chen, Q. Zhu, and Y. C. Soh, "Smartphone inertial sensor-based indoor localization and tracking with ibeacon corrections," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 4, pp. 1540–1549, 2016.

[3] S. Lee, B. Kim, H. Kim, R. Ha, and H. Cha, "Inertial sensor-based indoor pedestrian localization with minimum 802.15. 4a configuration," *IEEE Transactions on Industrial Informatics*, vol. 7, no. 3, pp. 455–466, 2011.

[4] P. Bahl and V. N. Padmanabhan, "Radar: An in-building rf-based user location and tracking system," in *Proceedings of the IEEE Annual Joint Conference of the Computer and Communications Societies*, vol. 2, pp. 775–784, 2000.

[5] F. Zhao, H. Luo, X. Zhao, Z. Pang, and H. Park, "Hyfi: Hybrid floor identification based on wireless fingerprinting and barometric pressure," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 1, pp. 330–341, 2017.

[6] J. M. Pak, C. K. Ahn, Y. S. Shmaliy, and M. T. Lim, "Improving reliability of particle filter-based localization in wireless sensor networks via hybrid particle/fir filtering," *IEEE Transactions on Industrial Informatics*, vol. 11, no. 5, pp. 1089–1098, 2015.

[7] H. Zou, X. Lu, H. Jiang, and L. Xie, "A fast and precise indoor localization algorithm based on an online sequential extreme learning machine," *Sensors*, vol. 15, no. 1, pp. 1804–1824, 2015.

[8] H. Lim, L.-C. Kung, J. C. Hou, and H. Luo, "Zero-configuration indoor localization over ieee 802.11 wireless infrastructure," *Wireless Networks*, vol. 16, no. 2, pp. 405–420, 2010.

[9] M. M. Atia, A. Noureldin, and M. J. Korenberg, "Dynamic online-calibrated radio maps for indoor positioning in wireless local area networks," *IEEE Transactions on Mobile Computing*, vol. 12, no. 9, pp. 1774–1787, 2013.

[10] L. M. Ni, Y. Liu, Y. C. Lau, and A. P. Patil, "Landmarc: indoor location sensing using active rfid," *Wireless Networks*, vol. 10, no. 6, pp. 701–710, 2004.

[11] P. Krishnan, A. Krishnakumar, W.-H. Ju, C. Mallows, and S. Gamt, "A system for lease: Location estimation assisted by stationary emitters for indoor rf wireless networks," in *Proceedings of the IEEE Annual Joint Conference of the Computer and Communications Societies*, vol. 2, pp. 1001–1011, 2004.

[12] J. Yin, Q. Yang, and L. M. Ni, "Learning adaptive temporal radio maps for signal-strength-based location estimation," *IEEE Transactions on Mobile Computing*, vol. 7, no. 7, pp. 869–883, 2008.

[13] S. Sorour, Y. Lostanlen, S. Valaee, and K. Majeed, "Joint indoor localization and radio map construction with limited deployment load," *IEEE Transactions on Mobile Computing*, vol. 14, no. 5, pp. 1031–1043, 2015.

[14] A. Rai, K. K. Chintalapudi, V. N. Padmanabhan, and R. Sen, "Zee: zero-effort crowdsourcing for indoor localization," in *Proceedings of the ACM Annual International Conference on Mobile Computing and Networking*, pp. 293–304, 2012.

[15] G. Shen, Z. Chen, P. Zhang, T. Moscibroda, and Y. Zhang, "Walkie-markie: Indoor pathway mapping made easy," in *Proceedings of the USENIX conference on Networked Systems Design and Implementation*, pp. 85–98, 2013.

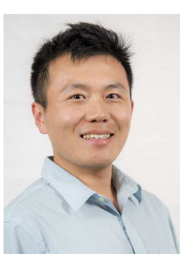
[16] H. Zou, B. Huang, X. Lu, H. Jiang, and L. Xie, "A robust indoor positioning system based on the procrustes analysis and weighted extreme learning machine," *IEEE Transactions on Wireless Communications*, vol. 15, no. 2, pp. 1252–1266, 2016.

[17] 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.

[18] T. Roos, P. Myllymäki, H. Tirri, P. Misikangas, and J. Sievänen, "A probabilistic approach to wlan user location estimation," *International Journal of Wireless Information Networks*, vol. 9, no. 3, pp. 155–164, 2002.

[19] X. Chai and Q. Yang, "Reducing the calibration effort for probabilistic indoor location estimation," *IEEE Transactions on Mobile Computing*, vol. 6, no. 6, pp. 649–662, 2007.
[20] B. Ferris, D. Haehnel, and D. Fox, "Gaussian processes for signal strength-based location estimation," in *Proceeding of Robotics: Science and Systems*, 2006.
[21] A. M. Hossain and W.-S. Soh, "Cramer-rao bound analysis of localization using signal strength difference as location fingerprint," in *Proceedings of the IEEE Annual Joint Conference of the Computer and Communications Societies*, pp. 1–9, 2010.
[22] N. Bargshady, N. A. Alsindi, K. Pahlavan, Y. Ye, and F. O. Akgul, "Bounds on performance of hybrid wifi-uwb cooperative rf localization for robotic applications," in *Proceedings of the IEEE International Symposium on Personal, Indoor and Mobile Radio Communications Workshops*, pp. 277–282, 2010.
[23] S.-S. Jan, S.-J. Yeh, and Y.-W. Liu, "Received signal strength database interpolation by kriging for a wi-fi indoor positioning system," *Sensors*, vol. 15, no. 9, pp. 21377–21393, 2015.
[24] S. J. Pan, J. T. Kwok, Q. Yang, and J. J. Pan, "Adaptive localization in a dynamic wifi environment through multi-view learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 1108–1113, 2007.
[25] OpenWrt. https://openwrt.org/.
[26] Libpcap. http://www.tcpdump.org/.
[27] C. K. Williams and C. E. Rasmussen, "Gaussian processes for machine learning," *MIT Press*, 2006.
[28] S. Kumar, R. M. Hegde, and N. Trigoni, "Gaussian process regression for fingerprinting based localization," *Ad Hoc Networks*, vol. 51, pp. 1–10, 2016.
[29] Y. Qi and H. Kobayashi, "On relation among time delay and signal strength based geolocation methods," in *Proceedings of the IEEE Global Telecommunications Conference*, vol. 7, pp. 4079–4083, 2003.
[30] A. Mahtab Hossain, Y. Jin, W.-S. Soh, and H. N. Van, "Ssd: A robust rf location fingerprint addressing mobile devices' heterogeneity," *IEEE Transactions on Mobile Computing*, vol. 12, no. 1, pp. 65–77, 2013.
[31] J. G. Park, D. Curtis, S. Teller, and J. Ledlie, "Implications of device diversity for organic localization," in *Proceedings of the IEEE Annual Joint Conference of the Computer and Communications Societies*, pp. 3182–3190, 2011.
[32] J. C. Gower, "Generalized procrustes analysis," *Psychometrika*, vol. 40, no. 1, pp. 33–51, 1975.
[33] K. Kaemarungsi and P. Krishnamurthy, "Analysis of wlans received signal strength indication for indoor location fingerprinting," *Pervasive and mobile computing*, vol. 8, no. 2, pp. 292–316, 2012.
[34] M. Jin, R. Jia, and C. Spanos, "Apec: Auto planner for efficient configuration of indoor positioning system," in *Proceedings of the International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies*, pp. 100–107, 2015.

Han Zou received the B.Eng. (First Class Honors) and Ph.D. degrees in Electrical and Electronic Engineering from the Nanyang Technological University, Singapore, in 2012 and 2016, respectively. He is currently a Postdoctoral Scholar with the Department of Electrical Engineering and Computer Sciences at the University of California, Berkeley, CA, USA. His research interests include ubiquitous computing, statistical learning, signal processing and data analytics with applications in occupancy sensing, indoor localization, cyber-physical systems, smart buildings and Internet of things.



Ming Jin is currently pursuing the Ph.D. degree in electrical engineering and computer science with the University of California, Berkeley. His current research interests include data-efficient algorithms for improving cyber-physical human system efficiency and resilience. He was the recipient of the Siebel scholarship (2018), the Best Paper Award at the International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services (2016), the Best Paper Award at the International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies (2015), the Electronic and Computer Engineering Department Scholarship (2012), the School of Engineering Scholarship (2012), and the University Scholarship at the Hong Kong University of Science and Technology (2012).

Hao Jiang obtained the B.E and Ph.D. degrees from the School of Information Science and Engineering, Xiamen University, China, in 2008 and 2013, respectively. He is currently an Associate Professor in the College of Electrical Engineering and Automation, Fuzhou University, Fuzhou, China. His research interests lie in localization system, sensor network, fiber optic sensor, and evolutionary algorithm.

Lihua Xie received the B.E. and M.E. degrees in electrical engineering from Nanjing University of Science and Technology in 1983 and 1986, respectively, and the Ph.D. degree in electrical engineering from the University of Newcastle, Australia, in 1992. Since 1992, he has been with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, where he is currently a professor and served as the Head of Division of Control and Instrumentation from July 2011 to June 2014. He held teaching appointments in the Department of Automatic Control, Nanjing University of Science and Technology from 1986 to 1989 and Changjiang Visiting Professorship with South China University of Technology from 2006 to 2011.

Dr Xie's research interests include robust control and estimation, networked control systems, multi-agent networks, and unmanned systems. He has served as an editor of IET Book Series in Control and an Associate Editor of a number of journals including IEEE Transactions on Automatic Control, Automatica, IEEE Transactions on Control Systems Technology, and IEEE Transactions on Circuits and Systems-II. Dr Xie is a Fellow of IEEE and Fellow of IFAC.

Costas J. Spanos received the EE Diploma from the National Technical University of Athens, Athens, Greece, in 1980, and the M.S. and Ph.D. degrees in electrical and computer engineering from Carnegie Mellon University, Pittsburgh, PA, USA, in 1981 and 1985, respectively.

In 1988, he joined the Faculty at the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA, USA. He has served as the Director of the Berkeley Microlab, the Associate Dean for Research in the College of Engineering, and the Chair of the Department of Electrical Engineering and Computer Sciences. He works on statistical analysis in the design and fabrication of integrated circuits, and on novel sensors and computer-aided techniques in semiconductor manufacturing. He also works on statistical data-mining techniques for energy efciency applications. He has participated in two successful startup companies: Timbre Tech (acquired by Tokyo Electron) and OnWafer Technologies (acquired by KLA-Tencor). He is the Director of the Center of Information Technology Research in the Interest of Society and the Chief Technical Officer for the Berkeley Educational Alliance for Research in Singapore.