SoiCP: A Seamless Outdoor–Indoor Crowdsensing Positioning System

Zan Li, Xiaohui Zhao, Fengye Hu, Zhongliang Zhao, José Luis Carrera Villacrés, and Torsten Braun

***Abstract*—Seamless outdoor–indoor positioning plays a critical role in many emerging applications, e.g., large-coverage user navigation in cities, smart buildings, and analytics of user spatial location big data. It is still challenging to construct a large-scale seamless outdoor–indoor positioning system due to the limited coverage of indoor positioning. In this paper, we propose a seamless outdoor–indoor crowdsensing positioning (SoiCP) system in which a radio map is automatically constructed based on crowdsourcing pedestrian dead reckoning (PDR) traces without professional site surveying. The constructed radio map is robust to inaccurate PDR traces and does not rely on prior knowledge of floor plans. In SoiCP, the crowdsensed radio map is obtained by a proposed three-step trace matching algorithm. This algorithm leverages building gates and WiFi fingerprints as landmarks to merge the noisy crowdsourcing traces and accurately construct the user walking paths. Moreover, following the crowdsensed radio map, SoiCP uses an enhanced particle filter to fuse PDR, GPS, and WiFi fingerprinting for seamless outdoor–indoor positioning with high accuracy. The comprehensive real-world experiments in two large-scale shopping malls demonstrate that SoiCP can effectively crowdsense the walking paths and track moving users with high accuracy.**

***Index Terms*—Crowdsensing, indoor positioning, particle filter, trace matching.**

I. Introduction

In RECENT years, the location information of people has become increasingly important for many emerging applications, such as large-coverage location-based services (LBS), smart buildings, and analytics of user spatial location big data. Although the global positioning system (GPS) provides satisfactory outdoor positioning services, it does not work indoor. WiFi is currently the dominant wireless local area networking standard and is becoming the leading technology for indoor positioning [1]–[3]. WiFi fingerprinting based on received signal strength indicator (RSSI) has been widely used in commercial indoor positioning systems because of its high accuracy and ubiquitousness. Generally, a traditional fingerprinting system requires a radio map comprised of a massive number of WiFi fingerprints. Each WiFi fingerprint consists of an RSSI list and its ground truth location label. To build up radio maps, professionals consume huge amount of time to conduct site surveying for the ground truth location labels [4]. Therefore, at present, the traditional fingerprinting systems are merely deployed at certain buildings where a site survey has been conducted by professionals. They cannot achieve large coverage of indoor positioning services for a massive number of buildings to meet the requirement of seamless outdoor–indoor positioning, e.g., in the scenario of smart cities.

To avoid intensive training efforts in fingerprinting, it is becoming interesting for researchers to crowdsense radio maps based on crowdsourcing user traces. Crowdsensing indoor positioning systems collect location-related information from crowdsourcing users instead of professionals, such as WiFi RSSI, pedestrian dead reckoning (PDR) traces, and GPS, to automatically build radio maps. Several researchers [5]–[9] utilize floor plans and landmarks to estimate the locations of the crowdsourcing users for labeling radio maps. However, floor plans are normally not available for large-scale positioning. Hence, crowdsensing walking paths without floor plans has a potential for automatic construction of radio maps in ubiquitous positioning. Trace matching algorithms are the critical technique for crowdsensing walking paths by adjusting the locations of the raw traces based on limitations among the traces defined by indoor landmarks such as radio signals and magnetic field. Although some studies [10], [11] have been conducted for trace matching and claim good accuracy, they are derived and verified under the conditions that the rotation errors of the raw traces are small and Gaussian distributed, which may not be realistic in some real-world crowdsourcing scenarios.

In this paper, we propose a seamless outdoor–indoor crowdsensing positioning (SoiCP) system without site surveying, as shown in Fig. 1. SoiCP requires neither priori knowledge of a floor plan nor manual initialization of PDR traces to crowdsense a radio map. It is robust to biased rotation of PDR traces caused by magnetic interference. Additionally, we design an enhanced particle filter to fuse PDR, GPS, and WiFi fingerprinting for seamless outdoor–indoor positioning. Our main contributions are summarized in two aspects as follows.

A. Crowdsensing Radio Map

We propose a three-step (intra, inner, and inter) trace matching algorithm as shown in Fig. 2 to merge noisy user

Manuscript received March 5, 2019; revised May 7, 2019; accepted June 3, 2019. Date of publication June 7, 2019; date of current version October 8, 2019. This work was supported in part by the National Natural Science Foundation of China under Grant 61801189, and in part by the Swiss National Science Foundation via the Intelligent Mobility Services Project under Grant 200021_184690. *(Corresponding author: Zan Li.)*

Z. Li, X. Zhao, and F. Hu are with the College of Communication Engineering, Jilin University, Changchun 130012, China (e-mail: zanli@jlu.edu.cn; xhzhao@jlu.edu.cn; hufy@jlu.edu.cn).

Z. Zhao, J. L. C. Villacrés, and T. Braun are with the Institute of Computer Science, University of Bern, 3012 Bern, Switzerland (e-mail: zhao@inf.unibe.ch; carrera@inf.unibe.ch; braun@inf.unibe.ch).

Digital Object Identifier 10.1109/JIOT.2019.2921561

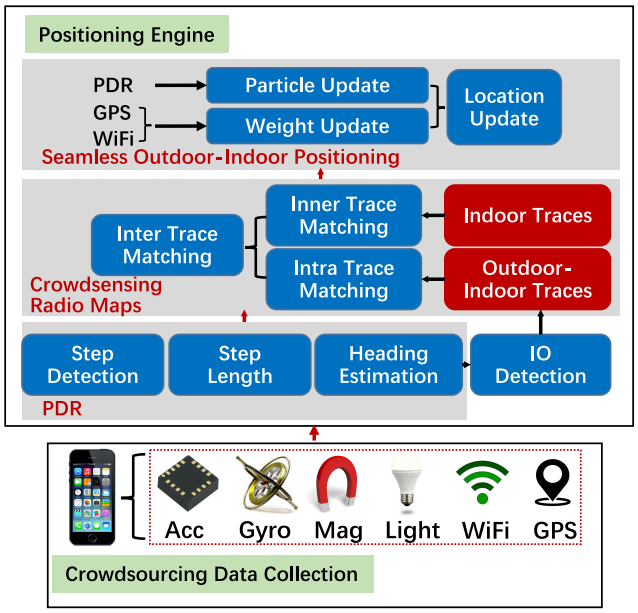


Fig. 1. SoiCP system.

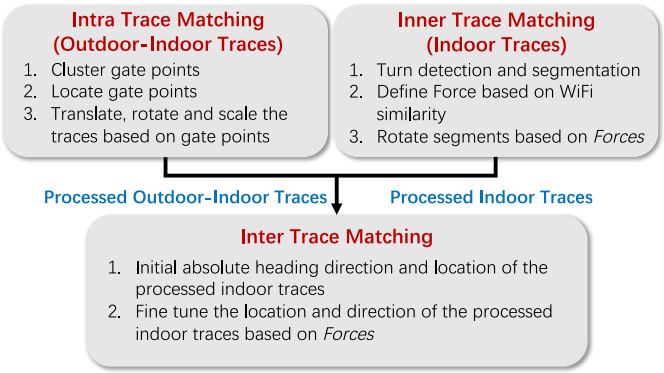


Fig. 2. Three-step trace matching algorithm.

traces collected by crowdsourcing for automatic construction of indoor walking paths. This algorithm relies on two kinds of ubiquitous landmarks, i.e., building gates and WiFi fingerprints. In the algorithm, intratrace matching merges the opportunistic outdoor–indoor PDR traces in which users walk through a target building to generate parts of walking paths according to the building gates as landmarks. An unsupervised learning approach, density-based spatial clustering of applications with noise (DBSCAN) [12], is adopted based on WiFi similarity to cluster and locate the gate points which are defined as the step points detected as building gates on the traces. Then, we design an inner trace matching algorithm to compensate the turning errors of the indoor PDR traces by minimizing an indicator named "force" which is defined as a weighted summation of physical distances among the WiFi fingerprints with high similarity. Finally, we design an intertrace matching algorithm by merging the processed indoor PDR traces to the outdoor–indoor ones. We also formulate the intertrace matching problem as an optimization problem to minimize the forces between the indoor and outdoor–indoor PDR traces by translating and rotating the indoor PDR traces. Note that the indoor PDR traces are preprocessed based on an enhanced magnetic interference detection method to reduce the rotation errors. The crowdsensed walking paths are used to label the WiFi fingerprints in the radio map.

B. Seamless Outdoor–Indoor Positioning

We design an enhanced particle filter to fuse PDR, GPS, and WiFi fingerprinting for seamless outdoor–indoor positioning according to the crowdsensed radio map. In the particle filter, we propose enhanced methods for both system and observation models. For the system model, an adaptive heading direction estimation algorithm relying on the aforementioned enhanced magnetic interference detection algorithm is proposed to design an enhanced PDR for updating particles. For the observation model, GPS is used outdoor to update the associated weights of the particles in which a GPS accuracy detection algorithm and an adaptive observation model are proposed. The estimated locations based on WiFi fingerprinting are taken as indoor measurements to update the associated weights. Moreover, an enhanced indoor and outdoor detector (IO detector) is proposed to switch the observation models.

According to our comprehensive experiments conducted in two large-scale shopping malls, SoiCP crowdsenses walking paths with mean accuracies of 2.8 and 2.6 m. In indoor environments, SoiCP can track moving users with mean accuracies of 3.0 and 2.9 m by integrating PDR, GPS, and WiFi fingerprinting, which outperforms WiFi fingerprinting and PDR-GPS fusion algorithms.

In the remainder of this paper, the related work is reviewed in Section II. The problems and challenges for crowdsensing positioning are discussed in Section III. Our main contributions in SoiCP are presented in Sections IV–VI. Section IV presents an overview of the system and the methods for raw user trace generation. Section V introduces the three-step trace matching algorithm for crowdsensing radio maps. Section VI presents the proposed particle filter for seamless outdoor–indoor positioning. Section VII describes the experiments and evaluation results in the large-scale shopping malls. Finally, Section VIII concludes this paper.

II. Related Work

A. Pedestrian Dead Reckoning

PDR is a key technique used in both crowdsensing radio maps and fusion-based positioning. PDR is comprised of three parts, i.e., step detection, step length estimation, and heading direction estimation. For step detection, one step is detected based on a pair of peak and valley in the measured accelerometer data [13]. However, noisy accelerometer data may degrade its accuracy. For step length estimation, a linear model for the step frequency and length is often used [14]. Heading direction estimation obtained by a compass or a gyroscope is the most challenging part in PDR. A compass designed by a magnetic sensor can obtain the absolute direction but is subject to magnetic interference inside buildings whose direction deviation may be larger than 100° [15]. The heading direction can also be obtained by integrating the angle velocity from gyroscope based on an initial angle as a reference, which is

TABLE I
DIFFERENT CROWDSENSING RADIO MAP SYSTEMS

	Floor Plan	Additional Devices	Landmarks	Absolute Locations	Magnetic Interference
SoiCP	Not needed	Not needed	WiFi and building gates	Obtained by gates and GPS	Handled
Walkie-Markie	Not needed	Not needed	WiFi	Not provide	No
PiLoc	Not needed	Not needed	WiFi	Not provide	No
FinLoc	Not needed	Bluetooth Nodes	Bluetooth	Obtained by Bluetooth nodes	No
UnLoc	Needed	Not needed	Seed landmarks, magnetic and WiFi	Obtained by seed landmarks	No

normally accurate for a short time but prone to accumulated errors in long-time measurements. There are many studies conducted in fusion of compass and gyroscope to improve the accuracy of heading direction estimation. X-AHRS [16] fuses gyroscope, accelerometer, and magnetometer to estimate Euler angles in which a parameter β is introduced to tune the impact of magnetometer and gyroscope on the estimated Euler angles. β is determined by the magnetic interference inside buildings but the interference is difficult to know. Li *et al.* [17] and Hong *et al.* [18] proposed to obtain heading errors based on floor plans which are however unavailable for some positioning scenarios, especially the crowdsourcing cases. A magnetic interference detection algorithm is proposed in [19] by comparing the Euler angles estimated by magnetometer and gyroscope. The algorithm can only detect the local distorted magnetic field in limited areas but cannot handle the entire distorted magnetic field affected by the iron structures of buildings.

B. Crowdsensing Radio Maps

To avoid labor-intensive site surveying, crowdsensing radio map approaches have recently been proposed. Zee [6] is an early crowdsensing indoor positioning system with zero efforts leveraging the movement of users and a floor plan to filter infeasible locations over time. UnLoc [7] is another crowdsensing indoor positioning system relying on some fixed landmarks with known locations in a floor plan to achieve trace matching. LIFS [5] leverages crowdsourcing PDR traces and a floor plan to automatically construct a 3-D radio map. Yu *et al.* [20] detected the door of each room as a reference location and cluster the fingerprints which are linked to their corresponding physical locations based on a floor plan. Ramchandran [21] leveraged a floor plan and wearable sensors to get the ground truth locations of the data collectors. RCILS [22] abstracts the indoor map as the semantics graph in which the edges are the possible user paths and the vertexes are the locations where users may take special activities. RCILS requires a floor plan for matching the graph to the target building. Zhuang *et al.* [23] adopted a commercial PDR algorithm to obtain the accurate user traces and construct the radio map. Li *et al.* [24] proposed a robust PDR algorithms. They leveraged big data techniques to obtain large amount of user traces and mark these traces with an accuracy indicator used to improve positioning accuracy. However, this approach requires large amount of time to collect enough PDR traces for positioning. FineLoc [25] leverages bluetooth to efficiently merge the user traces. Benefitting from the small coverage of bluetooth, the traces can be accurately merged. To know the absolute location of the generated walking paths in the target building, some bluetooth nodes with known locations are required. However, the floor plans or additional devices, e.g., wearable sensors, commercial PDR systems, and bluetooth nodes in the aforementioned works, may not be available for some crowdsourcing positioning scenarios especially for smart cities with a large number of buildings.

Much research has been conducted in crowdsensing radio map without assistant of floor plans and additional devices. Walkie-Markie [11], PiLoc [10], and its extension MPiLoc [26] are the most similar systems as this paper which leverage WiFi similarity to merge the user traces. Precisely, leveraging crowdsourcing PDR traces and WiFi access points as landmarks, Walkie-Markie generates the indoor walking paths of buildings. It needs to find the correct locations of the WiFi access points on each trace based on their largest RSSIs. However, due to the instability of WiFi RSSI, the locations of WiFi access points may not be reliable and correspondingly influence the generated walking paths. On the other hand, Walkie-Markie requires the direction errors of PDR traces following a zero-mean Gaussian distribution and a measurement with long time for enough user traces is necessary for trace matching. Additionally, the absolute locations of the walking paths are not provided in Walkie-Markie. PiLoc clusters user traces based on their shapes and WiFi similarity, and merges the traces by averaging the corresponding step points in the same cluster. Moreover, the authors further extend PiLoc to support multifloor radio map crowdsensing in a system called MPiLoc. For PiLoc and MPiLoc, the direction errors of PDR traces are assumed to be small and follow a zero-mean Gaussian distribution. In this case, they can cluster the traces with their shapes and merges the traces by simply averaging the corresponding step points in the same cluster. Large direction errors will significantly affect the trace clustering and merging. On the other hand, the authors mention that the absolute locations of the walking paths can be obtained by some occasional GPS locations near windows, but they do not provide a clear solution. In practice, indoor GPS signals obtained from windows are normally unreliable. Different from Walkie-Markie, PiLoc and its extension MPiLoC, we target on a crowdsensing radio map system without assistant of floor plans, in which the direction errors of the raw traces are random and the absolute locations of the walking paths are obtained.

Table I lists the differences among some recent crowdsensing radio map systems.

C. Seamless Indoor and Outdoor Positioning

To achieve large coverage of LBSs, seamless indoor and outdoor positioning fusing multimodel sensors is required.

Bayesian filters are often utilized for sensor fusion such as Kalman filter (KF), extended Kalman filter (EKF), and particle filter. In indoor environments, Li *et al.* [27] adopted a Kalman filter to fuse inertial sensors and WiFi fingerprints for indoor positioning. Moreover, they extended their solution to support 3-D positioning by fusing barometers. However, their solution only considers positioning in indoor environments but not work for outdoor environments. In outdoor environments, Kim *et al.* [28] and Kourogi *et al.* [29] adopted Kalman filters for GPS-inertial measurement unit (IMU) fusion. The covariance of GPS measurements in the Kalman filters is set based on GPS accuracy detected by the distribution of satellites and signal to noise ratio (SNR) of the received GPS signals. For indoor environments, WiFi signals are often integrated with IMU sensors for indoor positioning. References [30]–[32] fuse IMU sensors and WiFi RSSI by particle filters. Weyn *et al.* [33] fused GPS, WiFi, and global system for mobile communications (GSM) signals in the measurement models of particle filters with different weights to achieve a seamless indoor and outdoor positioning without distinguishing the users' indoor and outdoor status. Hansen *et al.* [34] investigated four strategies for GPS and WiFi fingerprinting handover.

Nowadays, with various sensors integrated in smartphones, indoor and outdoor detection (IO detection) is more often used as a switching engine to switch the indoor and outdoor positioning techniques. Morillo *et al.* [35] and Hardegger *et al.* [36] proposed combined approaches with radio signals and inertial sensors to achieve outdoor exit detection. However, in those work [35], [36], the detection accuracy is not given. According to our experience, these approaches relying only on radio signals and inertial sensors are normally not accurate enough for real-time switching indoor and outdoor positioning. Li *et al.* [37] designed an IO detection by integrating light sensor, cellular network information, and magnetic sensor with a hidden Markov model (HMM) estimator, which achieves a detection precision of higher than 90%. Radu *et al.* [38] proposed a semisupervised learning solution by integrating more information, such as battery temperature and microphone detected noise amplitude, and they declared a detection precision of 93.3% with high robustness. The high detection accuracy in [37] and [38] mainly relies on the light sensors. Although IO detection based on light sensors is reliable, light sensors may not be always available, especially when smartphones are placed in pockets.

We propose a preliminary approach for crowdsensing radio maps based on building gates and WiFi fingerprints as landmarks in [39] (a conference paper). The work of [39] is the basis of this paper. In this paper, we extend the crowdsensing walking paths algorithm in [39] to support trace matching with multiple gate points. Then, leveraging an enhanced magnetic inference detection algorithm, we improve intratrace matching by a joint estimation of rotation angle and locations. Moreover, we provide an accurate seamless outdoor–indoor positioning algorithm based on an enhanced particle filter in this paper. We conduct comprehensive experiments to test the effectiveness of the SoiCP system.

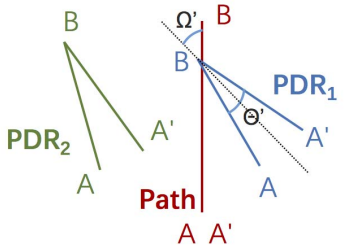


Fig. 3. PDR errors.

III. Problem Statement and Technical Challenges

To design a seamless outdoor–indoor positioning system without site surveying and priori knowledge of floor plans, we need to handle the following issues.

A. Crowdsensing Radio Map

1) PDR Errors: Due to the inaccuracy of inertial sensors and PDR algorithms, the errors of the collected PDR traces can be divided into three categories. We take Fig. 3 as an instance, in which a user walks along the path from points A to B and back to point A (marked as A') twice. Two PDR traces, namely PDR_1 and PDR_2, are obtained according to PDR algorithms. First, due to the drift of gyroscope, the turning angles at turn B cannot be accurately measured, which produces *turning errors* θ'. Therefore, the shapes of PDR traces are different from the ground truth paths. Second, the PDR traces may rotate with certain angles owing to the distorted indoor magnetic field, which is referred to as *rotation errors* Ω'. Third, because of the step length errors, the lengths of PDR traces can also be different from the ground truth lengths.

2) Uncertainty of Absolute Locations: The collected PDR traces only indicate the relative movement of users without absolute locations, e.g., latitude and longitude in the World Geographic Reference System (GEOREF). For a crowdsourcing system, it is unrealistic to initialize each PDR trace with a known starting point. Even if certain PDR traces may opportunistically obtain GPS locations around a target building, these PDR traces are generally not accurate enough to produce walking paths because of low GPS accuracy near tall buildings.

3) Trace Matching: Due to the PDR errors and the uncertainty of absolute locations, the PDR traces may represent different walking paths even if they are collected from the same ground truth path. As shown in Fig. 3, PDR_1 and PDR_2 are separated even if they are collected from the same path. To recover the walking paths of a target building, we need to design trace matching algorithms to merge (parts of) the PDR traces collected from the same ground truth path. Typically, trace matching can be achieved by leveraging some fixed landmarks with known locations in the target building, such as stairs, elevators, and doors. However, in a real-world crowdsensing positioning scenario, the locations of these landmarks are difficult to obtain because floor plans are normally unavailable.

B. Seamless Outdoor–Indoor Positioning

1) Sensor Fusion: Particle filter is a powerful tool to fuse multiple positioning techniques for seamless outdoor and indoor positioning. A properly designed particle filter is required to adaptively switch or fuse different positioning techniques to achieve higher accuracy.

2) Indoor and Outdoor Detection: Accurate IO detection is critical for switching indoor and outdoor positioning techniques. Light sensors can accurately detect indoor and outdoor status with short delay. However, in certain cases that the light sensors are unavailable, radio signals such as cellular networks and WiFi are used for IO detection, which, however, may encounter large delay due to the instability of radio signals.

3) GPS Accuracy Detection: In outdoor environments, GPS is fused with inertial sensors in particle filters. The accuracy of GPS determines its impact on the final locations. GPS accuracy reported from Android phones is calculated by many factors, e.g., distribution of the observed satellites, SNR of the received GPS signals, and pseudo ranges. However, the Android-reported GPS accuracy is still not reliable according to some preliminary tests.

4) Heading Direction Estimation: PDR positioning based on inertial sensors are used in a particle filter for both outdoor and indoor positioning. As the most challenging part, heading direction estimation is subject to indoor magnetic interference and finally affects the positioning accuracy.

C. Target and Challenges of This Paper

In this paper, we mainly aim to answer two questions.

1) Is it possible to design a trace matching algorithm to effectively merge the noisy crowdsourcing traces with large PDR errors and uncertain absolute locations?
2) How can we efficiently fuse PDR, GPS, and WiFi fingerprinting to achieve seamless outdoor–indoor positioning?

To answer and solve the aforementioned questions, we must handle the following challenges, respectively, for crowdsensing radio map and seamless outdoor–indoor positioning.

1) *Noisy PDR Traces:* As mentioned above, it is challenging to obtain PDR traces with high accuracy. Especially, the distortion of indoor magnetic field will significantly affect the estimation of heading direction. For both crowdsensing radio map and seamless outdoor–indoor positioning, PDR algorithms play an important role. Hence, compensation of heading direction errors resulting from the indoor magnetic field is a key to the whole system.
2) *Noisy GPS Locations:* GPS locations are used for assigning absolute locations to the user traces on the step of crowdsensing radio map and also for outdoor positioning on the step of seamless outdoor–indoor positioning. However, GPS locations are normally noisy, especially around the tall buildings, and its accuracy is difficult to obtain. Hence, it is necessary to design a reliable GPS accuracy detection method.
3) *Landmarks for Trace Matching:* WiFi signals are often used for tracing matching as landmarks due to their ubiquitousness. However, they are normally unstable and lack of absolute locations to recover the walking paths. Hence, our designed trace matching algorithm with WiFi signals should be robust to the instability of WiFi signals. Moreover, in addition to WiFi signals, we need to find another kind of landmarks to further improve trace matching and obtain the absolute locations of walking paths.
4) *IO Detection:* IO detection is critical to switch the indoor and outdoor positioning techniques in seamless outdoor–indoor positioning. Light sensors are normally reliable to design an IO detector but they can only be used when the light intensity significantly changes between indoor and outdoor environments. Otherwise, the traditional solutions with ubiquitous signals, such as WiFi signals, magnetic field, and GPS signals are normally unreliable. Therefore, to design a reliable IO detector is another significant issue for seamless indoor and outdoor positioning.

IV. Overview of Seamless Outdoor–Indoor Crowdsensing Positioning System

A. System Overview

Fig. 1 presents an overview of the proposed SoiCP system. This prototype system runs on smartphones and a cloud for crowdsourcing data collection and positioning engine. The part of crowdsourcing data collection is responsible for collecting massive sensor data from accelerometer, gyroscope, magnetometer, WiFi, GPS, and light sensors based on a self-designed Android application. In the current system, the users need to click a button in the Android application to start the data collection. Then, all sensor data are stored in the smartphone locally. In all our measurements, the accelerometer, gyroscope, and magnetometer are sampled with a rate of 50 Hz. The light sensor is sampled in every 350 ms. The WiFi signals are scanned in every 3 s and the GPS locations are obtained in every second. The data from different sensors are stored in separated files. Each piece of sensor data is timestamped with the linux time of the smartphone when it is sampled. Hence, all the sensor data in the separated files can be synchronized for future processing.

The sensor data from all users are first collected in a local working station, which can be a desktop or a laptop. Then, the collected sensor data are sent to a remote cloud server placed in the University of Bern where one virtual machine (VM) is assigned for the SoiCP system and the positioning engine runs. The VM is installed with an Ubuntu system in which Python is adopted for processing data. The positioning engine of SoiCP consists of two parts: 1) crowdsensing indoor radio maps and 2) seamless outdoor–indoor positioning based on multisensor fusion. The part of crowdsensing indoor radio maps focuses on automatic construction of indoor radio maps with zero effort for site surveying. A three-step trace matching algorithm runs in this part to merge the crowdsourcing user traces and generate the indoor walking paths. Each fingerprint in the generated radio map is labeled with the location on

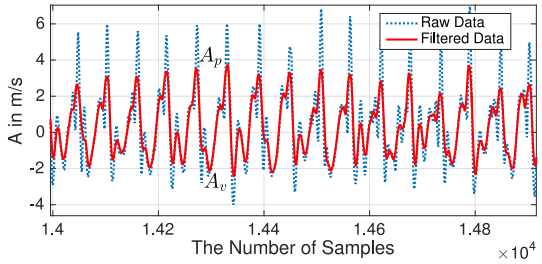


Fig. 4. Variation of acceleration during walking.

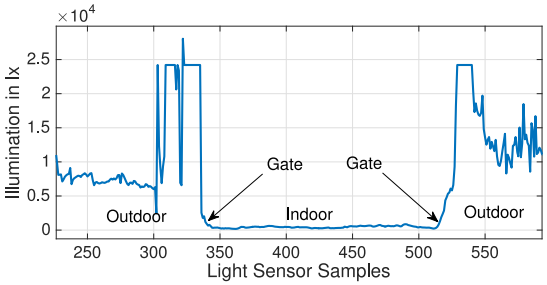


Fig. 5. Variation of light intensity outdoor and indoor.

the crowdsensed indoor walking paths. Based on the crowdsensed radio map, an enhanced particle filter fusing PDR, GPS, and WiFi fingerprinting is utilized for seamless outdoor–indoor positioning with high accuracy.

B. Raw User Trace Generation

To obtain the user traces, we design an offline PDR algorithm applied on the collected IMU sensor data (accelerometer and gyroscope), which consists of step detection, step length estimation, and heading direction estimation.

1) Step Detection: As shown in Fig. 4, human walking generates periodic variations of the measured accelerometer data [40]. Therefore, we detect peaks and valleys from the magnitude of triaxial accelerometer data as steps. To prevent the misrecognition of the steps due to noisy accelerometer data, we first smooth the magnitude of accelerometer data via a moving average filter and subtract the gravity. In consequence, we detect peaks and valleys as in Fig. 4. The magnitude of a peak is marked as A^p and detection time as T^p. The magnitude of a valley is marked as A^v and detection time as T^v. When the following two conditions are met, a pair of neighbor valley and peak is recognized as a step:

$$\begin{cases} \Delta A = A^p - A^v > 0.5 \text{ m/s}^2 \\ \Delta T = T^v - T^p > 0.15 \text{ s}. \end{cases} \tag{1}$$

2) Step Length Estimation: We estimate the step length by a linear model [14]

$$L = a \times f + b \tag{2}$$

where L is the step length, f is the step frequency, and (a, b) are the coefficients calibrated for individuals. We set both a and b be 0.25, which are the optimal coefficients without priori knowledge of user features, such as lengths and weights [14].

3) Heading Direction Estimation: We adopt the X-AHRS filter [16] fusing gyroscope and accelerometer to estimate the Euler angles. Magnetometer is not used at this step to avoid the heading direction errors due to severe indoor distortion of magnetic field. With only gyroscope and accelerometer, each PDR trace may encounter the turning error θ' and the large rotation error Ω' introduced in Section II but can better keep its original shape.

We classify the generated PDR traces into indoor and outdoor–indoor traces relying on light sensors. As shown in Fig. 5, outdoor light intensity is significantly stronger than that of indoor during daytime. Note that it is reasonable to only conduct the crowdsourcing data collection during daytime, which is different from real-time positioning. The light intensity on an indoor PDR trace is higher than $1000lx$ all through. For an outdoor–indoor PDR trace, the light intensity is smaller than $1000lx$ in the middle part but larger than $1000lx$ at both ends.

V. Crowdsensing Indoor Radio Maps

A three-step trace matching algorithm described in Fig. 2 is designed in SoiCP to merge the crowdsourcing user traces to recover the walking paths and construct the radio map.

A. Intratrace Matching

At the first step, intratrace matching merges the outdoor–indoor PDR traces based on the building gates as landmarks because the users walking through the building will definitely cross certain gates. At this step, we cluster the detected gate points on each trace, estimate their locations, and merge the traces according to them.

1) Gate Landmark Detection and Clustering: We detect $2N$ gates on N traces based on the sudden change of light sensor data as shown in Fig. 5. Since the ground truth gates which these detected gates belong to are unknown, we have to cluster these $2N$ detected gates for trace matching. To cluster the gate landmarks, we make use of WiFi similarity around the gates. We first search for the corresponding WiFi RSSI list around each detected gate landmark and then obtain their universal set of the MAC addresses with length M. We further define a feature vector to represent each gate landmark, in which each feature is the WiFi RSSI of one MAC address as

$$\mathbf{RSSI}_n = \left\{\text{RSSI}_{n,1}, \text{RSSI}_{n,2}, \ldots, \text{RSSI}_{n,M}\right\} \tag{3}$$

where n represents the nth gate point. We further form a WiFi feature matrix with a size of $2N \times M$ containing the feature vectors of $2N$ gate landmarks. Because the Euclidean distances of the feature vectors are generally small for nearby locations, we utilize DBSCAN [12] to cluster the gate points because it does not require to specify the number of clusters representing the number of gates.

DBSCAN is a commonly used density-based clustering algorithm which can efficiently generate clusters of arbitrary shapes based on a density-based notion of clusters. It uses a neighborhood radius ϵ and minimum number of points in the neighborhood *MinPts* to classify a set of points into core points in clusters, border points in clusters and outliers. A core

point is defined as the one which has at least *MinPts* points in its ϵ neighborhood. If one noncore point is reachable from a core point, it is a border point. Otherwise, the noncore point is regarded as an outlier.

DBSCAN sequentially clusters the target points as follows. Starting from an arbitrary point p unvisited, DBSCAN finds all points in ϵ neighborhood of p. If p is a core point, all its ϵ-neighborhood points are regarded as in a cluster. For each point in the cluster, all the density-reachable points are added into the cluster. A final cluster is obtained by repeating this process until no new density-reachable points are found. If p is not a core point, we move to the next point.

2) Gate Landmark Location: After clustering the gate landmarks, we decide the location of each gate by processing each outdoor–indoor PDR trace based on the opportunistic GPS locations. We, respectively, find the GPS locations with the highest Android-reported accuracy at each end [the last GPS location ($\mathbf{G}^i$) before the user enters the building and the first one ($\mathbf{G}^o$) after she/he exits] which meet the following condition

$$\begin{cases} P < P^{\text{Thre}} \\ \Delta v = |v - v'| <= 1 \text{ m/s} \\ \Delta\phi = |\phi - \phi'| <= 10^\circ \end{cases} \tag{4}$$

where P, v', and ϕ' are the Android-reported GPS accuracy, speed, and bearing angle, respectively. P^{Thre} is a threshold for the Android-reported GPS accuracy and empirically set to 6 m. v and ϕ are the speed and bearing angle calculated according to the adjacent GPS locations. Because the Android-reported GPS speed and bearing angle are calculated based on the Doppler effects and GPS locations are estimated based on the ranges, these two processes are independent [41]. Hence, we leverage their differences Δv and $\Delta\phi$ to assist the detection of good GPS locations. Note that both $\mathbf{G}^i$ and $\mathbf{G}^o$ are the vectors consisting of their longitudes and latitudes.

Then, $\mathbf{G}^i$ and $\mathbf{G}^o$ are used as pins to process the PDR traces. We search for the PDR points $\mathbf{P}^i$ and $\mathbf{P}^o$ whose timestamps are most close to $\mathbf{G}^i$ and $\mathbf{G}^o$. As shown in Fig. 6, we must translate, rotate, and scale the PDR trace to achieve

$$\mathbf{G}^i = \mathbf{P}^i \text{ and } \mathbf{G}^o = \mathbf{P}^o. \tag{5}$$

We obtain the rotation angle ω and the scaling factor α by solving

$$\mathbf{G}^o = \alpha\mathbf{S}\left(\mathbf{P}^o - \mathbf{P}^i\right) + \mathbf{G}^i \tag{6}$$

where $\mathbf{S}$ is the rotation matrix defined as

$$\mathbf{S} = \begin{bmatrix} \cos(\omega) & -\sin(\omega) \\ \sin(\omega) & \cos(\omega) \end{bmatrix}.$$

Therefore, we translate the starting point of the PDR trace to $\mathbf{G}^i$, rotate the trace around $\mathbf{G}^i$ with an angle of ω and scale it with a factor α.

After processing the outdoor–indoor PDR traces with GPS locations, we get the coordinates (latitude and longitude) of the gate points on each trace. Although the accurate GPS locations has been selected based on the proposed enhanced GPS accuracy detection algorithm in (4), the coordinates of the gate points on each trace are still noisy. Therefore, we further average the coordinates of the gate landmarks in the same cluster to obtain their ground truth coordinates.

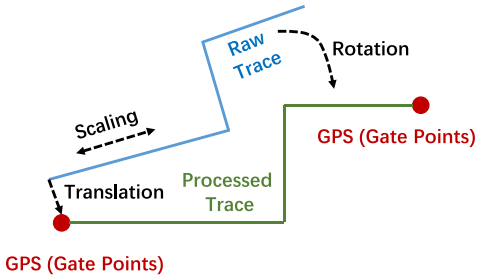


Fig. 6. Intratrace matching for outdoor–indoor PDR traces.

3) Intratrace Matching via Gate Landmarks: After defining the number and locations of gate points, we merge the outdoor–indoor traces based on them. Assuming that the two gate points on one user trace are, respectively, classified into the ith and kth clusters whose locations are defined as $\mathbf{D}^i$ and $\mathbf{D}^k$, $\mathbf{D}^i$ and $\mathbf{D}^k$ are used as pins (landmarks) to merge the traces. We first search for the PDR points $\mathbf{P}^i$ and $\mathbf{P}^k$ whose timestamps are most close to $\mathbf{D}^i$ and $\mathbf{D}^k$. Then, we must translate, rotate and scale the PDR trace to achieve

$$\mathbf{D}^i = \mathbf{P}^i \text{ and } \mathbf{D}^k = \mathbf{P}^k. \tag{7}$$

We obtain the rotation angle Ω^{intra} and the scaling factor α^{intra} by solving

$$\mathbf{D}^i = \alpha^{\text{intra}}\mathbf{S}'\left(\mathbf{P}^i - \mathbf{P}^k\right) + \mathbf{D}^k \tag{8}$$

where

$$\mathbf{S}' = \begin{bmatrix} \cos\left(\Omega^{\text{intra}}\right) & -\sin\left(\Omega^{\text{intra}}\right) \\ \sin\left(\Omega^{\text{intra}}\right) & \cos\left(\Omega^{\text{intra}}\right) \end{bmatrix}.$$

Finally, we translate the starting point of the PDR trace ($\mathbf{P}^k$) to $\mathbf{D}^k$, rotate the trace around $\mathbf{D}^k$ with an angle of Ω^{intra} and scale it with a factor of α^{intra}.

Note that Ω^{intra} is used to compensate the rotation errors Ω' as mentioned in Section II via the gate landmarks. α^{intra} is used to compensate the step length errors. Each outdoor–indoor PDR trace is assigned with latitude and longitude according to the locations of the building gates thanks to the opportunistic GPS locations.

B. Inner Trace Matching

Intratrace matching can generate parts of the walking paths in limited subareas which are normally not enough to cover the entire building. Therefore, we further utilize the indoor PDR traces to extend the coverage of the walking paths. For inner trace matching, the target is to compensate the turning errors θ' of the indoor PDR traces mentioned in Section II.

1) Turn Detection and Segmentation: We partition each indoor PDR trace into several line segments based on the turn angles larger than 60°. We define the nth turn as $\mathbf{T}_n$. Each segment can rotate around its first attached turn. For example, in Fig. 3, we can partition PDR_1 or PDR_2 into two segments AB and BA' based on the turn at point B, respectively. Segment BA' can be rotated around turn B.

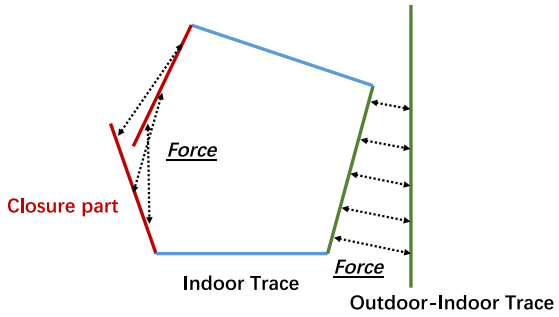


Fig. 7. Force definition.

2) Force Definition: To compensate θ', we consider the scenario that the user walks in a closed loop. This means that the user walks back to the paths defined as *"closure parts"* which he already passed. In these closure parts, the points in different segments but with similar WiFi RSSI lists should be close to each other in the ground truth paths. However, because of the inaccurate turn angles estimated by gyroscope, these points may be far away from each other in the measured PDR trace. Therefore, to compensate θ', we rotate each segment with an indicator, named force. Force represents a weighted physical distance between any pair of points belonging to different segments whose WiFi RSSI lists are with high similarity as shown in Fig. 7. The force is defined as

$$F_{i,j} = \begin{cases} \frac{E_{i,j}D_{i,j}}{\Lambda} & \text{when } E_{i,j} > 0.7 \text{ and } D_{i,j} > 10 \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

where $E_{i,j}$ is the cosine similarity of WiFi RSSI lists and $D_{i,j}$ is the physical distance between the ith and jth points in the PDR trace. Λ is a constant factor. The cosine similarity of WiFi RSSI lists is calculated by

$$E_{i,j} = \frac{\sum_k \left(\text{RSSI}_{i,k}\text{RSSI}_{j,k}\right)}{\sqrt{\sum_k \text{RSSI}_{i,k}^2}\sqrt{\sum_k \text{RSSI}_{j,k}^2}} \tag{10}$$

where $\text{RSSI}_{i,k}$ is the RSSI of the kth access point at the ith point. Note that we only consider the forces whose WiFi similarity are high but the physical distances are large, i.e., $E_{i,j} > 0.7$ and $D_{i,j} > 10$. Then, the overall force of the target PDR trace is calculated by

$$F = \sum_{i,j} F_{i,j}. \tag{11}$$

According to its definition, the force between two points with high WiFi similarity is enhanced when their physical distance increases. This violates the fact that the locations with high WiFi similarity should be nearby. Therefore, the inner trace matching is to minimize the overall force F to obtain

$$\theta' = \text{argmin}_\theta F. \tag{12}$$

3) Segment Rotation: To minimize the force, we keep rotating each segment within $[-10°, 10°]$ at a step of $1°$ around its first turn. The updated location of the nth turn is calculated by

$$\mathbf{T}'_n = \mathbf{S}''\left(\mathbf{T}_n - \mathbf{T}'_{n-1}\right) + \mathbf{T}'_{n-1} \tag{13}$$

where

$$\mathbf{S}'' = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

where $\mathbf{T}'_n$ is the updated location of the nth turn after rotation. Then, we calculate the points on the $(n+1)$th segment rotating around $\mathbf{T}'_n$ with θ as

$$\mathbf{L}'_i = \mathbf{S}''\left(\mathbf{L}_i - \mathbf{T}'_n\right) + \mathbf{T}'_n \tag{14}$$

where $\mathbf{L}_i$ is the location of the ith point on the indoor PDR trace and $\mathbf{L}'_i$ is the updated location. Then, we calculate the physical distance $D_{i,j}(\theta)$ between two points on the updated trace as

$$D_{i,j}(\theta) = \left\|\mathbf{L}'_j - \mathbf{L}'_i\right\|_2 \tag{15}$$

where $\|\cdot\|_2$ is two-norm of the vector. $D_{i,j}(\theta)$ is only related to θ because $\mathbf{L}'_i$ and $\mathbf{L}'_j$ are determined by θ based on (13) and (14). Therefore, after each rotation, we calculate the force by (11) and find θ' to minimize the force. Finally, each turn in the PDR trace is compensated with θ' and the coordinates of each point in the PDR trace is updated.

C. Intertrace Matching

After compensating the turning errors of the indoor PDR trace, intertrace matching is used to find the absolute heading direction and location of the indoor PDR trace. At this step, we leverage the WiFi fingerprints to match the indoor PDR trace with the outdoor–indoor ones based on a joint absolute heading direction and location estimation. We first decide the landmarks on the traces shared by the indoor and outdoor–indoor traces. For each point on the indoor PDR trace, we search for the point with highest cosine similarity of WiFi RSSI lists on the outdoor–indoor PDR traces and form a pair of anchor points. In those anchor point pairs, we select M^{WiFi} pairs with the highest similarity and M^{WiFi} is empirically set to 15.

We define the force in the ith pair of anchor nodes as

$$F'_i = \frac{E_i D_i}{\Lambda} \tag{16}$$

where E_i is the cosine similarity of WiFi RSSI lists and D_i is the physical distance between the ith pair of the anchor points. Then, the overall force is

$$F' = \sum_{i=1}^{M^{\text{WiFi}}} F'_i. \tag{17}$$

Similar as in Section V-B, our target in intertrace matching is to minimize F'. Because F'_i in (16) is only decided by the location $\mathbf{L}$ and the direction Ω^{inter} of the indoor PDR trace which determine D_i, intertrace matching translates and rotates the indoor PDR trace to minimize F' to obtain

$$\left\{\mathbf{L}', \Omega^{\text{inter'}}\right\} = \text{argmin}_{\{\mathbf{L},\Omega^{\text{inter}}\}} F'\left(\mathbf{L}, \Omega^{\text{inter}}\right). \tag{18}$$

To find the optimize solution of target (18), we need a proper initialization of the location and heading direction since (18) is not a global minimization.

1) Absolute Heading Direction Initialization via Enhanced Magnetic Interference Detection Algorithm: The PDR traces generated in Section IV only use gyroscope and accelerometer. Therefore, the absolute heading direction referenced to the global north is unknown. In this section, we design an offline procedure to fuse the magnetic field and estimate the absolute direction of the indoor PDR trace via an enhanced magnetic interference detection algorithm. The heading direction estimated by magnetometer is normally unreliable in indoor environments due to the influence of the iron materials. The influence can be classified into local and entire magnetic distortion.

Local magnetic distortion is caused by the surrounding iron materials such as elevators, computers, or doors in a limited indoor area and the estimated heading direction by fusing these magnetic field will be significantly affected. Therefore, we must detect the distorted areas and eliminate their influence. To detect the distorted magnetic field, we exploit two empirical observations: the magnitude of the distorted magnetic field is normally larger; the changes of the attitudes estimated, respectively, by magnetometer and gyroscope are significantly different in a short period in the distorted area [19]. Therefore, we select parts of the indoor PDR trace whose magnetic fields meet the following conditions:

$$\begin{cases} \mathrm{Mag}_k < K_M \\ \Delta\mathrm{Mag}_k < K_\Delta \end{cases} \tag{19}$$

where Mag_k is the magnitude of the kth magnetic sample and K_M is an empirical constant threshold. $\Delta\mathrm{Mag}_k$ is defined as the variance of the difference between the yaw angles estimated by magnetometer with accelerometer and gyroscope with accelerometer as

$$\Delta\mathrm{Mag}_k = \mathrm{var}\left(\mathbf{\Omega}_k^{\mathrm{mag}} - \mathbf{\Omega}_k^{\mathrm{gyro}}\right) \tag{20}$$

where $\mathbf{\Omega}_k^{\mathrm{mag}}$ is a moving window containing yaw angles estimated by magnetometer and accelerometer centering at the kth sample. $\mathbf{\Omega}_k^{\mathrm{gyro}}$ is for the yaw angles estimated by gyroscope and accelerometer. K_Δ in (19) is an empirical constant threshold. The selected magnetic field with condition (19) are processed to estimate the entire magnetic distortion as follows.

Entire magnetic distortion is caused by the biased indoor magnetic field from the global one due to the iron structure of building. We make use of the processed outdoor–indoor PDR traces in Section V-A as the reference to estimate the biased heading direction using magnetometer and accelerometer. The biased angle is calculated by

$$\Delta\Omega = \frac{\sum_{j=1}^{L'}\left(\Omega_j^{\mathrm{mag}} - \Omega_j^{\mathrm{oi}}\right)}{L'} \tag{21}$$

where Ω_j^{mag} is the yaw angle on the jth step estimated by magnetometer and accelerometer after filtering the local distorted magnetic field. Ω_j^{oi} is the corresponding yaw angle on the processed outdoor–indoor PDR trace in Section V-A and L' is the length of the angle vector.

After obtaining $\Delta\Omega$, we compute the average value of the differences between the yaw angles estimated by gyroscope and magnetometer (with accelerometer)

$$\Delta\Omega' = \frac{\sum_{j=1}^{L'}\left(\Omega_j^{\mathrm{gyro}} - \Omega_j^{\mathrm{mag}}\right)}{L'}. \tag{22}$$

Then the difference between the heading direction estimated by gyroscope with accelerometer and the global north is calculated by

$$\Delta\Omega^{\mathrm{initial}} = \Delta\Omega' + \Delta\Omega. \tag{23}$$

Finally, the indoor PDR trace is rotated with $\Delta\Omega^{\mathrm{initial}}$ to find its absolute direction.

2) Location Initialization: The location of the indoor PDR trace is initialized based on the difference about the centroids of the anchor points between the indoor and outdoor–indoor PDR traces as

$$\mathbf{L}' = \mathbf{L} + \frac{\sum_{i=1}^{M^{\mathrm{WiFi}}}\left(\mathbf{P}_i^{\mathrm{oi}} - \mathbf{P}_i^{\mathrm{in}}\right)}{M^{\mathrm{WiFi}}}$$

where $\mathbf{P}_i^{\mathrm{in}}$ is the ith anchor point on the indoor PDR trace and $\mathbf{P}_i^{\mathrm{oi}}$ is the corresponding anchor point on the outdoor–indoor PDR traces.

3) Fine Tune Location and Direction: According to the experimental results, we find that the aforementioned two initialization methods achieve good performance. Hence, the location and direction just require fine adjustment. The location of the indoor PDR trace is adjusted at a step of 1 m to west, east, south, and north within 5 m and the heading direction is adjusted at a step of 1° in both clockwise and anti-clockwise direction within 15°. We search for the optimized combination of the location and heading direction in these 600 states (5 m × 4 Directions × 15° × 2 Directions) to minimize F'.

Finally, we rotate and translate the indoor PDR trace based on $\Omega^{\mathrm{inter'}}$ and $\mathbf{L}'$.

D. Construction of Radio Map

According to the above procedure, we automatically generate the walking paths to label the radio map. We mark the location of the ith step point in the generated walking paths as $\mathbf{L}_i^{\mathrm{WP}} = (x_i^{\mathrm{WP}}, y_i^{\mathrm{WP}})$. For the ith step point, we obtain aN RSSI list comprised of all the scanned APs as $\mathbf{RSSI}_i = (\mathrm{RSSI}_i^1, \mathrm{RSSI}_i^2, \ldots, \mathrm{RSSI}_i^N)$ where N is the total number of the scanned APs. The radio map is constructed by labeling the RSSI lists with the corresponding locations of their step points in the walking paths as

$$\mathbf{RadioMap} = \begin{pmatrix} \mathbf{L}_1^{\mathrm{WP}} & \mathbf{RSSI}_1 \\ \mathbf{L}_2^{\mathrm{WP}} & \mathbf{RSSI}_2 \\ \cdots & \\ \mathbf{L}_M^{\mathrm{WP}} & \mathbf{RSSI}_M \end{pmatrix} \tag{24}$$

where M is the total number of step points in the generated walking paths.

VI. Seamless Outdoor–Indoor Positioning Based on Multisensor Fusion

We consider the problem of tracking a mobile device over time given a stream of noisy inertial sensor data and the

location measurements from outdoor GPS and indoor WiFi fingerprinting. We solve this problem by fusing these information by an enhanced particle filter.

A. Basic Concept of Particle Filter

At time k, we have an unknown system state vector $\mathbf{x}_k$ including the target's location and a discrete sequence of noisy measurement vectors $\mathbf{z}_{1:k}$ taken at times $1, \ldots, k$, describing the locations from GPS or WiFi fingerprinting.

To define the tracking problem, we consider the evolution of the state sequence $\{\mathbf{x}_k, k \in \mathbb{N}\}$ given by

$$\mathbf{x}_k = \mathbf{f}_k(\mathbf{x}_{k-1}, \mathbf{v}_k). \tag{25}$$

$\mathbf{f}_k$ is a possibly nonlinear function to define the evolution process from state $\mathbf{x}_{k-1}$ to state $\mathbf{x}_k$, which is called system model. $\mathbf{v}_k$ is an independent and identically distributed (i.i.d.) process noise sequence, and $\mathbb{N}$ is the set of natural numbers. The tracking objective is to recursively estimate $\mathbf{x}_k$ from the measurements

$$\mathbf{z}_k = \mathbf{h}_k(\mathbf{x}_k, \mathbf{u}_k). \tag{26}$$

$\mathbf{h}_k$ is a possibly nonlinear function of the relation between $\mathbf{x}_k$ and the corresponding measurements $\mathbf{z}_k$, which is called observation model. $\mathbf{u}_k$ is an i.i.d. measurement noise sequence.

From the perspective of Bayesian estimation, our goal is to calculate the posterior probability density function (PDF) $p(\mathbf{x}_k|\mathbf{z}_{1:k})$. In particle filter, a sequential importance learning (SIS) algorithm [42] is adopted to obtain $p(\mathbf{x}_k|\mathbf{z}_{1:k})$, which is a Monte Carlo (MC) method. Based on the principle of importance sampling [42], $p(\mathbf{x}_k|\mathbf{z}_{1:k})$ can be approximated by

$$p(\mathbf{x}_k|\mathbf{z}_{1:k}) \approx \sum_{i=1}^{N_s} w_k^i \delta\left(\mathbf{x}_k - \mathbf{x}_k^i\right) \tag{27}$$

where $\mathbf{x}_k^i$ is the ith particle and w_k^i is the *associated weight*. N_s is the total number of particles. Assuming that the initial PDF $p(\mathbf{x}_0|\mathbf{z}_0) \equiv p(\mathbf{x}_0)$ is known, we can obtain $p(\mathbf{x}_k|\mathbf{z}_{1:k})$ recursively via predicting and updating the particles, respectively.

For the predicting stage, the particles $\mathbf{x}_{k-1}$ are updated based on the system model in (25) to obtain the current $\mathbf{x}_k$. Afterwards, in the updating stage, the associated weight w_k^i for each particle is updated. For the bootstrap particle filter (BPF) [42], [43], which is commonly used and efficiently implementable, the associated weights are updated by

$$w_k^i \propto w_{k-1}^i p\left(\mathbf{z}_k|\mathbf{x}_k^i\right) \tag{28}$$

where w_k^i are only determined by the *likelihood* function of $p(\mathbf{z}_k|\mathbf{x}_k^i)$ obtained based on the observation model in (26).

Importantly, a particle filter is normally prone to the sample degeneracy problem, which often results in serious performance degradation. To deal with this problem, resampling is typically adopted [44]. A suitable measurement of the degeneracy is the effective sample size $N_{\text{eff}} = 1/\sum_{i=1}^{N_s}(w_k^i)^2$. As soon as N_{eff} is smaller than $0.5 \times N_s$, the degeneracy is considered to be serious and a suitable resampling method should be adopted. A systematic resampling method [44] is adopted in this paper, because of its high accuracy and efficient implementation.

In this paper, we propose an enhanced particle filter for seamless indoor and outdoor positioning. We provide enhanced solutions for both system model and observation model in the particle filter. For the system model, we design an enhanced PDR algorithm to model the evolution of the state sequence. For the observation model, we propose two kinds of likelihoods for outdoor with GPS and indoor with WiFi fingerprints. An enhanced IO detector is adopted for switching the likelihoods. In the following sections, we introduce the enhanced particle filter in the aspects of system model and observation model.

B. Pedestrian Dead Reckoning-Based System Model

1) System Model: In the system, the state $\mathbf{x}_k$ is defined as

$$\mathbf{x}_k = \left[x_k, y_k\right] \tag{29}$$

where x_k is the latitude and y_k is the longitude of the target user at the kth step. The particles are updated based on the system model when a step is detected in the PDR algorithm. The system model relies on the step length l and the heading direction θ estimated in the PDR algorithm to update the latitude and longitude, i.e.,

$$\begin{cases} x_k = x_{k-1} + \frac{l\cdot\sin(\theta+\Delta\Phi)}{110540} + n_x \\ y_k = y_{k-1} + \frac{l\cdot\cos(\theta+\Delta\Phi)}{111320\cdot\cos(x_{k-1})} + n_y \end{cases} \tag{30}$$

where $\boldsymbol{n} = [n_x, n_y]$ is the Gaussian noise vector on the latitude and longitude, and $\Delta\Phi$ is the local declination. Note that the constants 110 540 and 111 320 reflect the earth's oblateness.

2) Real-Time PDR With Enhanced X-AHRS: For the PDR algorithm driving the system model, we adopt the same step detection and step length estimation methods as in Section IV but for the heading direction estimation we fuse data from magnetometer to obtain the absolute direction because the absolute heading direction is required and offline processing is not available for real-time indoor tracking.

We design an enhanced X-AHRS filter with adaptive β to fuse data from gyroscope, accelerometer and magnetometer where β is a tradeoff parameter to tune the impact of magnetometer data [16]. First, the biased angle of $\Delta\Omega$ in (21) due to the entire magnetic distortion is compensated. Then, we utilize condition (19) to detect the local distorted magnetic filed. In the distorted area, β is set to 0.01. Otherwise, β is set to 0.05. Note that in X-AHRS [16] the estimated heading direction will be more affected by magnetometer with increasing β.

C. GPS and WiFi Fingerprinting-Based Observation Model

After updating the particles, the associated weights are updated according to the location observations (GPS or WiFi fingerprinting). Different observation models are designed for indoor and outdoor environments.

1) Indoor and Outdoor Detection: IO detector is used to switch the observation models based on GPS or WiFi fingerprinting. For IO detector, we first adopt the proximity sensor and clock in a smartphone to check the availability of light sensor. If there are not obstacles in front of the smartphone detected by the proximity sensor during the daytime (8:00 A.M. to 5:00 P.M.), light intensity detected by the light

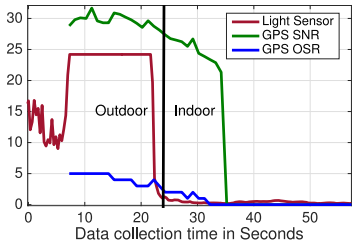


Fig. 8. Light intensity, SNR, and OSR of GPS changes in indoor and outdoor (light intensity is obtained by light sensor data divided by 1000 to scale the data.)

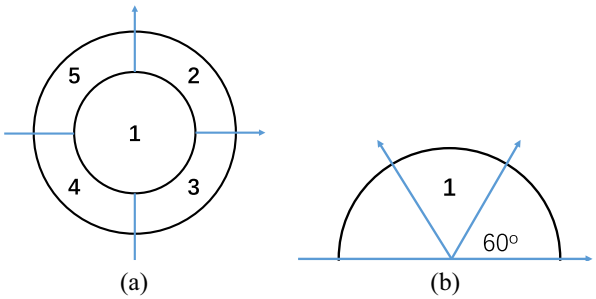


Fig. 9. OSR regions. (a) Five regions in OSR. (b) Region 1 in OSR.

sensor is used in IO detector. If light intensity is larger than 1000lx, the user is detected outdoor. Otherwise, the user is indoor. During the night (8:00 P.M. to 5:00 A.M.), we find that light intensity indoor is typically higher than that outdoor, especially for shopping malls. If light intensity is larger than 500lx, the user is detected indoor. Otherwise, the user is outdoor. Note that IO detector based on light intensity, namely light IO detector, is most accurate and with lowest latency compared with the other signals. Therefore, in this paper, light intensity is the only parameter used for IO detector as long as it is available.

If the light sensor is not available, we provide an IO detector integrating GPS signals and magnetic field. We, respectively, design three detectors, namely SNR detector, OSR detector, and MAG detector, in which the first two detectors rely on GPS signals and the last one relies on the magnetic field.

1) *SNR Detector:* According to Fig. 8, the mean SNR of GPS signals indoor is clearly smaller than that outdoor. Therefore, we calculate the mean SNR of GPS signals from all the detected satellites. If the mean SNR is smaller than 26 dBm, the SNR detector outputs indoor. Otherwise, it outputs outdoor. Note that if the number of the detected satellites is smaller than 5, the SNR detector outputs indoor.
2) *OSR Detector:* We design a novel parameter open sky ratio (OSR) to indicate the distribution of the detected satellites in the sky. According to the direction angle and elevation angle of GPS satellites, we divide the celestial sphere into five regions as shown in Fig. 9(a). The satellite with elevation angle over 60° is in region 1 as shown in Fig. 9(b). The remaining four regions are determined by direction angle and each region occupies 90°. If there is a satellite with an SNR exceeding 20 dBm in a certain area, the sky in the area is considered to be unobstructed. The OSR takes a value of 0–5. As shown in Fig. 8, the OSR value clearly decreases when a user walks from outdoor to indoor. In this paper, we set an OSR threshold of 2 for indoor (equal to or smaller than 2) and outdoor switching (larger than 2).
3) *MAG Detector:* Similar as the observations of magnetic field in [37], the variance of magnetic field indoor is clearly larger than that outdoor. Therefore, we implement an MAG detector the same as in [37] but with a moving window of 4 s for calculating the variance of magnetic field. We refer to [37] for more details about MAG detector.

Finally, we combine these three detectors based on a voting approach, namely SNR-OSR-MAG IO detector. If more than two detectors vote for indoor, the IO detector outputs indoor. Otherwise, the IO detector outputs outdoor.

2) Outdoor Adaptive Likelihood From GPS: Once a user is detected outdoor, GPS locations are used to define $p(\mathbf{z}_k|\mathbf{x}_k^i)$ for updating associated weights. Assuming that the latitude and longitude observations are independent, the likelihood is

$$p(\mathbf{z}_k|\mathbf{x}_k^i) = p\left(x_k^{\text{GPS}}|x_k^i\right)p\left(y_k^{\text{GPS}}|y_k^i\right) \tag{31}$$

where $\mathbf{z}_k = [x_k^{\text{GPS}}, y_k^{\text{GPS}}]$ are the latitude and longitude observed from GPS. The individual likelihoods on latitude and longitude are assumed to follow the Gaussian distribution:

$$\begin{cases} p\left(x_k^{\text{GPS}}|x_k^i\right) \sim N\left(0, \sigma^2\right) \\ p\left(y_k^{\text{GPS}}|y_k^i\right) \sim N\left(0, \sigma^2\right) \end{cases} \tag{32}$$

where σ is the standard deviation of the Gaussian noise for GPS positioning and set based on the GPS accuracy to tune the contribution of the GPS observation on the estimated location. The associated weights are only updated when condition (4) meets (P^{Thre} is set to 20 m in this part) and σ is set as

$$\sigma = 0.000025 \times P \tag{33}$$

where P is the Android-reported GPS accuracy.

3) WiFi Fingerprinting-Based Indoor Likelihood: Once a user is detected indoor, the user location estimated by WiFi fingerprinting is used to define $p(\mathbf{z}_k|\mathbf{x}_k^i)$ for updating the associated weights. Similar as the GPS likelihood, this likelihood can be written as

$$p(\mathbf{z}_k|\mathbf{x}_k^i) = p\left(x_k^{\text{WiFi}}|x_k^i\right)p\left(y_k^{\text{WiFi}}|y_k^i\right) \tag{34}$$

where $\mathbf{z}_k = [x_k^{\text{WiFi}}, y_k^{\text{WiFi}}]$ are the latitude and longitude estimated by WiFi fingerprinting with a commonly used weighted KNN algorithm [4]. The individual likelihoods on latitude and longitude are assumed to follow the Gaussian distribution as:

$$\begin{cases} p\left(x_k^{\text{WiFi}}|x_k\right) \sim N\left(0, \sigma'^2\right) \\ p\left(y_k^{\text{WiFi}}|y_k\right) \sim N\left(0, \sigma'^2\right) \end{cases} \tag{35}$$

where σ' is the standard deviation of the Gaussian noise for WiFi fingerprinting.

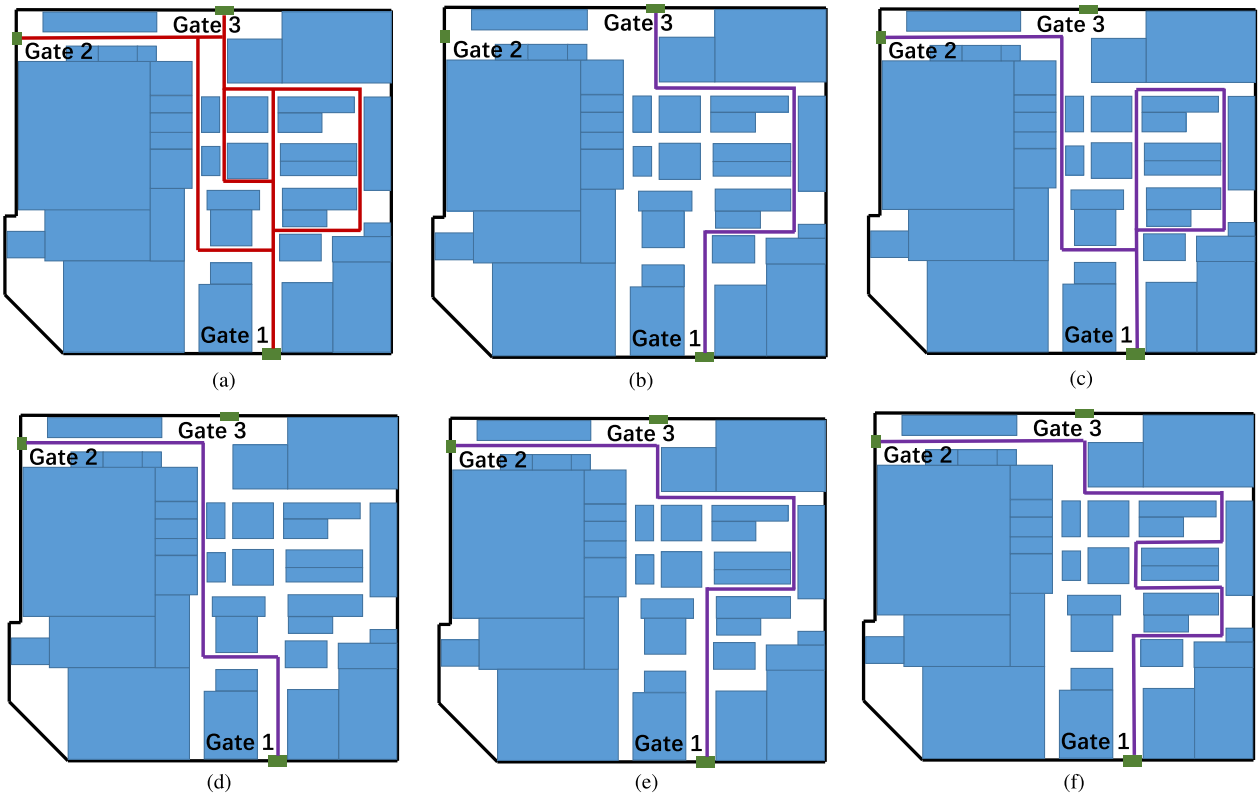


Fig. 10. Walking paths and testing paths in scenario 1. (a) Ground truth walking paths. (b) Testing path 1. (c) Testing path 2. (d) Testing path 3. (e) Testing path 4. (f) Testing path 5.

VII. Experiments and Evaluation Results

To evaluate the proposed SoiCP system, we conduct a set of comprehensive experiments in two shopping malls of 3600 m^2 (scenario 1) and 6000 m^2 (scenario 2) for crowdsensing indoor radio map and seamless outdoor–indoor positioning.

A. Experimental Setup

In our experiments, we collect raw sensor data from Huawei Mate8 and Mate10 smartphones with their accelerometer, gyroscope, magnetometer, light sensor, WiFi, and GPS. All these raw sensor data are sent and stored in a cloud server running in the University of Bern for offline processing including trace matching and positioning. All the algorithms are designed in Python. We first utilize the offline PDR algorithm to process the inertial sensor data to generate the user traces. Since all the raw sensor data are timestamped with their collecting time in smartphones, each step on the user traces is timestamped. Then, each detected step point in the traces is assigned with only one scanned WiFi list of the nearest timestamp. In our experiments, the users continuously walk but occasional stops during trace collection will not affect the trace matching algorithm because one step point is only attached with one WiFi list of the nearest timestamp.

In scenario 1, we collect seven outdoor–indoor PDR traces where five different users walk through the shopping mall and one indoor PDR trace where a user walks along the indoor paths for four rounds. Note that the outdoor–indoor PDR traces include long outdoor parts. The ground truth walking paths covered by these PDR traces are shown as the red lines in Fig. 10(a). These eight PDR traces are used to generate the walking paths and extract the radio map. Similar as scenario 1, we collect seven outdoor–indoor PDR traces and one indoor-outdoor PDR trace from five different users to construct the walking paths in scenario 2.

After constructing the radio map, we conduct testing experiments in which users, respectively, walk along five different paths indicated in Fig. 10(b)–(f) to evaluate the accuracy of the proposed seamless outdoor–indoor positioning algorithm based on the enhanced particle filter in scenario 1. Note that the testing paths 1–3 are entirely consistent with the walking paths in the radio map but parts of the testing paths 4 and 5 are different from the walking paths in the radio map. In total, the length of testing paths in scenario 1 is about 630 m and the duration is 449 s. The positioning accuracy of the SoiCP system is evaluated on 903 testing points along the five paths. To test the robustness of the system, we further evaluate the positioning algorithm in scenario 2 with five testing traces in Fig. 11(b)–(f). WiFi scanning is conducted in every 3 s and a GPS positioning in every second outdoor. Note that the users hold phones horizontally in all experiments.

To obtain the ground truth locations for evaluating the positioning accuracy, the moving paths are predefined, in which the coordinates of all turning points are measured. Then, the ground truth coordinates of the other positions along the paths are obtained by interpolation. To know the time when the user passes each ground truth position, we record the time passing each turn and keep the moving speed constant for the whole

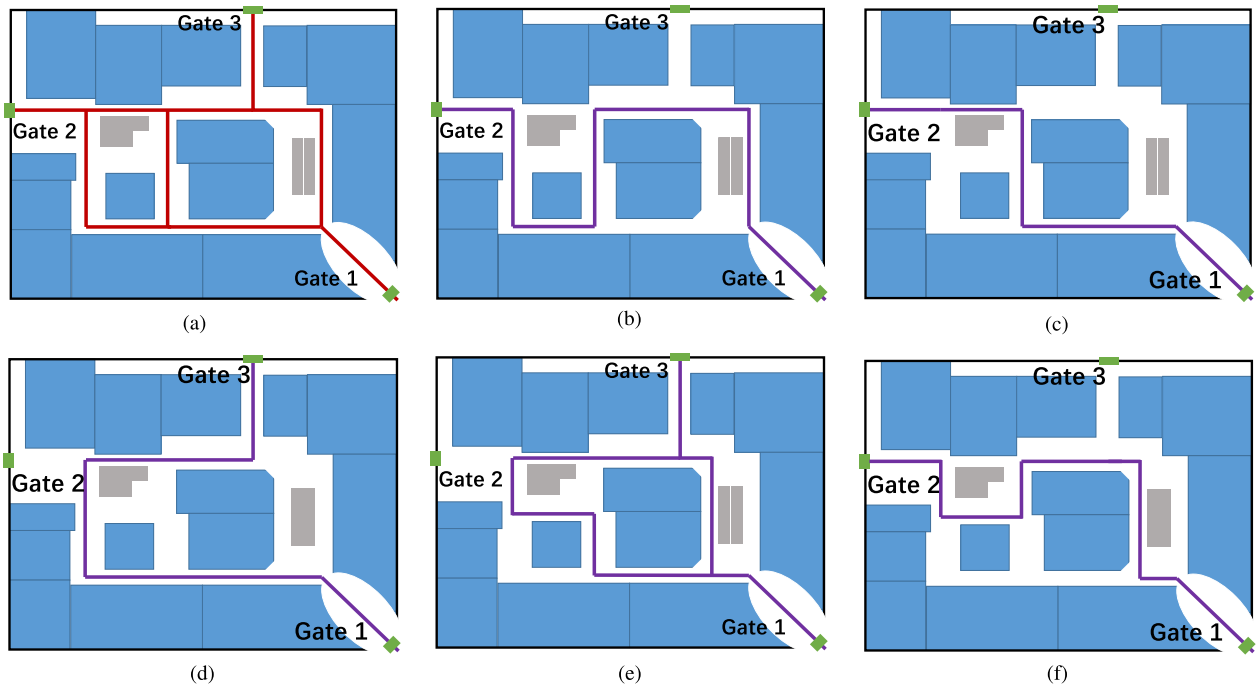


Fig. 11. Walking paths and testing paths in scenario 2. (a) Ground truth walking paths. (b) Testing path 1. (c) Testing path 2. (d) Testing path 3. (e) Testing path 4. (f) Testing path 5.

moving path. All the WiFi RSSI lists are timestamped and mapped to each ground truth position based on the time when the user passes.

B. Experiments

1) Preliminary Analysis of PDR: In SoiCP, we propose two kinds of PDR algorithms, i.e., offline and real-time PDR, respectively, in crowdsensing radio map and seamless outdoor–indoor positioning. We first evaluate the accuracy of the proposed step detection algorithm introduced in Section IV-B3 based on the testing traces in both Scenarios. According to Table II, we can detect the steps with an accuracy higher than 98%.

We further evaluate the heading direction errors of the proposed offline and real-time PDR algorithms compared with the traditional X-AHRS. For offline PDR, the heading directions of five traces are first estimated based on gyroscope and accelerometer in Section IV-B3. Then, the generated traces are further rotated offline based on the absolute heading direction initialization algorithm in Section V-C1. For real-time PDR, heading direction is calculated based on the enhanced X-AHRS algorithm with adaptive β in Section VI-B.

According to Tables II and III, the heading direction errors of the traditional X-AHRS without magnetic interference detection are up to 34° and 36°, respectively, in scenarios 1 and 2. Moreover, the heading directions are biased due to the entire magnetic distortion. Our proposed magnetic interference detection and compensation algorithms in both offline and real-time PDR can significantly reduce the heading direction errors as shown in Tables II and III. In scenario 1 the mean heading direction errors of offline PDR are within 13° and that of real-time PDR are within 14°. The similar findings can be found in scenario 2 as in Table III, i.e., the mean heading direction errors of offline PDR within 12° and that of real-time PDR within 13°.

TABLE II
STEP DETECTION AND MEAN DIRECTION ERRORS IN SCENARIO 1

	Trace 1	Trace 2	Trace 3	Trace 4	Trace 5
Ground truth steps	132	215	147	188	207
Estimated steps	133	217	149	191	210
Mean direction error (Offline-SoiCP)	10°	13°	7°	8°	9°
Mean direction error (Realtime-SoiCP)	14°	11°	9°	8°	10°
Mean direction error (X-AHRS)	34°	32°	28°	26°	29°

TABLE III
STEP DETECTION AND MEAN DIRECTION ERRORS IN SCENARIO 2

	Trace 1	Trace 2	Trace 3	Trace 4	Trace 5
Ground truth steps	163	152	171	220	155
Estimated steps	165	153	174	223	156
Mean direction error (Offline-SoiCP)	10°	9°	8°	7°	12°
Mean direction error (Realtime-SoiCP)	9°	10°	9°	10°	13°
Mean direction error (X-AHRS)	31°	30°	26°	28°	36°

2) Accuracy of Crowdsensing Walking Paths: Fig. 12(a)–(e) indicates the changes of the PDR traces by utilizing our proposed three-step trace matching for crowdsensing walking paths in SoiCP in scenario 1. Note that the O-I trace represents the outdoor–indoor trace.

Fig. 12(a) shows the original PDR traces. We initialize the starting points of all outdoor–indoor PDR traces to their first GPS locations. For the indoor PDR trace, we initialize the starting point to a random location near the target building. Note that the outdoor–indoor GPS traces include long outdoor

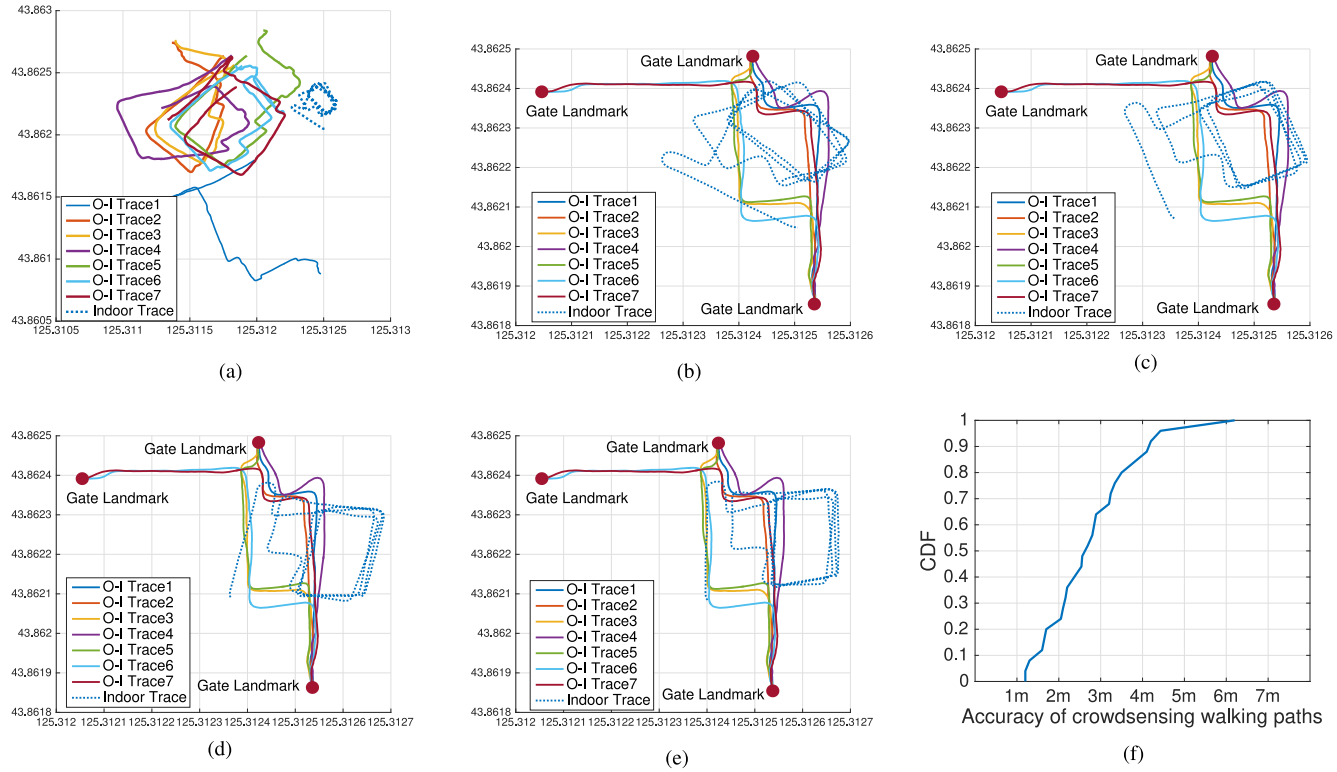


Fig. 12. Procedure of crowdsensing walking paths in scenario 1. (a) Original traces. (b) Intratrace matching. (c) Inner trace matching. (d) Intertrace matching (initialization). (e) Intertrace matching (final). (f) Errors of generated walking paths.

TABLE IV
ACCURACY OF DETECTED GATE LOCATIONS

Scenario 1	Gate number	Gate 1	Gate 2	Gate 3
	Location accuracy	$2.8m$	$2.04m$	$3.07m$
Scenario 2	Gate number	Gate 1	Gate 2	Gate 3
	Location accuracy	$2.43m$	$2.11m$	$2.72m$

parts. The raw PDR traces in Fig. 12(a) are noisy, randomly scattered around the target building, and cannot represent the ground truth walking paths.

In the three-step trace matching algorithm, the intratrace matching is to process the outdoor–indoor PDR traces. We detect the gate points based on the light sensors and adopt DBSCAN to cluster the gate points based on the Euclidean distances of WiFi RSSI lists. We successfully detect all 14 gate points on the seven outdoor–indoor PDR traces. Using DBSCAN, we successfully cluster the 14 gate points into three clusters with 100% accuracy. Then, we calculate the average location of each cluster as the location of each gate. Table IV shows that our proposed matching approach can accurately cluster and estimate the locations of the gates with accuracy in 2–3 m. Based on the gate points, we rotate, translate and scale the outdoor–indoor PDR traces as introduced in Section V-A and further remove their outdoor parts. The results in Fig. 12(b) illustrate that the outdoor–indoor PDR traces are well merged and parts of walking paths are generated.

Until now, we have not processed the indoor PDR trace. We can find that the shape of the indoor PDR trace is distorted compared with the ground truth walking path in Fig. 12(b). In SoiCP, we take inner trace matching to compensate the turning errors in the indoor PDR trace by minimizing the overall force described in Section V-B. Fig. 12(c) shows the processed indoor PDR trace by inner trace matching, which shows a more similar shape, i.e., smaller turning error, as the ground truth path than the raw indoor PDR trace in Fig. 12(b).

Finally, intertrace matching is used to match the processed indoor PDR trace with the outdoor PDR traces. For intertrace matching, the indoor PDR trace is first initialized with the location and heading direction introduced in Section V-C. Fig. 12(d) indicates the initialized indoor PDR trace rotated with an angle error of around 12° because of the distorted indoor magnetic field. After the initialization, the location and heading direction of the indoor PDR trace are further fine-tuned by the method introduced in Section V-C3. Fig. 12(e) shows the final generated walking paths in which the indoor PDR trace is well translated and rotated by minimizing the overall force in Section V-C. The crowdsensed walking paths based on the three-step trace matching in Fig. 12(e) match with the ground truth paths well as shown in Fig. 10(a).

Because the turning points on the PDR traces and ground truth walking paths are the most significant landmarks to obtain their distances, we calculate the errors of turning points on the generated walking paths referenced to the ones on the ground truth walking paths. The cumulative distribution function (CDF) of errors is given in Fig. 12(f). According to this figure, the median accuracy of the walking paths generated by our proposed algorithms achieves 2.7 m, the mean accuracy achieves 2.8 m, and 90% of errors are smaller than 4.1 m.

To evaluate the robustness of the proposed algorithm, we further crowdsense the walking paths in scenario 2 as shown

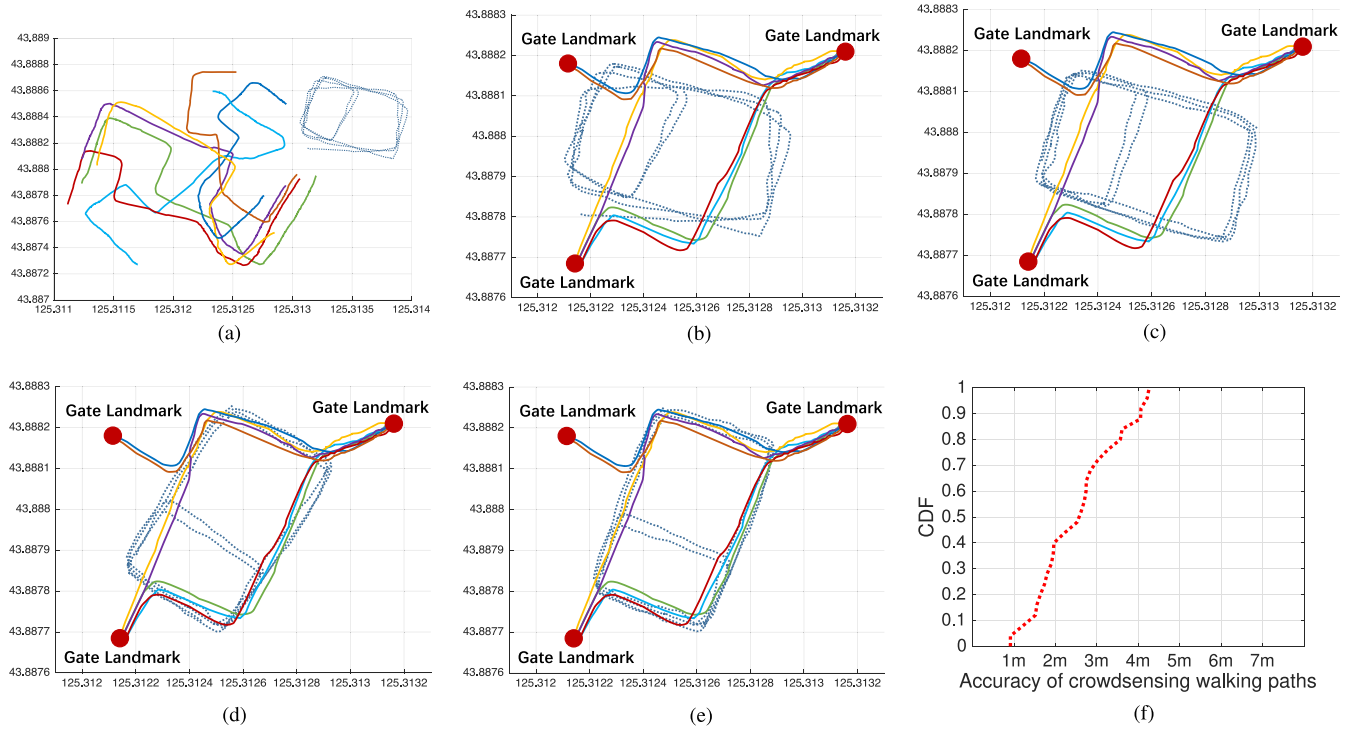


Fig. 13. Procedure of crowdsensing walking paths in scenario 2. (a) Original traces. (b) Intratrace matching. (c) Inner trace matching. (d) Intertrace matching (initialization). (e) Intertrace matching (final). (f) Errors of generated walking paths.

TABLE V
MEAN ACCURACY OF WALKING PATHS WITH DIFFERENT NUMBER OF APS

Scenario 1	Number of APs	294	200	150	100	50
	Mean accuracy	2.8m	2.8m	2.9m	3.2m	6.1m
Scenario 2	Number of APs	531	300	150	100	50
	Mean accuracy	2.6m	2.7m	3.3m	4.1m	7.0m

in Fig. 11(a). Fig. 13(a)–(e) indicates the process of our three-step trace matching for crowdsensing walking paths in scenario 2 and Fig. 13(f) shows the CDF of the errors for the generated walking paths in scenario 2. In this scenario, the median accuracy of walking paths achieves 2.6 m, the mean accuracy achieves 2.6 m, and 90% of the errors are smaller than 4.0 m, which demonstrate the robustness of our system in different scenarios.

Moveover, the aforementioned accuracy of walking paths are obtained based on all the available WiFi access points in both scenarios, i.e., 294 APs (scanned) in scenario 1 and 531 APs in scenario 2. Since the density of WiFi APs plays an important role in both inner and intratrace matching, we randomly remove certain number of APs in both scenarios to evaluate its impact on the accuracy of walking paths. Table V shows the mean accuracy of the generated walking paths in both scenarios with different number of APs. According to Table V, the mean accuracy of the generated walking paths degrades with less number of APs. On the other hand, we find that in scenario 1 when the number of APs is larger than 150, the performance of the trace matching algorithm is still robust and its accuracy is higher than 3 m. When the number of APs gets too small, both inner and intratrace matching cannot well merge the traces and hence the accuracy becomes significantly lower. For example, when the number of APs is 50, the mean accuracy reduces to 6.1 m. We observes the similar findings in scenario 2 with larger area. When the number of APs is larger than 300, the mean accuracy is higher than 3 m, and the performance is robust. However, when the number of APs is 50, the mean accuracy reduces to 7.0 m.

3) Accuracy of Seamless Outdoor–Indoor Positioning:

a) IO detector: In this part, we first evaluate the performance of the proposed IO detector. According to some preliminary tests, we find that the light IO detector can accurately detect the indoor and outdoor status of users with very short delay (within 0.8 s). On the other hand, if the light IO detector is unavailable, SNR-OSR-MAG IO detector can obtain accurate results (100%) in the sealed-off shopping malls (indoor) and outdoor environments with clear sky. However, it encounters larger delay than light IO detector when the user switches from outdoor to indoor, i.e., at the entrance of the shopping mall. Tables VI and VII show the delays of SNR-OSR-MAG IO detector when the users switch from outdoor to indoor on the testing traces. According to the tables, the delays range from 2.2 to 3.9 s. Additionally, we compare our proposed SNR-OSR-MAG IO detector to the Cellular-MAG detector [37]. Note that the light sensor is not used in Cellular-MAG detector which is different from [37] because SNR-OSR-MAG IO detector is only used when the light sensor is not available. Since the received power of cellular signals depends on the layout of base stations, we find that it decreases more slowly and is less stable compared with GPS signals in our testing scenarios, and hence introduces longer delay. According to the testing results in Tables VI and VII,

TABLE VI
DELAY OF IO DETECTOR IN SCENARIO 1

	Trace1	Trace2	Trace3	Trace4	Trace5	Overall
SNR-OSR-MAG	3.1s	2.2s	3.8s	3.1s	3.9s	3.22s
Cellular-MAG	5.1s	5.8s	4.9s	6.3s	4.2s	5.26s

TABLE VII
DELAY OF IO DETECTOR IN SCENARIO 2

	Trace1	Trace2	Trace3	Trace4	Trace5	Overall
SNR-OSR-MAG	2.6s	2.4s	3.1s	2.8s	2.5s	2.68s
Cellular-MAG	6.1s	7.2s	5.8s	6.1s	5.2s	6.08s

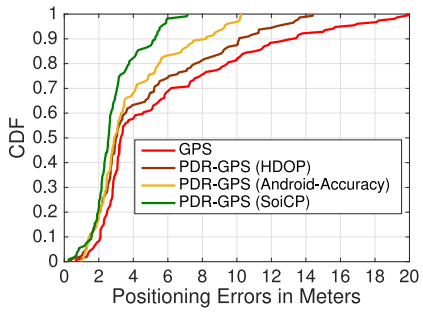


Fig. 14. Different positioning methods in outdoor environments.

the delays of SNR-OSR-MAG IO detector are clearly shorter than those of the Cellular-MAG detector. IO detector with shorter delay will switch the indoor and outdoor positioning techniques more smoothly.

b) Outdoor positioning accuracy: Then, we evaluate the accuracy of the proposed positioning algorithm in SoiCP based on the outdoor parts of the ten traces in the two testing scenarios. Fig. 14 shows the positioning accuracy. Note that PDR-GPS (SoiCP) is referred to the proposed seamless outdoor–indoor positioning algorithm with the enhanced GPS accuracy detection method in SoiCP. We further implement particle filters fusing GPS and PDR only relying on the Android-reported accuracy and horizontal dilution of precision (HDOP) as in the work of [29] for comparison. According to the evaluation results, GPS achieves a median accuracy of 3.1 m in our testing environments but the accuracy rapidly decreases when the users are approaching and leaving the buildings. We observe a maximum positioning error around 20 m. Based on the particle filters fusing GPS and PDR, the accuracy improves especially in the areas around the buildings, because the impact of GPS locations is reduced in the particle filters based on the detected GPS accuracy. In these three GPS accuracy detection algorithms, we find that Android-reported accuracy outperforms HDOP because the Android-reported accuracy considers more factors besides HDOP such as SNR and pseudo ranges. Our proposed solution in SoiCP significantly outperforms both PDR-GPS fusion with Android-reported accuracy and HDOP because our proposed enhanced GPS accuracy detection algorithm is more reliable and can better filter the inaccurate GPS locations in the particle filter.

c) Indoor positioning accuracy: With the radio map labeled by the walking paths in Fig. 12(e), we evaluate three positioning algorithms, PDR-GPS-WiFi fusion introduced in Section VI (under light IO detector), PDR-GPS fusion and WiFi fingerprinting. Note that the only difference between PDR-GPS-WiFi fusion and PDR-GPS fusion is that PDR-GPS fusion does not include WiFi fingerprinting in the observation model of the particle filter in the indoor environment.

TABLE VIII
MEAN POSITIONING ACCURACY IN SCENARIO 1

	Trace 1	Trace 2	Trace 3	Trace 4	Trace 5	Overall
PDR-GPS-WiFi	2.7m	2.5m	2.9m	3.4m	3.7m	3.0m
WiFi	3.4m	3.4m	3.5m	4.9m	4.4m	3.9m
PDR-GPS	12.4m	6.6m	5.3m	5.3m	6.4m	6.9m

CDFs of the positioning errors are shown in Fig. 15 in which Fig. 15(a) indicates the positioning errors for the traces 1–3, Fig. 15(b) for the traces 4 and 5, and Fig. 15(c) for all the traces in scenario 1. Table VIII illustrates the mean positioning errors for all five traces with three positioning algorithms.

First, PDR-GPS fusion cannot easily achieve high positioning accuracy since it merely relies on PDR in the indoor environment. Two reasons result in the low positioning accuracy: 1) accumulated positioning error of PDR and 2) low GPS positioning accuracy around the building. Besides the errors of PDR introduced in Section VII-B1, GPS positioning around the tall building is still difficult to achieve high accuracy, although we filter some inaccurate GPS locations based on the proposed GPS accuracy detection provided in Section VI. Therefore, the indoor PDR positioning is initialized with a GPS location with low accuracy and the positioning error is continuously accumulated. Because there is no any other information used to calibrate the indoor PDR positioning, GPS-PDR fusion can only achieve a mean positioning accuracy of 6.9 m for the indoor environment.

Then, we evaluate the positioning accuracy of WiFi fingerprinting. Note that the WiFi sampling rate is set one sample per 3 s and all the estimated locations between two WiFi samples are based on the first WiFi sample. We can find that the positioning accuracy of WiFi fingerprinting is from 3.4 to 4.9 m for the five testing traces which are significantly higher than that of the PDR-GPS fusion algorithm. However, due to the low sampling rate of WiFi, it cannot accurately track the user within a WiFi sampling duration (3 s). Additionally, the positioning accuracy of traces 1–3 is higher than that of traces 4 and 5 because parts of traces 4 and 5 are not consistent with the radio map. Moreover, due to the low sampling rate, the users' location can only update every 3 s, which introduce a long latency.

Our proposed PDR-GPS-WiFi fusion algorithm achieves a mean positioning accuracy of 3.0 m as shown in Table VIII and a median accuracy of 2.8 m as shown in Fig. 15(c). It obviously outperforms the other two positioning algorithms as shown in Fig. 15. Compared to WiFi fingerprinting, the locations of the user within one WiFi sampling interval are tracked by fusing PDR in PDR-GPS-WiFi. Hence the positioning accuracy is higher. According to Fig. 15(a) and (b), the improvement of accuracy for traces 4 and 5 is more significant than that of trace 1–3. This means that PDR-GPS-WiFi can better deal with the case where the user is not throughout walking on the paths of radio map than WiFi fingerprinting.

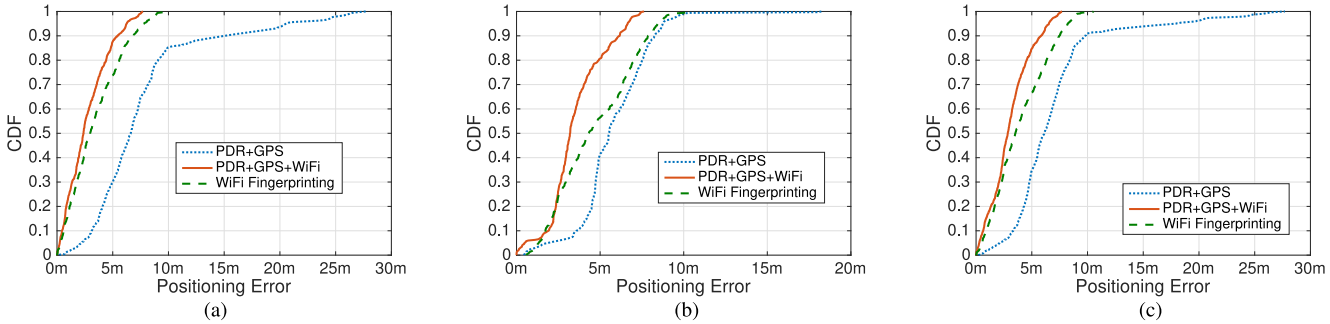


Fig. 15. Positioning errors in scenario 1. Positioning errors for (a) paths 1–3 and (b) paths 4 and 5. (c) Overall positioning errors.

TABLE IX
MEAN POSITIONING ACCURACY IN SCENARIO 2

	Trace 1	Trace 2	Trace 3	Trace 4	Trace 5	Overall
PDR-GPS-WiFi	2.8m	2.4m	2.7m	3.1m	3.4m	2.9m
WiFi	3.4m	3.1m	3.3m	3.7m	4.1m	3.5m
PDR-GPS	7.6m	6.1m	5.7m	6.8m	9.7m	7.2m

Compared to PDR-GPS, the positioning accuracy of PDR-GPS-WiFi is also higher because the WiFi fingerprinting locations can calibrate the PDR traces. Considering the updating rate, PDR-GPS-WiFi fusion algorithm can update locations as soon as the users walk a step (normally in 0.5 s) whose latency is much shorter than WiFi fingerprinting.

Finally, we evaluate the proposed PDR-GPS-WiFi fusion in scenario 2 and summarize the mean positioning accuracy in Table IX. According to Table IX, we observe similar results as in scenario 1 that our proposed PDR-GPS-WiFi fusion algorithm with a mean accuracy of 2.9 m clearly outperforms PDR-GPS (7.2 m) and WiFi fingerprinting (3.5 m) in this scenario. Moreover, the positioning accuracy of traces 1–3 based on both the PDR-GPS-WiFi and WiFi fingerprinting algorithms is higher than that of traces 4 and 5 because parts of traces 4 and 5 are not consistent with the radio map. These findings demonstrate the robustness of the proposed positioning algorithms.

VIII. CONCLUSION

In this paper, we present a novel system to achieve a SoiCP without site surveying. SoiCP can crowdsense the indoor walking paths based on the proposed three-step trace matching algorithm. The algorithm leverages gate points and WiFi fingerprints to merge noisy traces. According to the real-world experiments in two large-scale shopping malls, SoiCP can accurately generate the indoor walking paths with mean accuracies of 2.8 and 2.6 m without any priori knowledge of floor plans. The radio maps for WiFi fingerprinting can be generated by labeling the WiFi RSSI lists with its locations on the crowdsensed walking paths. Furthermore, in SoiCP system, we introduce an enhanced particle filter to fuse PDR, GPS, and WiFi fingerprinting for seamless outdoor–indoor positioning. Based on the crowdsensed radio map, we obtain a mean positioning accuracy of 3.0 m and a median accuracy of 2.8 m in the indoor environment of scenario 1, which significantly outperforms WiFi fingerprinting and the algorithm merely fusing GPS and PDR. Furthermore, we evaluate the proposed SoiCP in another large-scale shopping mall and obtain the similar conclusion, demonstrating the robustness of the system. SoiCP facilitates the deployment of seamless outdoor–indoor positioning and can be used for many LBSs with large coverage.

In the future work, we plan to reduce the power consumption of the system in both the terminal and server sides. In the current prototype system, all the sensor samples are uploaded to a cloud server for data processing and evaluating our proposed algorithms. We find that the power consumption of sending these data from both terminal side and working station side is high, and we need large storage in the cloud. We can qualify the power consumption of the system and move parts of processing to a local edge server to save the power consumption similar to our previous work [**?**]. For example, we plan to move the algorithm of PDR in a local edge server (e.g., the working station) and hence only the data of user traces will be sent to the cloud server for trace matching.

ACKNOWLEDGMENT

The authors would like to thank all the volunteers who participated in the experiments for crowdsourcing data collection.

REFERENCES

[1] X. Tian, R. Shen, D. Liu, Y. Wen, and X. Wang, "Performance analysis of RSS fingerprinting based indoor localization," *IEEE Trans. Mobile Comput.*, vol. 16, no. 10, pp. 2847–2861, Oct. 2017.
[2] W. Xue, K. Yu, X. Hua, Q. Li, W. Qiu, and B. Zhou, "APs' virtual positions-based reference point clustering and physical distance-based weighting for indoor Wi-Fi positioning," *IEEE Internet Things J.*, vol. 5, no. 4, pp. 3031–3042, Aug. 2018.
[3] D. Liang, Z. Zhang, and M. Peng, "Access point reselection and adaptive cluster splitting-based indoor localization in wireless local area networks," *IEEE Internet Things J.*, vol. 2, no. 6, pp. 573–585, Dec. 2015.
[4] Z. Li, T. Braun, and D. C. Dimitrova, "A time-based passive source localization system for narrow-band signal," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2015, pp. 4599–4605.
[5] Z. Yang, C. Wu, and Y. Liu, "Locating in fingerprint space: Wireless indoor localization with little human intervention," in *Proc. Mobicom*, 2012, pp. 269–280.
[6] A. Rai, K. K. Chintalapudi, V. N. Padmanabhan, and R. Sen, "Zee: Zero-effort crowdsourcing for indoor localization," in *Proc. Mobicom*, 2012, pp. 293–304.
[7] H. Wang, S. Sen, A. Elgohary, M. Farid, M. Youssef, and R. R. Choudhury, "No need to war-drive: Unsupervised indoor localization," in *Proc. MobiSys*, 2012, pp. 197–210.
[8] B. Zhou, Q. Li, Q. Mao, W. Tu, X. Zhang, and L. Chen, "ALIMC: Activity landmark-based indoor mapping via crowdsourcing," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 5, pp. 2774–2785, Oct. 2015.

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

[10] C. Luo, H. Hong, and M. C. Chan, "PiLoc: A self-calibrating participatory indoor localization system," in *Proc. Int. Symp. Inf. Process. Sensor Netw. (Ipsn)*, 2014, pp. 143–154.

[11] G. Shen, Z. Chen, P. Zhang, T. Moscibroda, and Y. Zhang, "Walkie–Markie: Indoor pathway mapping made easy," in *Proc. NSDI*, 2013, pp. 85–98.

[12] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. 2nd Int. Conf. Knowl. Disc. Data Min.*, 1996, pp. 226–231.

[13] Z. Li, D. B. Acuña, Z. Zhao, J. L. Carrera, and T. Braun, "Fine-grained indoor tracking by fusing inertial sensor and physical layer information in WLANs," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2016, pp. 1–7.

[14] F. Li, C. Zhao, G. Ding, J. Gong, C. Liu, and F. Zhao, "A reliable and accurate indoor localization method using phone inertial sensors," in *Proc. ACM Conf. Ubiquitous Comput.*, 2012, pp. 421–430.

[15] M. H. Afzal, V. Renaudin, and G. Lachapelle, "Assessment of indoor magnetic field anomalies using multiple magnetometers," in *Proc. Int. Tech. Meeting Satellite Division Inst. Navig.*, 2010, pp. 525–533.

[16] S. Madgwick, A. Harrison, and R. Vaidyanathan, "Estimation of IMU and MARG orientation using a gradient descent algorithm," in *Proc. IEEE Int. Conf. Rehabil. Robot.*, 2011, pp. 1–7.

[17] X. Li, J. Wang, and C. Liu, "Heading estimation with real-time compensation based on Kalman filter algorithm for an indoor positioning system," *ISPRS Int. J. Geo-Inf.*, vol. 5, no. 6, p. 98, 2016.

[18] F. Hong, Y. Zhang, Z. Zhang, M. Wei, Y. Feng, and Z. Guo, "WaP: Indoor localization and tracking using WiFi-assisted particle filter," in *Proc. IEEE Conf. Local Comput. Netw.*, 2014, pp. 210–217.

[19] P. Zhou, M. Li, and G. Shen, "Use it free: Instantly knowing your phone attitude," in *Proc. 20th Annu. Int. Conf. Mobile Comput. Netw. (MobiCom)*, 2014, pp. 605–616.

[20] N. Yu, C. Xiao, Y. Wu, and R. Feng, "A radio-map automatic construction algorithm based on crowdsourcing," *Sensors*, vol. 16, no. 4, p. 504, 2016.

[21] K. Ramchandran, "An integrated approach to indoor and outdoor localization," M.S. thesis, Air Force Res. Lab., University of California at Berkeley, Berkeley, CA, USA, 2017.

[22] B. Zhou, Q. Li, Q. Mao, and W. Tu, "A robust crowdsourcing-based indoor localization system," *Sensors*, vol. 17, no. 4, p. 864, 2017.

[23] 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.

[24] Y. Li, Z. He, Z. Gao, Y. Zhuang, C. Shi, and N. El-Sheimy, "Toward robust crowdsourcing-based localization: A fingerprinting accuracy indicator enhanced wireless/magnetic/inertial integration approach," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 3585–3600, Apr. 2019.

[25] X. Tong, K. Liu, X. Tian, L. Fu, and X. Wang, "FineLoc: A fine-grained self-calibrating wireless indoor localization system," *IEEE Trans. Mobile Comput.*, to be published.

[26] C. Luo, H. Hong, M. C. Chan, J. Li, X. Zhang, and Z. Ming, "MpiLoc: Self-calibrating multi-floor indoor localization exploiting participatory sensing," *IEEE Trans. Mobile Comput.*, vol. 17, no. 1, pp. 141–154, Jan. 2018.

[27] Y. Li, Z. Gao, Z. He, P. Zhang, R. Chen, and N. El-Sheimy, "Multi-sensor multi-floor 3D localization with robust floor detection," *IEEE Access*, vol. 6, pp. 76689–76699, 2018.

[28] Y. Kim, S. Lee, S. Lee, and H. Cha, "A GPS sensing strategy for accurate and energy-efficient outdoor-to-indoor handover in seamless localization systems," *Mobile Inf. Syst.*, vol. 8, no. 4, pp. 315–332, 2012.

[29] M. Kourogi, N. Sakata, T. Okuma, and T. Kurata, *Indoor/Outdoor Pedestrian Navigation With an Embedded GPS/RFID/Self-Contained Sensor System*. Heidelberg, Germany: Springer, 2006.

[30] W. W.-L. Li, R. A. Iltis, and M. Z. Win, "A smartphone localization algorithm using RSSI and inertial sensor measurement fusion," in *Proc. Glob. Commun. Conf.*, 2014, pp. 3335–3340.

[31] J. Carrera, Z. Li, Z. Zhao, T. Braun, and A. Neto, "A real-time indoor tracking system in smartphones," in *Proc. ACM Int. Conf. Model. Anal. Simulat. Wireless Mobile Syst.*, 2016, pp. 292–301.

[32] 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, 2013.

[33] M. Weyn, M. Klepal, and Widyawan, *Adaptive Motion Model for a Smart Phone Based Opportunistic Localization System* (LNCS 5801). Heidelberg, Germany: Springer, 2009, pp. 50–65.

[34] R. Hansen, R. Wind, C. S. Jensen, and B. Thomsen, "Seamless indoor/outdoor positioning handover for location-based services in streamspin," in *Proc. 10th Int. Conf. Mobile Data Manag. Syst. Services Middleware*, 2009, pp. 267–272.

[35] L. M. S. Morillo, J. A. O. Ramírez, J. A. A. García, and L. Gonzalez-Abril, "Outdoor exit detection using combined techniques to increase GPS efficiency," *Expert Syst. Appl.*, vol. 39, no. 15, pp. 12260–12267, 2012.

[36] M. Hardegger, S. Mazilu, and G. Troster, "Continuous indoor and outdoor tracking with a stand-alone wearable system," in *Proc. 12th Workshop Position. Navig. Commun. (WPNC)*, 2015, pp. 1–6.

[37] M. Li, P. Zhou, Y. Zheng, Z. Li, and G. Shen, "IODetector: A generic service for indoor/outdoor detection," *ACM Trans. Sensor Netw.*, vol. 11, no. 2, pp. 1–29, Dec. 2014.

[38] V. Radu, P. Katsikouli, R. Sarkar, and M. K. Marina, "Poster: Am I indoor or outdoor?" in *Proc. 20th Annu. Int. Conf. Mobile Comput. Netw. (MobiCom)*, 2014, pp. 401–404.

[39] Z. Li, X. Zhao, and H. Liang, "Automatic construction of radio maps by crowdsourcing PDR traces for indoor positioning," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2018, pp. 1–6.

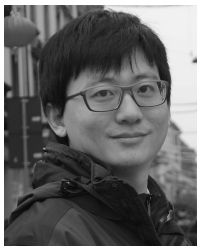
[40] D. Sathyamoorthy, S. Shafii, Z. M. Amin, A. Jusoh, and S. Z. Ali, "Evaluation of the accuracy of global positioning system (GPS) speed measurement via GPS simulation," vol. 8, no. 2, pp. 121–128, 2015.

[41] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 174–188, Feb. 2002.

[42] Z. Li and T. Braun, "Passively track WiFi users with an enhanced particle filter using power-based ranging," *IEEE Trans. Wireless Commun.*, vol. 16, no. 11, pp. 7305–7318, Nov. 2017.

[43] A. J. Haug, *Bayesian Estimation and Tracking: A Practical Guide*. Hoboken, NJ, USA: Wiley, 2012.

[44] J. L. Carrera, Z. Zhao, M. Wenger, and T. Braun, "MEC-based UWB indoor tracking system," in *Proc. 15th Annu. Conf. Wireless On-Demand Netw. Syst. Services*, 2019, pp. 1–8.



Zan Li received the Ph.D. degree from the University of Bern, Bern, Switzerland, in 2016.

He was a Senior Algorithm Engineer with Alibaba Group, Beijing, China, from 2016 to 2017. Since 2017, he has been a Lecturer with the College of Communication Engineering, Jilin University, Changchun, China.

Dr. Li was a recipient of the Best Paper Award from the WMNC 2014, and the Fritz-Kutter Award (the best Ph.D. dissertation in Swiss Universities) in 2016 for his Ph.D. dissertation.

Xiaohui Zhao received the Ph.D. degree in control theory from the Universite de Technologie de Compiegne, Compiègne, France, in 1993.

He is a Full Professor with the College of Communication Engineering, Jilin University, Changchun, China. His current research interests include signal processing and wireless communication, and cognitive radio.

Fengye Hu received the B.S. degree from the Department of Precision Instrument, Xi'an University of Technology, Xi'an, China, in 1996, and the M.S. and Ph.D. degrees in communication and information systems from Jilin University, Changchun, China, in 2000 and 2007, respectively.

He is currently a Full Professor with the College of Communication Engineering, Jilin University. He serves as a Visiting Scholar of Electrical and Electronic Engineering with Nanyang Technological University, Singapore. His current research interests include wireless body area networks, MIMO-OFDM, cognitive radio, and space time communication.

Zhongliang Zhao received the Ph.D. degree from the University of Bern, Bern, Switzerland, in 2014.

He was a Senior Researcher with the University of Bern. He has been active as a multiple work package leader in the EU FP7 Project Mobile Cloud Networking, a co-PI of the Sino-Swiss Science and Technology Cooperation Project M3WSN, and the Technical Coordinator of the Swiss National Science Foundation Project SwissSenseSynergy.

José Luis Carrera Villacrés received the B.S.E. degree from National Polytechnic School, Quito, Ecuador, and the M.Sc. degree in computer sciences from the Swiss Joint Master of Science in Computer Science Program, University of Neuchâtel, Neuchâtel, Switzerland, University of Fribourg, Fribourg, Switzerland, and University of Bern, Bern, Switzerland, in 2015. He is currently pursuing the Ph.D. degree with the Institute of Computer Sciences, University of Bern.

His current research interests include artificial intelligence, machine learning, indoor localization, and distributed systems.

Torsten Braun received the Ph.D. degree from the University of Karlsruhe, Karlsruhe, Germany, in 1993.

Since 1998, he has been a Full Professor with the Institute of Computer Science, University of Bern, Bern, Switzerland. He has been the Vice President of the Swiss Research and Education Network Provider Foundation, Zürich, Switzerland, since 2011.

Dr. Braun was a recipient of the Best Paper Award from the LCN 2001, the WWIC 2007, the EE-LSDS 2013, the WMNC 2014, the ARMSCC 2014 Workshop, and the GI-KuVS Communications Software Award in 2009.