Collaborative SLAM based on Wifi Fingerprint Similarity and Motion Information

Ran Liu, Sumudu Hasala Marakkalage, Madhushanka Padmal, Thiruketheeswaran Shaganan, Chau Yuen, Yong Liang Guan, and U-Xuan Tan

Abstract—Simultaneous localization and mapping (SLAM) has been extensively researched in past years particularly with regard to range-based or visual-based sensors. Instead of deploying dedicated devices that use visual features, it is more pragmatic to exploit the radio features to achieve this task, due to their ubiquitous nature and the widespread deployment of Wi-Fi wireless network. This paper presents a novel approach for collaborative simultaneous localization and radio fingerprint mapping (C-SLAM-RF) in large unknown indoor environments. The proposed system uses received signal strengths (RSS) from Wi-Fi access points (AP) in the existing infrastructure and pedestrian dead reckoning (PDR) from a smart phone, without a prior knowledge about map or distribution of AP in the environment. We claim a loop closure based on the similarity of the two radio fingerprints. To further improve the performance, we incorporate the turning motion and assign a small uncertainty value to a loop closure if a matched turning is identified. The experiment was done in an area of 130 meters by 70 meters and the results show that our proposed system is capable of estimating the tracks of four users with an accuracy of 0.6 meters with Tango-based PDR and 4.76 meters with a step counter-based PDR.

I. INTRODUCTION

With growing applications of the Internet of Things (IoT), recent research shows an increasing interest in indoor positioning due to the rapid demand of location-based services, such as indoor guidance and asset tracking [1]–[4]. To perform indoor positioning, the knowledge of the existing infrastructure must be provided in advance (for example a map of the environment or locations of the beacons). In the scenario of emergency response in disaster areas or large scale environments, such kind of knowledge is not available or difficult to obtain beforehand, which makes the indoor positioning challenging. Therefore, recent researchers are focusing on developing efficient methods and technologies to simultaneously localize mobile devices (robots and smartphones) and generate a map of the environment [5]–[8]. The underlying problem is well known by the term, Simultaneous Localization and Mapping

R. Liu, S. H. Marakkalage, C. Yuen, and U-X. Tan are with the Engineering Product Development Pillar, Singapore University of Technology and Design, 8 Somapah Rd, Singapore, 487372. R. Liu is also with the School of Information Engineering, Southwest University of Science and Technology, 59 Qinglong Road, Mianyang, China, 621010. {ran_liu, yuenchau, uxuan_tan}@sutd.edu.sg.

M. Padmal and T. Shaganan are with the Department of Electronic and Telecommunication Engineering, University of Moratuwa, Sri Lanka, 10400.

Y. L. Guan is with the School of Electrical and Electronic Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore, 639798 eylguan@ntu.edu.sg.

(SLAM). Extensive researches have been done with regard to visual-based [9] or range-based sensors [10].

Loop closure detection is elementary to any SLAM system. It denotes a situation that the mobile device has entered a previously-visited location, which permits to correct the accumulated odometry error. In order to perform loop closure detection in SLAM, dedicated devices (i.e., laser range finders or cameras) are required to measure the similarity of observations by scan matching [11] or feature matching [9], which are usually computationally expensive. However, growing popularity of Wi-Fi wireless networks provide a new opportunity to detect loop closure and perform SLAM in a different way.

Most existing buildings with Wi-Fi network deployed can be exploited for localization and mapping with low hardware requirements and computational cost (for example with the ubiquitous IoT devices like normal smartphones) due to their ubiquitous nature of in-built sensing capabilities [1], [12]. The current signal-strength-based SLAM requires an analytical model to feature the radio signal distribution over distance [6], [13]. However, it is not practical to build such a model due to many multiple path issues in uncontrolled environments. On the other hand, radio fingerprinting [2], [14], [15], represents a location with a collection of radio signals from Wi-Fi access points, which is considered to be more robust against the signal distortions. Therefore, we adopt this technique to simultaneously determine the location of a user and create a radio map of the environment.

In addition to the Wi-Fi network, a typical indoor environment consists of many landmarks, such as turnings, elevators, rooms, and doors, which can be also considered as features for the positioning of a device. These landmarks can be recognized through inertial sensors, which are available in most commercial off-the-shelf smartphones [16], [17]. In contrast to large location uncertainty of radio fingerprints due to the distortion of signals, such kind of landmarks can better confine the location of a device and enhance the positioning accuracy of fingerprinting-based approaches [15], [18].

In opposite to the feature map or occupancy map built by laser range finders or visual cameras, our goal is to build a map (in particular a radio map) with radio fingerprint as feature, and use that for the positioning. To ensure a good positioning accuracy in large scale environment, a fine-grained radio map is required [1], [2] and it will be time consuming to create such a map with a single user. Therefore, a low cost method (e.g., acquire fingerprints via crowdsensing by multiple users) to create the radio map is a necessity.

This paper presents a system that fuses the pedestrian dead

reckoning from a smartphone, and received signal strength (RSS) measurements from surrounding Wi-Fi access points (AP), to estimate the trajectory of multiple users and map the radio signals in unknown environment via a collaborative fashion, using graph SLAM technique. To further improve the accuracy, we incorporate the turning features and reduce the uncertainty of loops inferred based on radio fingerprints similarity. The proposed approach requires neither the map of the environment nor the locations of the access points. We tested the system under two different dead reckoning systems, one is based on Tango that has a high motion tracking accuracy through vision-based odometry, and the second one is based on step counter using on-board inertial sensors that has a poor motion tracking accuracy. By leveraging on in-built sensing capabilities from smart phones and crowdsensing nature, our system can generate a radio fingerprint map in a large indoor environment at low cost as compared with traditional site surveying methods.

We summarize the contributions of this paper as:

- We present a solution that incorporates Wi-Fi fingerprint and dead reckoning information for crowdsensing SLAM in unknown indoor environments;
- We propose an algorithm that automatically learns a model to characterize the uncertainty of a loop based on the degree of similarity using the short term odometry measurement;
- We integrate the turning features to further reduce the uncertainty of radio fingerprint-based loop closures and improve the overall accuracy;
- We throughly evaluate our approach in one building at our campus with an area of approx. 9000 square meters with two different pedestrian dead reckoning systems.

We organize the rest of this paper as follows. The related work is discussed in Section II. Section III formulates the problem and explains the detail of the proposed system. Section IV presents the experimental results. Conclusions with possible directions of future work are made in Section V.

II. RELATED WORK

Over the past decades, indoor positioning shows a growing popularity due to the increasing demand of location-aware applications [1], [2]. A large number of researches have been performed regarding indoor positioning given a reference of the infrastructure (i.e., map of the environment or distributions of beacons). Obtaining and maintaining such kind of information is challenging, particularly in large scale environments [19] or emergency response for example search and rescue in disaster scenes [20], [21]. A solution to this problem is SLAM (Simultaneous Localization and Mapping), which has been investigated extensively in robotics community. In this section, we present a summary of the related work in SLAM, using different kinds of techniques. Throughout the years, many techniques and algorithms have been proposed, mainly including filtering-based solutions (for example the Kalman filter [5] and the particle filter [22]) and graph-based solutions [7], [23].

Depending on the types of sensors used, one can classify the SLAM into laser-based SLAM, visual-SLAM, magnetic-SLAM, WifiSLAM, and FootSLAM. Laser-based SLAM uses laser-range finders to create a structural map of an environment. The detection of loop closure is achieved by scan matching. Visual-SLAM methods, utilize cameras like Kinect or Tango [24] to construct a 3D model of the indoor scene. Bundle adjustment [25] is another popular technique for SLAM that uses visual images and has been used in commercialized SLAM systems such as Google's Project Tango [26]. Magnetic-SLAM systems, exploit digital magnetic compass for localization and mapping of a device [27]. The loop closure is inferred by examining spatial similarity of a sequence of magnetic measurements. For example, authors in [18] correlated motion patterns with the magnetic field to address the SLAM problem. A unique magnetic fingerprint may not be guaranteed due to the distortion of the environment, which makes this solution challenging for real applications.

WifiSLAM [6], [13] techniques use the radio signal and motion data of the device for localization and signal strength mapping in unknown environments. With SLAM technique, the hassle of site surveying can be avoided, and radio map can be created and updated conveniently whenever needed. For example, authors in [6] solved the WifiSLAM problem by mapping the high-dimensional signal strength into a two-dimensional latent space with a Gaussian process. Authors in [13] proposed a generalized and effective algorithm to solve the WifiSLAM using GraphSLAM algorithm. Both approaches assume the signal strength at two close locations are similar and the measurement likelihood can be modeled as a Gaussian process. In contrast, our approach does not require any model to describe the signal strength distribution, and the closeness of locations is determined by the similarity of the radio fingerprints.

FootSLAM [28] uses inertial-based measurements to determine the underlying building structure. No ranging or visual measurement were required; the only features used are the probability distributions of human motions at different locations. Several extensions, for example ActionSlam [29], incorporate location-orientated actions (for example entering elevators or door opening) as features to compensate for the IMU drifting error. Additional information [30] can be further incorporated into FootSLAM, for example a prior map or signal strength from a Wi-Fi access point. Authors in [31] proposed SenseWit that utilizes inertial measurements to generate a floorplan by identifying featured locations (turning, water dispenser, and door) in indoor space.

When the indoor environment becomes huge, generating the radio map with single mobile device becomes time consuming. The power of crowd comes into play in this scenario. Mobile crowdsensing is a popular computing paradigm, which enables ubiquitous devices to collect sensing data at large scales [32]–[34]. This technique can be utilized to unleash the potential of mobile phones of people who move inside the indoor environment [35]. Prior research have harnessed the power of crowdsensing to reconstruct indoor floor plans by combining user mobility traces, images of landmarks, and Wi-Fi fingerprints [36], [37]. Localization by combination of 6-DOF

gyro-odometry and Wi-Fi localization has been done in [38], using multiple robots. Authors in [39] proposed an approach to utilize pairwise distance measures between users to reduce the positioning error in fingerprinting-based approaches.

Our system combines crowdsensed RSS from Wi-Fi APs and dead reckoning information from a phone to localize a device and build a radio map of the environment. To improve the accuracy, we additionally incorporate turning features extracted from users' trajectories into our system. Section III explains the details of the system implementation.

III. COLLABORATIVE SLAM BASED ON POSE GRAPH OPTIMIZATION

We present a novel approach that incorporates radio fingerprint measurement and motion information for collaborative SLAM in an unknown environment. The approach presented here does not require any prior knowledge about the map or the distribution of the access points nor does it need a labor-intensive phase to collect the measurements in the existing infrastructure. Our approach features a cost-effective alternative to estimate the trajectory of multiple users in unknown environments. A radio map is created simultaneously, which can be used as reference to localize other users afterwards.

In our proposed collaborative framework, user walks in the environment and collects the radio measurements. Our approach merges the tracks from different users, performs loop closure detection, and optimizes the graph to generate a consistent radio fingerprint map. The data collected will be shared among all users through the server and each user will contribute to certain part of the map. The collaborative approach will accelerate the conventional way of map building. With our collaborative approach, the mapping of a building will become easier, since all users will participate in the map creation, which eliminates the expensive on-site survey phase in the conventional way of fingerprint map generation.

The goal is to estimate the entire trajectories from observations (i.e., Wi-Fi observations and motion measurements) taken from different users at different times without a prior knowledge about the environment. The problem addressed here is known as SLAM, which has been well studied in the field of robotics [10], [22], [40], [41]. Among those, the graph-based approach, which formulates the problem as maximum likelihood estimation using pose graphs, is regarded as one of the most effective way to solve SLAM problem. Based on the raw sensor measurements, graph-based SLAM [22] creates a graph, where nodes denote the poses of the users and edges decode the constraints between two nodes. The problem turns into graph optimization, which determines the best configuration of the poses by considering all constraints in the graph.

Loop closure is important for any SLAM system and is considered as one of the main challenges in implementing a SLAM system in large-scale environment. It represents a situation that users have revisited a previously observed location. Since the odometry will inevitably drift for long term run, loop closure allows to correct the accumulated odometric errors and create a consistent map of the scene. The loop closure problem has been researched extensively using visual-based or ranging-based sensors [9], [42], which are usually costly and computationally expensive. Instead of using the dedicated devices to perform loop closure detection, we focus on the radio fingerprints, which are available in existing indoor infrastructures and can be easily retrieved from every smartphone.

Formally, let $\mathbf{x}_{1:T}^k = \{\mathbf{x}_1^k, ..., \mathbf{x}_T^k\}$ be the path of user k we would like to estimate up to time T, where $\mathbf{x}_t^k = (x_t^k, y_t^k, \theta_t^k)$ represents the global 2D location and heading of the user at time t. Let $\mathbf{z}_{m,i}^{n,j}$ and $\Sigma_{m,i}^{n,j}$ represent the mean and covariance of a measurement (i.e., constraint or edge) between node $\mathbf{x}_i^m$ and $\mathbf{x}_j^n$. We use $\mathbb{C}$ to denote the set containing all pairs of constraints. $\hat{\mathbf{z}}_{m,i}^{n,j}(\mathbf{x}_i^m, \mathbf{x}_j^n)$ is the prediction of a measurement based on the current configuration of node $\mathbf{x}_i^m$ and $\mathbf{x}_j^n$. The graph-based SLAM aims to find the best configuration $\mathbf{x}^*$ to meet the following criteria:

$$
\begin{aligned}
\mathbf{x}^* = \arg\min_{\mathbf{x}} \sum_{(\mathbf{x}_i^m, \mathbf{x}_j^n) \in \mathbb{C}} &(\mathbf{z}_{m,i}^{n,j} - \hat{\mathbf{z}}_{m,i}^{n,j}(\mathbf{x}_i^m, \mathbf{x}_j^n))^{\mathsf{T}} \Sigma_{m,i}^{n,j}{}^{-1} \\
&\times (\mathbf{z}_{m,i}^{n,j} - \hat{\mathbf{z}}_{m,i}^{n,j}(\mathbf{x}_i^m, \mathbf{x}_j^n))
\end{aligned}
\tag{1}
$$

In particular for graph-based SLAM, $\mathbf{z}_{m,i}^{n,j}$ is known as edge or constraint which represents a rigid-body transformation between node $\mathbf{x}_i^m$ and $\mathbf{x}_j^n$. The transformation is a 3×1 vector which encodes the 2D translation (i.e., $x_{m,i}^{n,j}$ and $y_{m,i}^{n,j}$) and the rotation $\theta_{m,i}^{n,j}$. The constraint can be either sequential odometry measurement (i.e., odometry-based constraint) or loop closure (observation-based constraint), which is determined by aligning the sensor observations at two non-consecutive poses. Since observations are usually erroneous, all constraints are additionally parameterized with a certain degree of uncertainty (i.e., $\Sigma_{m,i}^{n,j}$). For laser range finders, the transformation $\mathbf{z}_{m,i}^{n,j}$ can be determined by matching two scans [11] using model-based registration, for example iterative closest point (ICP). Given a signal strength measurement from an AP, it is straightforward to know if an area has been visited by a user, since each reported RSS value is associated with a unique MAC address. However, estimating the precise transformation $\mathbf{z}_{m,i}^{n,j}$ between two observations turns out to be tricky, since radio signal neither reports distance nor bearing, and the detection range of an AP can be up to 50 meters, which is usually much larger than the accumulated error of a pedestrian dead reckoning system.

The distance to an access point can be approximated by the signal strength via an analytical signal-to-distance model. This model is used by some researchers to address the SLAM problem [6], [13], [43] and indoor positioning [1], [2]. However, obtaining such a model is usually not practical, as the propagation of signal will be distorted by many environmental factors (for example, multiple path or obstruction from obstacles). Instead of modeling them explicitly, this paper represents the location with radio fingerprint, which consists of the address of detected device and the measured signal strength. These fingerprints are location-dependent and are assumed to be unique to describe a location in an infrastructure. The closeness of two locations can be determined by comparing
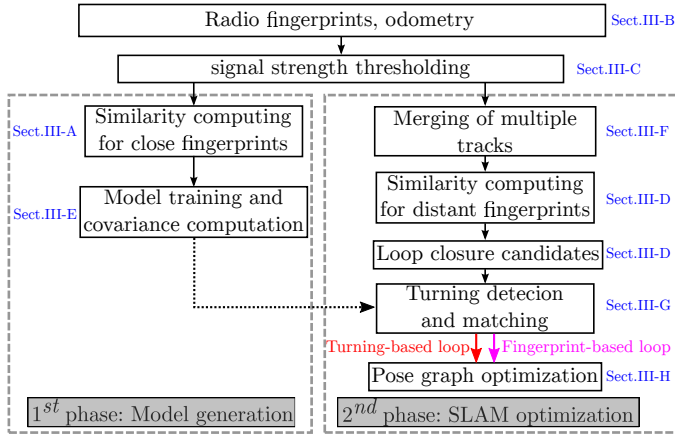
Fig. 1. Illustration of our collaborative simultaneous localization and radio fingerprint mapping (C-SLAM-RF) system. The proposed system automatically creates a radio fingerprint map of an environment using the radio and odometry measurements from a group of users. The fingerprint-based and turning-based loops are identified and incorporated into a graph-based SLAM algorithm for optimization.

the degree of similarity of the fingerprints.

We claim a loop closure if the similarity between two radio measurements at $\mathbf{x}_i^m$ and $\mathbf{x}_j^n$ reaches a threshold ϑ_s. We then infer that their locations are the same and add a constraint $\mathbf{z}_{m,i}^{n,j}$, with all elements zero, to the graph. Actually, the two locations are unlikely to be exactly the same, which will produce a small amount of error to the loop closure. The error can be compensated by associating a covariance $\Sigma_{m,i}^{n,j}$ to the constraint. A choice of this can be a diagonal matrix by setting small values on the main diagonal. Our solution is a careful examination of the uncertainty of a loop based on the degree of similarity in a training phase. Based on training data, we automatically learn a nonparametric model to characterize the degree of similarity conditioned on the distance of two locations.

The uncertainty of the loop closure inferred from radio fingerprints is very high, we therefore exploit the turning features to improve the accuracy. We identify the turnings using motion information, match the turnings, and assign the loop closure with a small covariance if a match is found. Figure 1 gives an overview of the system. We will describe the details of each component in our proposed solutions in the following subsections. We summarize the notations of the symbols used in this paper in Table I.

A. Radio Fingerprints and the Similarity

Radio fingerprinting represents location with radio signals from radio-based sensors, for example Wi-Fi APs, Bluetooth beacons, and RFID tags. These fingerprints are robust against location-dependent distortions as compared to the model-based approaches, since the propagation of the radio signal in an environment is hard to predict due to the blockage of obstacles and multipath fading issue. This is quite similar to appearance-based approach, where the scene is represented by a number of visual features. Extracting visual features involves a large amount of computation, while this process can be ignored for

TABLE I
SUMMARY OF IMPORTANT VARIABLES USED IN THIS PAPER

Symbol	Meaning
$\mathbf{x}_t^k$	the pose of user k at time t, i.e., $\mathbf{x}_t^k = (x_t^k, y_t^k, \theta_t^k)$
$\mathbf{x}_{1:T}^k$	the trajectories of user k up to time T
$\mathbf{z}_{m,i}^{n,j}$	the transformation (i.e., translation and rotation) between node $\mathbf{x}_i^m$ and $\mathbf{x}_j^n$
$\Sigma_{m,i}^{n,j}$	the covariance of the transformation $\mathbf{z}_{m,i}^{n,j}$ between node $\mathbf{x}_i^m$ and $\mathbf{x}_j^n$
$\mathbf{F}_t^k$	the radio fingerprint and pose of user k recorded at time t, i.e., $\mathbf{F}_t^k = (\mathbf{f}_t^k, \mathbf{x}_t^k)$
$\mathbf{f}_t^k$	the radio fingerprint at pose $\mathbf{x}_t^k$, which consists of the RSS from L APs: $\mathbf{f}_t^k = \{f_{t,1}^k, ..., f_{t,L}^k\}$
$\mathbf{s}_{m,i}^{n,j}$	the cosine similarity between $\mathbf{f}_i^m$ and $\mathbf{f}_j^n$
$d(\mathbf{x}_i^m, \mathbf{x}_j^n)$	the relative distance between pose $\mathbf{x}_i^m$ and $\mathbf{x}_j^n$ where fingerprints $\mathbf{F}_i^m$ and $\mathbf{F}_j^n$ are recorded
$\theta(\mathbf{x}_i^m, \mathbf{x}_j^n)$	the relative orientation between pose $\mathbf{x}_i^m$ and $\mathbf{x}_j^n$ where fingerprints $\mathbf{F}_i^m$ and $\mathbf{F}_j^n$ are recorded
ϑ_s	the threshold used to claim a loop closure based on the cosine similarity $\mathbf{s}_{m,i}^{n,j}$
ϑ_r	the threshold used to filter out the low RSS measurement
r	the binning size to train the model (i.e., the uncertainty given a measured similarity)
$\{s_k, d_k\}_{k=1}^K$	the K training samples, where s_k denotes the similarity and d_k denotes the distance between a fingerprint pair
$\mathbf{b}(s, r)$	a set of samples that sits in an interval r around a similarity s
$var(d\|s)$	variance of the samples given a similarity s
$c(\mathbf{b}(s, r))$	the number of samples that sits in an interval r around a similarity s
a_t^k	the acceleration measurement of user k at time t
s	the step length of the step counter-based pedestrian dead reckoning
$R_t^k(\tau)$	the normalized auto-correlation of the accelerometer data for lag τ at the tth sample of user k
$\mu_t^k(\tau)$	mean of the sequential acceleration measurements for a lag τ at time t
$\sigma_t^k(\tau)$	standard deviation of the sequential acceleration measurements for a lag τ at time t
c_t^k	the step counter at timestamp t of user k
w	the window size to segment the inertial measurements for turning detection and matching
$\mathbb{C}_t^k$	the track segmentation of user k at time t
$\overline{\mathbb{C}}_t^k$	the relative positions of $\mathbb{C}_t^k$ with respect to $\mathbf{x}_t^k$
$\theta_t^{k-}, \theta_t^{k+}$	the mean orientation of poses with timestamps smaller or larger than t in segmentation $\mathbb{C}_t^k$
ϑ_f	the threshold used to claim a valid turning match

the radio fingerprint, since the AP can be regarded as the unique feature for the positioning.

We represent a fingerprint of user k at time t as a pair $\mathbf{F}_t^k = (\mathbf{f}_t^k, \mathbf{x}_t^k)$. $\mathbf{x}_t^k = (x_t^k, y_t^k, \theta_t^k)$ denotes the odometry at time t when user k traverses the environment. $\mathbf{f}_t^k$ represents the radio measurement at location $\mathbf{x}_t^k$, which consists of the RSS values from L access points: $\mathbf{f}_t^k = \{f_{t,1}^k, ..., f_{t,L}^k\}$. Let L_i^m and L_j^n denote the number of detections in $\mathbf{f}_i^m$ and $\mathbf{f}_j^n$, respectively. $L_{m,i}^{n,j} = \left|\mathbf{f}_i^m \cap \mathbf{f}_j^n\right|$ is used to represent the common APs in both $\mathbf{f}_i^m$ and $\mathbf{f}_j^n$. The similarity function $\mathrm{sim}(\mathbf{F}_i^m, \mathbf{F}_j^n)$ yields a positive value, representing the similarity between two radio measurements, namely $\mathbf{f}_i^m$ and $\mathbf{f}_j^n$. We apply the cosine similarity which has been extensively used in the literature [44] [45].

$$\mathbf{s}_{m,i}^{n,j} = \mathrm{sim}(\mathbf{F}_i^m, \mathbf{F}_j^n) = \frac{\sum_{l=1}^{L_{m,i}^{n,j}} f_{i,l}^m f_{j,l}^n}{\sqrt{\sum_{l=1}^{L_i^m} \left(f_{i,l}^m\right)^2}\sqrt{\sum_{l=1}^{L_j^n} \left(f_{j,l}^n\right)^2}} \quad (2)$$

We refer the readers to [2] [46] for a comparison between different similarity measures.

B. Pedestrian Dead Reckoning

The spatial relationship between sequential poses in Equation 1 can be determined by the odometry measurements, which is known as odometry-based constraint. Nowadays, smartphones are equipped with various types of sensors, including IMU sensor, camera, light sensor, and proximity sensor. This enables one to implement a variety of pedestrian dead reckoning systems using different techniques. We evaluated our system under two pedestrian dead reckoning systems (PDR): Tango-based PDR using visual-inertial odometry (VIO) and step counter-based PDR using accelerometers and compass. The goal is to compare the approach under various tracking systems with different tracking accuracies.

Tango is developed by Google that uses visual-inertial odometry, to estimate the location of a device without GPS or any external referencing. It uses visual features with a combination of inertial measurements from accelerometer and gyroscope to track the movement of a device in 3D space. Lenovo Phab 2 Pro and Asus Zenfone AR are two examples of commercially available Tango phones.

Alternatively, the IMU sensor embedded inside a phone can be used to implement a step counter-based dead reckoning. A typical IMU system is comprised of accelerometer, gyroscope, and magnetometer for motion or orientation sensing. Following Zee [47], we implemented the step counting based on auto-correlation. Given the acceleration measurement a_t^k of user k at time t, the step counting is achieved by examining the periodic step patterns through normalized auto-correlation $R_t^k(\tau)$ for a lag τ:

$$R_t^k(\tau) = \frac{\sum_{l=0}^{l=\tau-1}\left[\begin{array}{c}(a_{t+l}^k - \mu_t^k(\tau)) \\ (a_{t+l+\tau}^k - \mu_{t+\tau}^k(\tau))\end{array}\right]}{\tau \sigma_t^k(\tau)\sigma_{t+\tau}^k(\tau)}, \quad (3)$$

where $\mu_t^k(\tau)$ and $\sigma_t^k(\tau)$ are mean and standard deviation of the sequential acceleration measurements $\{a_t^k, a_{t+1}^k, ..., a_{t+\tau-1}^k\}$.

The algorithm first identifies an optimal τ_{opt} to maximize $R_t^k(\tau)$. Then τ_{opt} is used as a replacement of τ to count further steps. Similar to Zee [47], the sampling rate of the IMU is 50Hz, we therefore set the initial searching window τ to $[40, 100]$.

During the walking of a person, we assume the phone is always held in front of him. Therefore, we use the magnetometer reading to approximate the orientation of the user. Let c_t^k and θ_t^k be the step counting and the orientation of user k at time t respectively, the position (i.e., x_t^k and y_t^k) of user k is determined by:

$$x_t^k = x_{t-1}^k + s \cdot (c_t^k - c_{t-1}^k) \cdot \cos(\theta_{t-1}^k) \quad (4)$$
$$y_t^k = y_{t-1}^k + s \cdot (c_t^k - c_{t-1}^k) \cdot \sin(\theta_{t-1}^k), \quad (5)$$

where s is the step length, which is assumed to be fixed throughout the experiments. The estimation of the pose in odometry frame works in recursive fashion. The initial values of x and y are set to zero. The initial headings of different PDR systems are treated differently. For Tango-based PDR, the initial heading of a user is set to zero. For step counter-based PDR, the heading is determined by the in-built magnetometer based on geographic cardinal directions. The magnetic materials in the building might affect the accuracy of the orientation estimation. In the future, we would like to compensate for the orientation by incorporating the gyroscope readings [48].

The odometry-based edge is determined based on the relative translation and rotation between the sequential odometry measurements. For a user k, the rigid-body transformation (Δx_t^k, Δy_t^k and $\Delta \theta_t^k$) between the pose at time $t-1$ and time t can be computed as:

$$\begin{bmatrix}\Delta x_t^k \\ \Delta y_t^k \\ \Delta \theta_t^k\end{bmatrix} = \begin{bmatrix}\cos(\theta_{t-1}^k) & -\sin(\theta_{t-1}^k) & 0 \\ \sin(\theta_{t-1}^k) & \cos(\theta_{t-1}^k) & 0 \\ 0 & 0 & 1\end{bmatrix}^\top \begin{bmatrix}x_t^k - x_{t-1}^k \\ y_t^k - y_{t-1}^k \\ \theta_t^k - \theta_{t-1}^k\end{bmatrix} \quad (6)$$

C. RSS Thresholding

The time required to compute the similarity in Equation (2) grows linearly with the number of APs in the two fingerprints. The computational time can be significant in densely AP covered environment, which is the typical case in modern office or commercial buildings. A large amount of computational time can be saved if the size of the measurements can be reduced. Therefore, we propose to filter out the RSS observation with value below a threshold ϑ_r.

Thresholding prunes observations with small RSS values, which represent spurious readings due to multiple propagation issues in indoor environment. In addition, larger RSS values indicate a location close to the access point with more confidence. These measurements are expected to better confine the location of the user. In the experimental section, we show that thresholding technique can provide a better accuracy while consuming less computational time.

D. Finding Loop Closure Candidates

To find the observation-based edge (i.e., constraint) between the non-consecutive poses in Equation 1, we need to perform

(a) Experimental snapshot in our campus

(b) Similarity and the distance variance in two different buildings
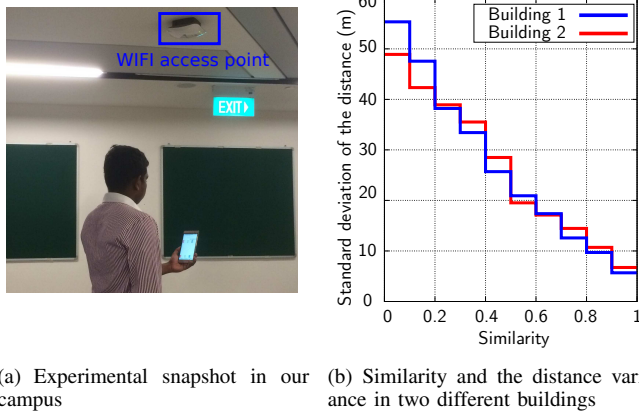
Fig. 2. (a) Snapshot of a person holding a phone and walking in the environment; (b) The trained variance model at two different buildings.
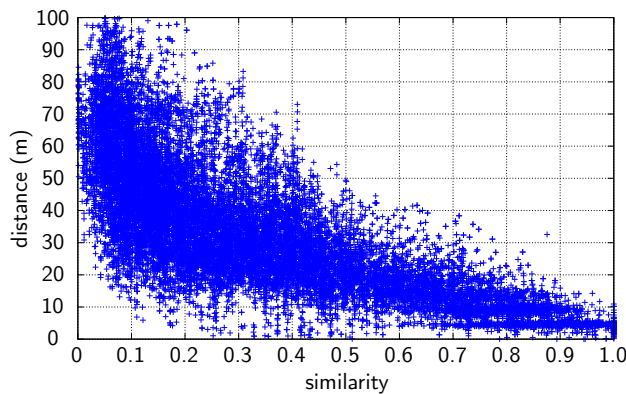

Fig. 3. Scatter plot of similarity versus distance from experimental data in building 1.

loop closure detection. The observation-based edge consists of two different types of edges, namely similarity-based edge and turning-based edge. In our approach, each fingerprint carries the odometry information where the fingerprint is recorded (i.e., x, y, and the orientation θ of a user). We first compute the relative distance $d(\mathbf{x}_i^m, \mathbf{x}_j^n)$ and orientation $\theta(\mathbf{x}_i^m, \mathbf{x}_j^n)$ between the odometric poses of two fingerprints $\mathbf{F}_i^m$ and $\mathbf{F}_j^n$. If these values are smaller than pre-defined thresholds (50 meters and 0.3 radians for distance and orientation respectively), we compute the similarity $\mathbf{s}_{m,i}^{n,j}$ between them. We add a tuple $< \mathbf{x}_i^m, \mathbf{x}_j^n, \mathbf{s}_{m,i}^{n,j} >$ as a candidate of the loop closure if the similarity $\mathbf{s}_{m,i}^{n,j}$ exceeds a threshold ϑ_s, which is one of the few parameters that has to be supplied by the user. The impact of ϑ_s on the performance is not too critical, as shown in our experiments. In most cases, $\vartheta_s = 0.7$ gives good results. We reject the similarity with values smaller than ϑ_s, to avoid false positive loop closures. To improve the accuracy of the system, we further check if this loop is a turning-based loop. We identify a turn by examining the orientation changes and checked the fitness of their respective tracks by a matching algorithm. If the fitness score is higher than a threshold, we consider this loop as a turning-based loop. The detail of the detection of turning-based loop can be found in Section III-G.

E. Model Training

To optimize Equation 1, an uncertainty estimation of the constraint is necessary for all edges in SLAM graph. For odometry-based edges, the parameter is obtained from the motion model. We now need to derive a model to represent the uncertainty of the observation-based edges. Our solution is to train such model by passing over the observation data (i.e., odometry and radio fingerprints), which is recorded by the smart phone as shown in Figure 1.

Our goal is to generate an uncertainty model to feature the distance variance of two radio fingerprints given their similarity. To build such a model, we need to know the true locations where the fingerprints are recorded or relative distance between the recorded positions, which is not possible without any external reference system. Although the error from odometry accumulates in the long term, it is sufficiently small for a short time of duration. In this work, we assume odometry is accurate enough for the distance traveled less than 100 meters, which is suitable for most inertial tracking platforms [48] [49] [50]. For example, authors in [49] evaluated the visual odometry with wide angle and fisheye cameras, and showed a relative positioning error of less than 1.4% with a distance of 538 meters traveled. Therefore, we compute the degree of similarity for close fingerprint pairs. These values are annotated with the distance between the two locations using the PDR system. As a result, we will get a set of K training samples: $\{s_k, d_k\}_{k=1}^K$, where s_k is the similarity and d_k is the distance of the fingerprint pair. Figure 3 shows the scatter plot of distance versus similarity in one of the buildings. We then train a model which features the variance of distance given a similarity by binning. That is, for a similarity value s, we compute the variance of the samples that sites in the small interval r around s:

$$var(d|s) = \frac{1}{c(\mathbf{b}(s,r))} \sum_{k \in \mathbf{b}(s,r)} d_k{}^2 \tag{7}$$

where $c(\mathbf{b})$ counts the number of samples in interval r. $var(d|s)$ denotes the variance of the distance d given a similarity s. Although binning is a simple way for smoothing, the computation is efficient, since assigning the sample into a bin is straightforward. One example of the variance computed in two different buildings is shown in Figure 2(b). The resulted variance is stored in a look up table, which could be used in the second stage of SLAM, as shown in Figure 1.

F. Merging Tracks at Different Times

To leverage the power of crowdsourcing, we utilize the measurements captured from different users to generate a radio map of the environment. This involves the loop closure detection between different users in order to correct their paths using the power of crowdsourcing. The tracks recorded from different users are based on different reference systems, for example different starting positions. The determination of orientation is different for the two pedestrian dead reckoning systems. The Tango-based PDR estimates the orientation by visual-inertial odometry based on the starting pose and the step counter-based PDR determines the orientation based on the

compass readings and is not relevant to any starting position. Therefore, these trajectories are needed to be merged into the same coordinate system to guarantee a robust loop closure detection between different users.

In this paper, we start the tracking by assuming all users passing by the same place. This is quite reasonable since users may pass through several key landmarks in an environment for example entrance of a building or the elevator. One might argue that in large buildings for example airport, not all users share the common place. It is possible to first build several sub-maps, and then merge them into a large and consistent map [51] [52] [53]. An edge is added to connect the first nodes in different tracks. For the transformation matrix $\mathbf{z}_{m,i}^{n,j}$, we set $x_{m,i}^{n,j}$ and $y_{m,i}^{n,j}$ to zero and the covariance is obtained by checking the covariance table as detailed in Section III-E. Due to the omnidirectional characteristics of antennas, the facing of a user has very little impact on the radio signals, therefore the radio fingerprint does not deliver any orientation information. This is the reason why Section III-E does not model orientation variance with respect to the similarity. Based on two fingerprint observations, we have no knowledge about how accurate is the relative orientation between two poses, we therefore set $\theta_{m,i}^{n,j}$ to zero and give a very large covariance value (i.e., 1000) to the edge, meaning that we are not able to infer the relative angle from two radio fingerprint observations. For users starting from arbitrary locations, we refer to [54] [55] [56] [57] to merge the paths based on the radio measurements and activity landmarks in our future work.

G. Turning Identification and Matching

A typical indoor infrastructure often contains various landmarks, such as turnings, elevators, and staircases. These landmarks are unique in an existing infrastructure and can be used as a good feature for loop closure identification in a SLAM system. In this case, a loop closure can be claimed if two landmarks match each other. Due to the physical constraint of an environment, such kind of loop closure provides lower positioning uncertainty as compared to fingerprinting-based loops. For example, the size of an elevator is usually less than 3 meters, and the turning radius during human walking is smaller than a typical corridor width (i.e., 5 meters), while the positioning error using fingerprinting-based approach is usually larger than 5 meters.

Here we only focus on the turning features in a trajectory, which is regarded as one of the most common indoor landmarks. A slide window is used to produce segmentations of the track. In particular, we define a segmentation $\mathbb{C}_t^k$ of user k at time t as a collection of sequential poses with a window size of w, i.e., $\mathbb{C}_t^k = \{\mathbf{x}_{t'}^k \mid -\frac{w}{2} \leq t' - t \leq \frac{w}{2}\}$. For each loop closure candidate $< \mathbf{x}_i^m, \mathbf{x}_j^n >$, we check if there are turnings at these poses by examining the orientation change in segment $\mathbb{C}_i^m$ and $\mathbb{C}_j^n$. If yes, we try to match the two segmentations $\mathbb{C}_i^m$ and $\mathbb{C}_j^n$ using ICP (Iterative Closest Point) [58]. If the fitness score (i.e., average of squared distances between the correspondence points) between $\mathbb{C}_i^m$ and $\mathbb{C}_j^n$ is smaller than a predefined threshold ϑ_f, we regard loop $< \mathbf{x}_i^m, \mathbf{x}_j^n >$ as the turning-based loop. Otherwise, this loop
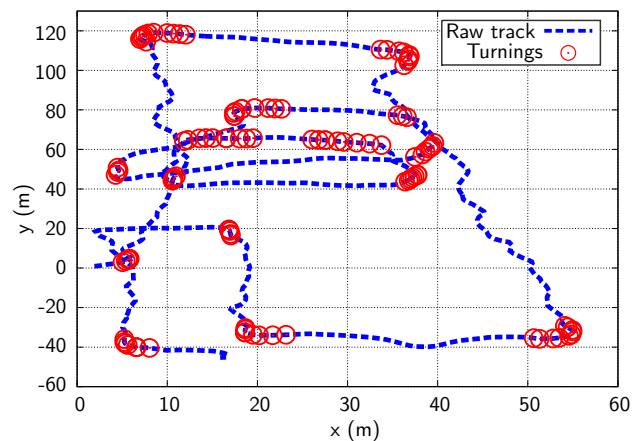

Fig. 4. Example of the track (blue color) and turnings detected (red circles) with a window size of 20 based on the approach described in Section III-G. In total, 111 turnings are identified. These turnings are further examined by turning matching module to find the potential turning-based loops.
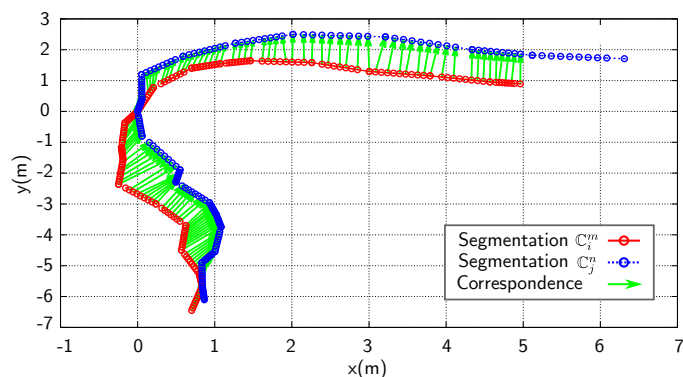

Fig. 5. Example of two track segmentations (i.e., $\mathbb{C}_i^m$ and $\mathbb{C}_j^n$ in blue and red color, respectively), and the correspondence points found using ICP (green color). A fitness score is calculated to determine if this candidate is a turning-based loop.

closure is referred to as fingerprint-based loop, as shown in Figure 1. The transformation $\mathbf{z}_{m,i}^{n,j}$ of both types of loop is set to zero. We treat the covariance matrix $\Sigma_{m,i}^{n,j}$ differently: for fingerprint-based loop, the covariance along x and y can be found at the looking up table computed previously in Section III-E; for turning-based loop, we set the covariance along x and y to 5.0, which is smaller than fingerprint-based loop (i.e., 8.0 in Figure 2(a) with the highest fingerprint similarity). The orientation covariance in $\Sigma_{m,i}^{n,j}$ is set to 1000, meaning that we do not have any knowledge about the orientation of the two poses by the radio observations or the matching of two turnings. The details of turning identification and matching will be described in subsequent parts:

1) Turning Identification: We first segment the compass data and find the potential turnings [16]. For the pose $\mathbf{x}_t^k$, we calculate the mean orientations for the poses with timestamps smaller and larger than t in the segmentation $\mathbb{C}_t^k$, i.e., $\theta_t^{k^-} = \mu\{\theta_{t'}^k \mid t' - t \leq 0\}$ and $\theta_t^{k^+} = \mu\{\theta_{t'}^k \mid t' - t \geq 0\}$. If $|\theta_t^{k^+} - \theta_t^{k^-}|$ is higher than a threshold (for example $\frac{\pi}{3}$ as suggested in [16]), a turning is identified. The window size w here has the impact on the performance of turning detection

and we show its impact on the accuracy in the experimental section. One example of the track and turnings detected are shown in Figure 4. A better approach to improve the accuracy of turning detection can be found in [59].

2) Turning Matching: The ICP aims to find a transformation (translation and rotation) between two point clouds that minimizes the sum of the square distance between the correspondence points. This approach has been extensively used to match 2D laser scans in the field of robotics.

To find the correct transformation using ICP, an appropriate initial transformation has to be provided, otherwise ICP will fall into the local minimum. Rather than using the global raw odometry for a segmentation $\mathbb{C}_t^k$, we use the relative translation between $\mathbf{x}_{t'}^k$ and $\mathbf{x}_t^k$, i.e., $\overline{\mathbb{C}}_t = \{\mathbf{T}_{\mathbf{x}_t^k}^{\mathbf{x}_{t'}^{k'}} \mid -\frac{w}{2} \leq t' - t \leq \frac{w}{2}\}$. Since the sampling rate of our pedestrian tracking system is too low (less than 1.0 HZ), we further interpolate the trajectory to get a large amount of locations for performing ICP. An illustration of the turning matching using ICP is shown in Figure 5.

We finally considers $< \mathbf{x}_i^m, \mathbf{x}_j^n >$ as a valid match (i.e., turning-based loop) if the fitness score is smaller than a threshold ϑ_f. Robust loop closure detection is essential to a SLAM system, as incorrect loop closures will ruin the consistency of trajectory and the map. Other heuristic approach can be applied to further examine the loops and filter out the suspicious ones. Authors in [60], for example, proposed an approach to group the close loop closures and check the temporal and spatial consistency for robust loop closure detection. However, this technique goes beyond the scope of this paper, hence, we add the loop closures without performing futher consistency check.

H. Pose Graph Optimization

The Equation 1, which represents a graph consisting of poses (i.e., nodes) and constraints (i.e., edges), is finally optimized through the pose graph optimization algorithm GraphSLAM. For the implementation, we choose Levenberg-Marquardt in g2o [23], which is freely available and is one of the state-of-the-art SLAM algorithms[1].

IV. Experimental Results

A. Experimental Details

We program two smart phones (Lenovo Phab 2 Pro with Android 6.0.1 and Sony Xperia Z3 with Android 5.0.2) to receive the signal strength from APs and perform pedestrian dead reckoning. In particular, the Lenovo phone uses the Tango for position tracking and the motion tracking data is recorded every five seconds due to its high tracking accuracy. We implement the step counting on a Sony phone and record the step counting and the compass readings every one second. To evaluate the performance of the proposed approach, we conducted experiments on the Level 3 of Building 2 at Singapore University of Technology and Design with a size of 130m×70m (see Fig. 6). This environment is comprised of corridors, concrete walls, and wide open space. We asked a

[1]https://github.com/RainerKuemmerle/g2o
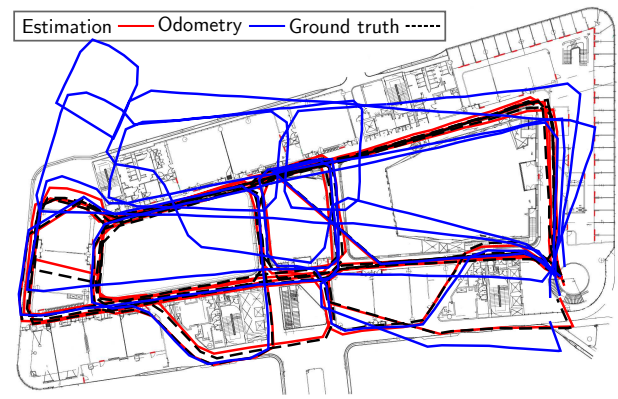
(a) Evaluation of Tango-based pedestrian dead reckoning system
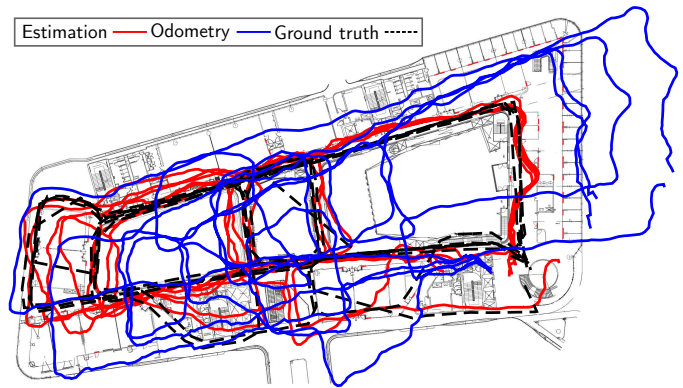
(b) Evaluation of step counter-based pedestrian dead reckoning system

Fig. 6. Ground truth, odometry, and estimated path with our proposed approach under two different pedestrian dead reckoning (PDR) systems, namely Tango-based PDR and step counter-based PDR.
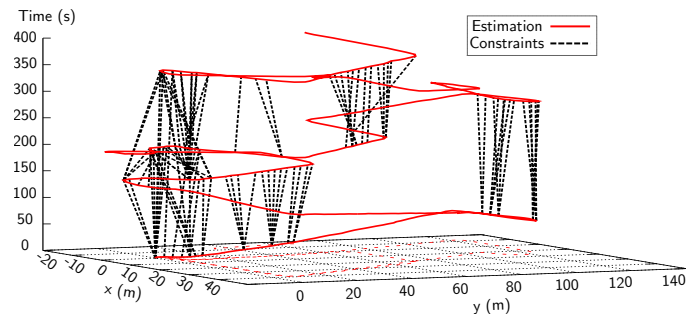
Fig. 7. Part of the estimated track over time using Tango-based pedestrian dead reckoning and fingerprint-based constraints inferred with a similarity threshold of $\vartheta_s = 0.7$.

person to hold two phones (i.e., Lenovo Phab 2 Pro and Sony) and walk in the building along different paths with a regular walking speed. Four tracks were recorded at different times to show the power of the collaborative SLAM. For each track, the user started from the same position. The total distance traveled is 1906 meters with a duration of 2179 seconds and a number of 1702 unique MAC addresses are detected. This results in four log files with a duration of 687, 582, 439, and 471 seconds respectively. The step length s is fixed to 0.7m throughout this paper. Fig. 2(a) shows a snapshot during the experiment.

(a) Visualization of track from user1 using Tango-based PDR

(b) Visualization of track from user2 using Tango-based PDR

(c) Visualization of track from user3 using Tango-based PDR

(d) Visualization of track from user4 using Tango-based PDR

(e) Visualization of track from user1 using step counter-based PDR

(f) Visualization of track from user2 using step counter-based PDR

(g) Visualization of track from user3 using step counter-based PDR

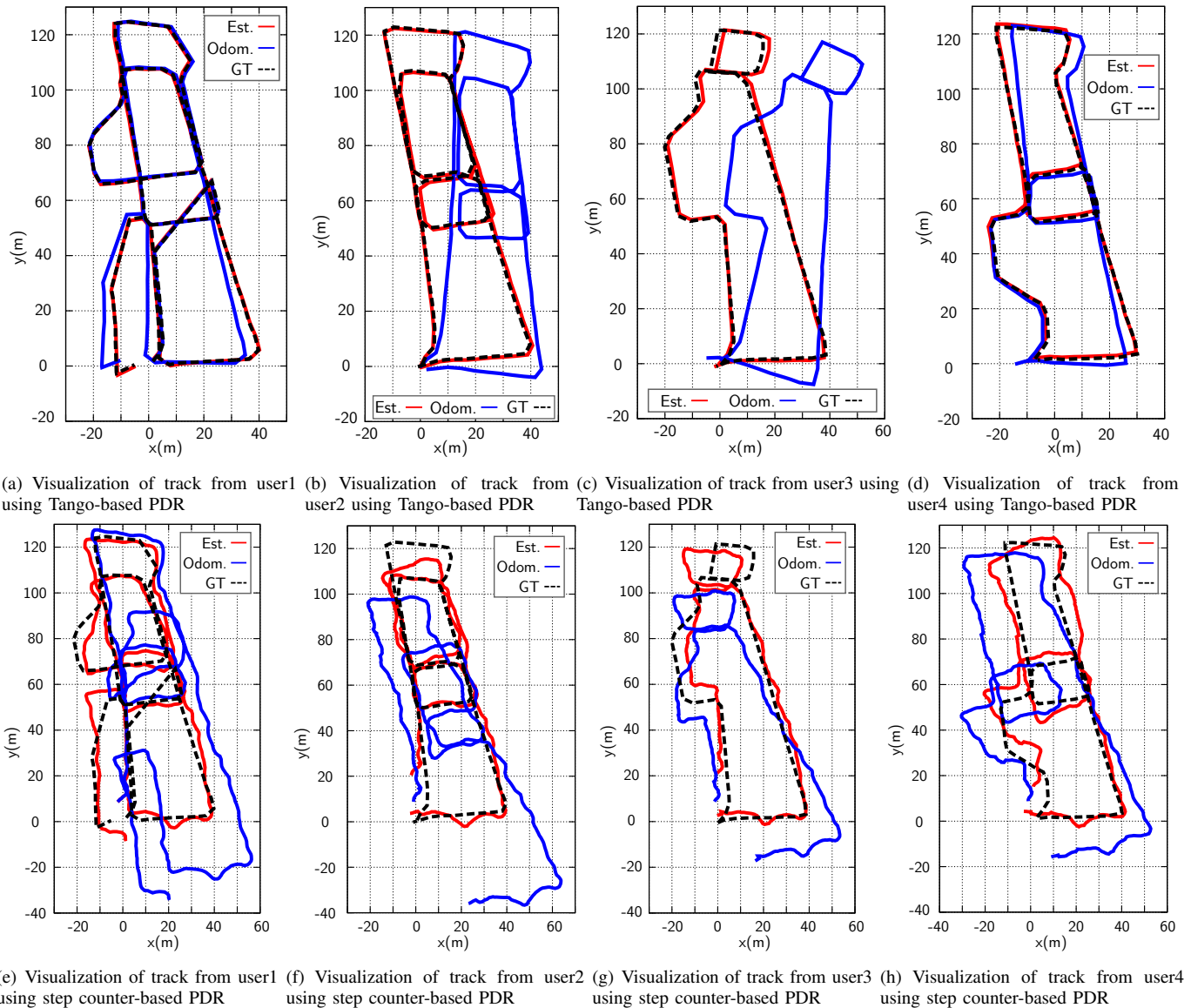(h) Visualization of track from user4 using step counter-based PDR

Fig. 8. A comparison between ground truth from point cloud-based SLAM, estimated track using our approach, and raw odometry of the three individual users using two different pedestrian dead reckoning systems (i.e., Tango and step counter).

TABLE II
THE ROLE OF PARAMETERS IN THE PROPOSED METHOD AND GUIDELINES ON HOW TO CHOOSE THEIR VALUES.

Parameter	Range	Guidelines to choose the parameters
ϑ_s	[0, 1]	Threshold used to claim a loop closure based on similarity between two radio fingerprints. A large ϑ_s will produce less loop closures and a small ϑ_s will result in false loop closures. A setting of $\vartheta_s = 0.7$ is recommended in our approach.
ϑ_r	[-90, -50]	Threshold used to filter out the low RSS measurements. The choice of ϑ_r depends on the distribution of the received signal strength in a radio environment. A value of $\vartheta_r = -70$ is suggested in our approach.
r	[0, 1]	Binning size to train the variance model. A small r requires more training time and training samples. A large r is not able to characterize the detail of the variance model. A setting of $r = 0.2$ produces the best results in our approach.
w	[0, 200]	Window size to identify a turn. A large w will result in more false turnings and a small w is not able to infer the correct turnings. $w = 40$ is suggested in our approach.
ϑ_f	[0, 1]	Threshold used to claim a valid turning match. A large ϑ_f will discard the true turning loop candidates and a small ϑ_f will incorporate too many false turning loops. $\vartheta_f = 0.5$ is appropriate for the application.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2019.2957293, IEEE Internet of Things Journal

TABLE III
EVALUATION OF THE SYSTEM PERFORMANCE WITH RESPECT TO DIFFERENT RSS THRESHOLDS ϑ_r. THE TABLE SHOWS THE POSITIONING ACCURACY (MEAN, STANDARD DEVIATION, MAXIMUM, AND MEDIAN) IN METERS, AVERAGE NUMBER OF MAC ADDRESSES DETECTED, AND AVERAGE COMPUTATIONAL TIME FOR LOOP CLOSURE DETECTION.

ϑ_r	Tango-based				Step counter-based			
	Mean± Std. Dev.	Max./Median	Number of MAC	Comput. time (s)	Mean± Std. Dev.	Max./Median	Number of MAC	Comput. time (s)
-90	1.39±1.51	7.86/1.37	179.23	24.50	14.95±7.09	37.72/14.85	129.36	426.37
-80	1.26±1.21	6.30/1.19	137.81	19.53	5.91±3.76	20.87/5.14	70.08	212.07
-75	1.06±0.49	4.53/0.94	81.29	12.31	5.40±3.60	21.52/4.56	43.36	127.67
-70	0.77±0.46	3.79/0.73	47.97	5.73	5.19±3.48	21.64/4.47	25.49	72.34
-65	0.92±0.44	3.87/0.84	28.94	3.29	5.43±3.43	20.40/4.70	13.41	37.78
-60	1.24±0.89	4.91/1.13	16.22	1.76	5.95±3.82	22.78/5.06	6.08	20.36
-50	2.44±0.95	5.81/2.12	3.29	0.37	11.51±5.82	35.03/11.52	0.67	3.08
odom.	8.03±7.80	31.55/4.22	NA	NA	15.39±7.89	41.23/14.58	NA	NA

B. Ground Truth and Accuracy Comparisons

To extract the ground truth as comparison, we optimized the track from Tango using GraphSLAM taken 3D point clouds as input. We implement loop closure detection based on the point cloud library (PCL) [61]. We identify Harris keypoints in a pair of point clouds and compute the corresponding SHOT (Signature of Histograms of OrienTations) descriptors [62]. We match these descriptors with k-nearest neighbors algorithm (k-NN) and find an initial transformation using SVD (singular value decomposition). The transformation is further refined by ICP. If the number of matched points exceeds a threshold (half size of the point cloud), a loop closure is confirmed and added to the graph as constraints. We treat this optimized path as the ground truth to evaluate the accuracy of our system.

We show the accuracy by the root mean square error (RMSE) between the ground truth and the estimation. Our experiments show that we are able to achieve an accuracy of 0.6 meters with Tango-based PDR and 4.76 meters of a step counter-based PDR with a size of 130m×70m, as shown in Figure 6. The optimized track is annotated with the radio measurement and can serve as the radio map for the positioning of another user. In Table II, we show the important parameters used in this paper and the remarks of how to choose these parameters. The final positioning error is calculated at the end of the process after the loop closure detection, turning matching, and pose graph optimization. Larger positioning errors are expected in a real time system, since we have to process the incoming data in a sequential way and provide pose estimation at regular intervals before loop closures are detected.

To implement a practical indoor localization system, the mechanism to deal with the change of the radio environment is a necessity. The evolution of the radio environment (for example adding or removing the access points) can be examined by looking at the signal variance at similar locations, as shown in [63] [64] [65] [66] [67]. Some MAC addresses might be static for a fixed duration of time, but are essentially mobile for example personal hotspots. A good way to address the mobile hotspots issue is to filter out the MAC addresses by the organizationally unique identifier (OUI), which is used to

uniquely identify a vendor[2]. The MAC address from a phone manufacturer should be removed from the detection list to prevent the uncertainty of incorporating the additional mobile hotspot observations. Another approach to filter out the mobile hotspots is to look at the spatial relations of the detected positions of a particular MAC address [68] [69] [70].

C. Impact of Different RSS Threshold ϑ_r

We examined the influence of RSS thresholding on the accuracy in this series of experiments. We set the similarity threshold $\vartheta_s = 0.7$ and use a binning size $r = 0.2$. We chose ϑ_r values between -90 and -50 to evaluate the mean accuracy, as listed in Table III. As compared to the raw odometry, our approach can effectively reduce the accumulated odometry error: with the setting of $\vartheta_r = -70$, our approach improves the positioning accuracy by 90.4% and 66.3% for Tango (from 8.03m to 0.77m) and step counter (from 15.39m to 5.19m) respectively. In addition, the accuracy of the PDR has a very high impact on the accuracy achieved with our SLAM system: Tango shows a good motion tracking performance and we achieved an accuracy of 0.77m with a threshold $\vartheta_r = -70$. The accuracy achieved with Tango is better than state-of-the-art fingerprinting-based approaches [1], [12]. While the step counter results in a large amount of accumulated odometry error and the accuracy obtained with our approach is worse (5.19m). A further investigation to the PDR system will help to improve the accuracy, which will be one of our future work. One has to note that the accuracy is achieved without training as opposed to the fingerprinting-based approaches, where a time-consuming phase to collect and annotate the fingerprints is prerequisite to guarantee a good positioning accuracy.

Table III also shows that we maintain a good accuracy with an RSS threshold between -75 and -65, while the computational time decreases considerably with the thresholding technique. As an example, for Tango-based system, a threshold of -70 reduces the computation time to 5.73 seconds as compared to a threshold of -90 (i.e., 24.28 seconds). At the same time, the accuracy even increases by 0.62 meters (error drops from 1.39m to 0.77m). A suitable threshold produces

[2]http://standards-oui.ieee.org/oui.txt

2327-4662 (c) 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

TABLE IV
EVALUATION OF THE PROPOSED APPROACH WITH TWO DIFFERENT
PEDESTRIAN DEAD RECKONING SYSTEMS UNDER DIFFERENT SETTINGS
OF SIMILARITY THRESHOLD ϑ_s. THE TABLE SHOWS THE POSITIONING
ACCURACY (MEAN AND STANDARD DEVIATION) IN METERS AND THE
NUMBER OF CONSTRAINTS INFERRED.

ϑ_s	Tango-based		Step counter-based	
	Mean ± Std. Dev	No. of constr.	Mean ± Std. Dev	No. of constr.
0.95	5.63±6.76	18	13.86±7.67	2098
0.9	2.15±0.71	94	6.29±3.67	4241
0.8	1.06±0.51	384	5.49±3.61	10582
0.7	0.77±0.46	715	5.19±3.48	17020
0.6	0.81±0.76	1150	5.35±3.34	23859
0.4	0.80±0.79	2198	5.49±3.66	42625
0.2	1.06±0.89	3494	5.94±4.02	66423
0.1	1.09±0.95	4517	6.09±3.94	85834

TABLE V
EVALUATION OF THE MEAN POSITIONING ACCURACY UNDER THE IMPACT
OF DIFFERENT SETTINGS OF ϑ_s AND ϑ_r FOR TANGO-BASED APPROACH.

ϑ_s	ϑ_r						
	-90	-80	-75	-70	-65	-60	-50
0.9	2.96	2.65	2.38	2.15	1.81	2.08	2.61
0.8	1.68	1.29	1.16	1.06	1.23	1.41	2.58
0.7	1.39	1.26	1.06	0.77	0.92	1.24	2.44
0.6	1.22	1.02	0.87	0.81	0.89	1.27	2.35
0.5	0.88	0.83	0.78	0.82	0.93	1.57	2.33
0.4	0.99	0.91	0.82	0.80	0.85	1.54	2.40
0.3	1.19	1.16	1.08	0.92	0.89	1.61	2.35
0.2	1.26	1.27	1.22	1.06	0.96	1.65	2.35
0.1	1.28	1.29	1.23	1.09	1.01	1.70	2.30

TABLE VI
EVALUATION OF THE POSITIONING ACCURACY (MEAN, STANDARD
DEVIATION, MEDIAN, AND MAXIMUM IN METERS) WITH TWO
PEDESTRIAN DEAD RECKONING SYSTEMS UNDER THE IMPACT OF
DIFFERENT CONFIGURATIONS OF BINNING SIZE r.

r	Tango-based		Step counter-based	
	Mean ± Std. Dev	Max. / Median	Mean ± Std. Dev	Max. / Median
1.0	1.03±0.63	4.29/0.89	6.23±3.84	24.83/5.59
0.8	0.85±0.50	4.22/0.85	5.71±3.50	21.97/5.03
0.6	0.84±0.39	3.62/0.79	5.32±3.54	21.78/4.60
0.4	0.81±0.62	4.26/0.76	5.22±3.51	21.93/4.46
0.2	0.77±0.46	3.79/0.73	5.19±3.48	21.64/4.47
0.1	0.79±0.40	3.59/0.86	5.23±3.79	20.97/4.13
0.05	0.81±0.39	4.00/0.87	5.29±3.52	21.48/4.50

result in a small number of constraints and a decrease of the accuracy. For Tango-based system, we obtain a mean accuracy of 0.77m with $\vartheta_s = 0.7$, which is an improvement of 86.3% as compared to the mean accuracy of 5.63m with $\vartheta_s = 0.95$. Yet, such an improvement is at the expense of a higher number of constraints added (i.e., 715 constraints with $\vartheta_s = 0.7$ as compared to 18 with $\vartheta_s = 0.95$). But the accuracy does not get improved with a threshold smaller than 0.6. One reason could be because a low similarity value will always come along with a very large covariance, and has very less strength to correct the odometric error. A setting of $\vartheta_s = 0.7$ seems to be a good trade off between the accuracy and the number of constraints inferred. Table V showed the accuracy by jointly optimizing the parameters ϑ_s and ϑ_r for Tango-based approach. As can be seen from this table, a careful examination of ϑ_s and ϑ_r will improve the accuracy. A too large or too small will obviously deteriorate the performance of our approach.

E. Impact of the Binning Size of Training

Next, we examined the influence of accuracy with respect to various binning sizes r. We chose RSS threshold $\vartheta_r = -70$ and similarity threshold $\vartheta_s = 0.7$. To evaluate accuracy under impact of different binning sizes, we set r to the following values $r = \{0.05, 0.1, 0.2, 0.4, 0.6, 0.8, 1.0\}$. In our approach, the uncertainty model is trained with all the collected data. We show a comparison of the results in Table VI. This table shows that the best choice of r is 0.2. The covariance estimated with a large r is usually too large to compensate for the error from the odometry. Optimizing r gives an improvement of 25.2% (0.77m and 1.03m for $r = 0.2$ and $r = 1.0$ respectively) for Tango-based PDR and 16.6% (5.19m and 6.23m for $r = 0.2$ and $r = 1.0$ respectively) for step counter-based PDR. Covariance of the loop is a key to optimize the pose graph, as it is the only information to measure how close the two locations are in a loop, therefore, a careful examination of the parameter will lead to an improvement of the accuracy. The covariance added here (see Figure 2(b)) is much smaller as compared to the accumulated odometry error (notice that the maximum positioning error of Tango and step counter in Table III are 31.55m and 41.23m respectively). This is why we are still able to correct the accumulated odometry error. The approach

a good accuracy, as it will filter out the suspicious radio signals. However, a threshold larger than -65 leads to a bad result (for example, 1.24 meters of accuracy with $\vartheta_r = -60$ for Tango-based system). The ground truth, estimation, and odometry of individual tracks using two different pedestrian dead reckoning systems are visualized in Figure 8. A part of estimated trajectory and the constraints inferred are shown in Figure 7.

D. Impact of Different Similarity Threshold ϑ_s

Next, we performed a series of experiments to examine the influence of accuracy with respect to different similarity thresholds ϑ_r. We show the results in Table IV. We fixed RSS threshold $\vartheta_r = -70$ and a binning size $r = 0.2$. We increased the similarity threshold ϑ_s from 0.1 to 0.95 to evaluate the accuracy and the number of constraints inferred. From Table IV, we can observe that the number of constraints is different for Tango and step counter-based PDR due to different sampling rates of the device (5 seconds for Tango and 1 second for step counter): Tango-based PDR offers a small number of constraints as compared to step counter-based PDR. In addition, the threshold has a high impact on the accuracy and the number of constraints. Applying a high threshold will

TABLE VII
COMPARISON OF MEAN POSITIONING ACCURACY UNDER DIFFERENT
SIMILARITY MODELS TRAINED AT TWO DIFFERENT BUILDINGS
REGARDING TANGO-BASED PEDESTRIAN DEAD RECKONING SYSTEM.

Model	Binning size r						
	0.05	0.1	0.2	0.4	0.6	0.8	1.0
Bld1	0.81	0.79	0.77	0.81	0.84	0.85	1.03
Bld2	0.84	0.81	0.78	0.85	0.86	0.91	1.07

presented here provides a way to automatically calibrate the uncertainty model with the odometry measurement. The model generated in different environments might be slightly different, as shown in Figure 2(b). To evaluate the accuracy under the impact of different similarity models, Table VII compared the results using the similarity models produced from two different buildings, namely Building1 (the one used for the verification of the positioning accuracy) and Building2 regarding to Tango-based approach. As can be seen from this table, the two models provide similar positioning accuracy, which proves our assumption that the similarity model can be applied to different environments.
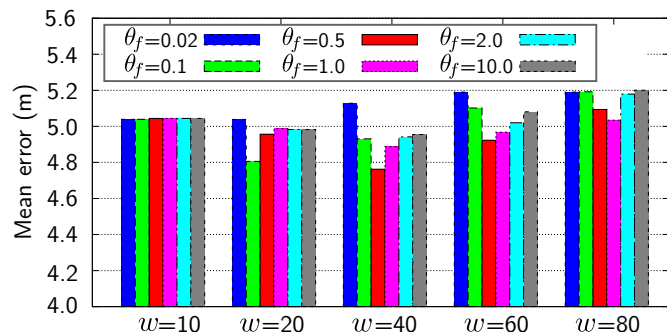
F. Impact of Turning Detection and Matching

We compared the accuracy with and without the integration of turning features in the next series of experiments. We fixed RSS threshold $\vartheta_r = -70$, similarity threshold $\vartheta_s = 0.7$, and binning size $r = 0.2$. We varied the setting of w and ϑ_f to evaluate their impact on the accuracy. Figure 9 shows a comparison of the results. This figure shows that the accuracy can be improved by additional integration of turning features: we obtain an improvement of 22.1% for the Tango-based pedestrian dead reckoning (from 0.77m to 0.6m with $w = 40$ and $\vartheta_f = 0.5$) and 8.3% for step counter-based dead reckoning (from 5.19m to 4.76m with $w = 40$ and $\vartheta_f = 0.5$). The improvement of Tango is slightly higher as compared to the step counter-based system. Since the odometry error of Tango is smaller than step counter, the turning feature here exhibits great capability to correct the drift error of odometry. However, for step counter-based PDR, the error is dominated by the odometry and the turning feature shows less improvement to the accuracy as compared to Tango-based PDR.

The number of turnings detected with different settings of w is shown in Table VIII. The constraints are mostly from the fingerprinting matching. As can be seen from Table VIII, only 26 turnings are detected with $w = 40$, which is much less than the number of fingerprinting-based constraints (715 for Tango-based approach with $\vartheta_s = 0.7$ as shown in Table IV). Due to the low sampling rate, Tango-based PDR leads to a small number of turnings as compared to the step counter-based PDR. A large value of w leads to an increasing number of turnings detected. From Figure 9, we can also observe that a window size $w = 40$ leads to the best accuracy for both systems. A too high or too small w obviously results in a less improvement to the accuracy.

In addition, Figure 9 shows that a too large or too small ϑ_f leads to a decrease of the accuracy. A suitable fitness threshold ϑ_f will help to remove the false turning-based loop
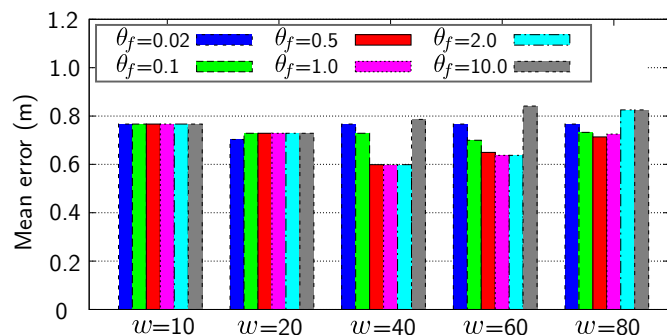
(a) Positioning accuracy of Tango-based pedestrian dead reckoning system under different settings of w

(b) Positioning accuracy of step counter-based pedestrian dead reckoning system under different settings of w

Fig. 9. The impact of turning detection and matching on average positioning accuracy with two different pedestrian dead reckoning systems under various settings of window size w and ϑ_f.

TABLE VIII
COMPARISON OF NUMBER OF TURNINGS DETECTED WITH TWO DIFFERENT
PEDESTRIAN DEAD RECKONING SYSTEMS UNDER VARIOUS SETTINGS OF
WINDOW SIZE w.

Window size w	10	20	40	60	80
Turnings (Tango)	0	6	26	45	78
Turnings (step counter)	0	423	1447	2666	3973

closures. A small ϑ_f will not be able to identify true matched turnings and results in a small number of turning-based loops and therefore has less strength to improve the accuracy. A large ϑ_f produces too many false turning-based loops, which leads to a deterioration of the result. As an example, for step counter-based PDR with $w = 40$, a setting of $\vartheta_f = 0.5$ gives an accuracy of 4.76m, which produces an improvement of 3.8% and 7.0% when compared to $\vartheta_f = 10.0$ (4.95m) and $\vartheta_f = 0.02$ (5.12m), respectively.

G. Computational Time

Finally, we evaluated the time consumption at each stage of our approach. The results are listed in Table IX. In our approach, we process the recordings in a batch matter, which means that we optimized the graph after all loop closures are identified. An Intel Core i5-4200M CPU with 2.5GHz frequency and 4GB RAM is used to process the measurements. We set $\vartheta_r = -70$, $\vartheta_s = 0.7$, $r = 0.2$, $w = 40$, and $\vartheta_f = 0.5$. As can be seen from Table IX, the entire data processing took 17.81 (6.08+5.73+5.93+0.07) seconds for tango-based

TABLE IX
EVALUATION OF THE COMPUTATIONAL TIME (IN SECONDS) IN EACH
STAGE USING TWO DIFFERENT PEDESTRIAN DEAD RECKONING SYSTEMS.

Stage	duration(s)	
	Tango	Step counter
Data recording (time per track)	544.75	544.75
Model training+ variance computation	6.08	23.58
Loop closure detection	5.73	72.34
Turning detection and matching	5.93	31.82
Pose graph optimization	0.07	1.65

dataset, which is almost 30 times faster than the data recording stage (544.75 seconds), while the processing time is much longer for step counter-based dataset (approx. 129.39 seconds, i.e., 23.58+72.34+31.82+1.65), due to its higher sampling rate. Additionally, optimization of the graph only took less than two seconds (0.07 and 1.65 seconds for Tango and step counter-based system respectively). In our current offline implementation, we compute the similarity of all pairs of Wi-Fi measurements in the entire dataset to find the potential loop closures. This module consumes too much time (72.34 seconds for the step counter version as shown in Table IX), and thus cannot be used for real time pose estimation. However, we believe the computation can be further optimized to make online implementation possible, e.g. we only need to compare the current Wi-Fi measurement with the previous measurements for the loop closure detection (estimated to be 72.34/2179 = 0.03 seconds), which is less than the current Wi-Fi sampling rate of one second interval. Nonetheless, we believe further optimization is needed to ensure online implementation in real time.

When more users are involved in the experiment, it will take longer time to run the algorithm due to the increasing number of nodes and the constraints in the graph. Still, our approach is efficient when compared to vision-based approaches, as vision-based approach requires heavy computational resources due to feature extraction and feature matching. In addition, there can be data association problems which will result in the loop closure failure. On other hand, the MAC address of the AP is unique. It is not necessary to run the optimization algorithm for each new Wi-Fi measurement. We suggest to run the optimization when a loop closure is detected or a certain amount of loop closures has been identified. Solutions to reduce the computational time can be found in [71] [72]. One might notice that the model training described in Section III-E takes long time. But this phase can be performed offline, and the learned model can be saved and applied to other users or different environments once it is generated.

V. CONCLUSIONS AND FUTURE WORK

In this paper, we presented a novel approach for collaborative simultaneous localization and radio fingerprint mapping (C-SLAM-RF) in unknown environments. The proposed system makes use of a pedestrian dead reckoning system and the RSS measurement from surrounding wireless access points. We further incorporate the motion features to improve the accuracy of the system. The proposed approach does not require any knowledge of the map and locations of the access points. The performance of our approach is evaluated in a large scale environment under two pedestrian dead reckoning systems with different motion tracking accuracies. Our results reveal that the accuracy of a pedestrian tracking system plays an important role in the accuracy of our approach. We obtained an accuracy of 0.6m and 4.76m for Tango and step counter-based pedestrian dead reckoning systems, respectively. The quality of the radio map will increase with more users involved in collecting the measurements due to the crowdsourcing nature of the proposed approach. One of our future work is to enhance the accuracy of step counter-based PDR by stride length estimation and the fusion of gyroscope measurement. Another direction would be the evaluation of the indoor positioning accuracy by applying the radio map constructed from our SLAM system.

REFERENCES

[1] A. Yassin, Y. Nasser, M. Awad, A. Al-Dubai, R. Liu, C. Yuen, and R. Raulefs, "Recent advances in indoor localization: A survey on theoretical approaches and applications," *IEEE Communications Surveys Tutorials*, vol. 19, no. 2, pp. 1327–1346, Secondquarter 2017.

[2] S. He and S.-H. G. Chan, "Wi-fi fingerprint-based indoor positioning: Recent advances and comparisons," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 466–490, 2016.

[3] X. Guo, S. Zhu, L. Li, F. Hu, and N. Ansari, "Accurate wifi localization by unsupervised fusion of extended candidate location set," *IEEE Internet of Things Journal*, pp. 1–1, 2019.

[4] W. Li, Z. Chen, X. Gao, W. Liu, and J. Wang, "Multi-model framework for indoor localization under mobile edge computing environment," *IEEE Internet of Things Journal*, pp. 1–1, 2019.

[5] M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit, "FastSLAM: A factored solution to the simultaneous localization and mapping problem," in *Proceedings of the Eighteenth National Conference on Artificial Intelligence (AAAI 2002)*, Alberta, Canada, July 28–August 01 2002, pp. 593–598.

[6] B. Ferris, D. Fox, and N. Lawrence, "WiFi-SLAM using gaussian process latent variable models," in *Proceedings of the 20th International Joint Conference on Artifical Intelligence*, San Francisco, CA, USA, January 06–12 2007, pp. 2480–2485.

[7] W. Burgard, C. Stachniss, G. Grisetti, B. Steder, R. Kümmerle, C. Dornhege, M. Ruhnke, A. Kleiner, and J. D. Tardós, "A comparison of slam algorithms based on a graph of relations," in *Proceedings of the 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, St. Louis, USA, October 11-15 2009, pp. 2089–2095.

[8] S. Kuutti, S. Fallah, K. Katsaros, M. Dianati, F. Mccullough, and A. Mouzakitis, "A survey of the state-of-the-art localization techniques and their potentials for autonomous vehicle applications," *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 829–846, April 2018.

[9] T. Taketomi, H. Uchiyama, and S. Ikeda, "Visual slam algorithms: a survey from 2010 to 2016," *IPSJ Transactions on Computer Vision and Applications*, vol. 9, no. 1, p. 16, June 2017.

[10] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard, "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age," *IEEE Transactions on Robotics*, vol. 32, no. 6, pp. 1309–1332, December 2016.

[11] F. Lu and E. Milios, "Globally consistent range scan alignment for environment mapping," *Autonomous Robots*, vol. 4, no. 4, pp. 333–349, October 1997.

[12] R. Liu, C. Yuen, T. N. Do, and U. X. Tan, "Fusing similarity-based sequence and dead reckoning for indoor positioning without training," *IEEE Sensors Journal*, vol. 17, no. 13, pp. 4197–4207, July 2017.

[13] J. Huang, D. Millman, M. Quigley, D. Stavens, S. Thrun, and A. Aggarwal, "Efficient, generalized indoor wifi graphslam," in *Proc. of the 2011 IEEE International Conference on Robotics and Automation (ICRA)*, Shanghai, China, May 09–13 2011, pp. 1038–1043.

[14] Y. Zhuang, J. Yang, L. Qi, Y. Li, Y. Cao, and N. El-Sheimy, "A pervasive integration platform of low-cost mems sensors and wireless signals for indoor localization," *IEEE Internet of Things Journal*, vol. 5, no. 6, pp. 4616–4631, December 2018.

[15] M. Zhou, Y. Tang, Z. Tian, L. Xie, and W. Nie, "Robust neighborhood graphing for semi-supervised indoor localization with light-loaded location fingerprinting," *IEEE Internet of Things Journal*, vol. 5, no. 5, pp. 3378–3387, October 2018.

[16] B. Zhou, Q. Li, Q. Mao, W. Tu, and X. Zhang, "Activity sequence-based indoor pedestrian localization using smartphones," *IEEE Transactions on Human-Machine Systems*, vol. 45, no. 5, pp. 562–574, October 2015.

[17] S. Kaiser and C. Lang, "Detecting elevators and escalators in 3d pedestrian indoor navigation," in *2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Alcala de Henares, Spain, October 04–07 2016, pp. 1–6.

[18] S. Wang, H. Wen, R. Clark, and N. Trigoni, "Keyframe based large-scale indoor localisation using geomagnetic field and motion pattern," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Daejeon, Korea, October 9-14 2016, pp. 1910–1917.

[19] A. Khalajmehrabadi, N. Gatsis, and D. Akopian, "Modern wlan fingerprinting indoor positioning methods and deployment challenges," *IEEE Communications Surveys Tutorials*, vol. 19, no. 3, pp. 1974–2002, thirdquarter 2017.

[20] R. Liu, C. Yuen, T. N. Do, W. Guo, X. Liu, and U-X. Tan, "Relative positioning by fusing signal strength and range information in a probabilistic framework," in *the Second IEEE International Workshop on Localization and Tracking: Indoors, Outdoors, and Emerging Networks (IEEE LION 2016)*, Washington, USA, December 04–08 2016.

[21] R. Liu, C. Yuen, T. N. Do, D. Jiao, X. Liu, and U-X. Tan, "Cooperative relative positioning of mobile users by fusing imu inertial and uwb ranging information," in *the 2017 IEEE International Conference on Robotics and Automation (ICRA 2017)*, Singapore, May 29-June 3 2017, pp. 5623–5629.

[22] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics (Intelligent Robotics and Autonomous Agents)*. The MIT Press, 2005.

[23] R. Kuemmerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard, "g2o: A general framework for graph optimization," in *Proceedings of the 2011 IEEE International Conference on Robotics and Automation (ICRA)*, Shanghai, China, May 09–13 2011, pp. 3607–3613.

[24] N. Engelhard, F. Endres, J. Hess, J. Sturm, and W. Burgard, "Real-time 3d visual slam with a hand-held rgb-d camera," in *Proc. of the RGB-D Workshop on 3D Perception in Robotics at the European Robotics Forum, Vasteras, Sweden*, vol. 180, 2011, pp. 1–15.

[25] B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon, "Bundle adjustment–a modern synthesis," in *Vision Algorithms: Theory and Practice, Lecture Notes in Computer Science (LNCS)*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2000, pp. 298–375.

[26] "Project Tango," 2017. [Online]. Available: https://get.google.com/tango/

[27] J. Jung, T. Oh, and H. Myung, "Magnetic field constraints and sequence-based matching for indoor pose graph slam," *Robotics and Autonomous Systems*, vol. 70, pp. 92–105, 2015.

[28] M. Angermann and P. Robertson, "Footslam: Pedestrian simultaneous localization and mapping without exteroceptive sensors–hitchhiking on human perception and cognition," *Proceedings of the IEEE*, vol. 100, no. Special Centennial Issue, pp. 1840–1848, May 2012.

[29] M. Hardegger, D. Roggen, S. Mazilu, and G. Trster, "Actionslam: Using location-related actions as landmarks in pedestrian slam," in *2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Sydney, Australia, Nov. 13–15 2012, pp. 1–10.

[30] M. G. Puyol, P. Robertson, and O. Heirich, "Complexity-reduced footslam for indoor pedestrian navigation," in *2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Sydney, Australia, Nov. 13–15 2012, pp. 1–10.

[31] Y. He, J. Liang, and Y. Liu, "Pervasive floorplan generation based on only inertial sensing: Feasibility, design, and implementation," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 5, pp. 1132–1140, May 2017.

[32] D. Sikeridis, B. P. Rimal, I. Papapanagiotou, and M. Devetsikiotis, "Unsupervised crowd-assisted learning enabling location-aware facilities," *IEEE Internet of Things Journal*, vol. 5, no. 6, pp. 4699–4713, December 2018.

[33] C. Wu, Z. Yang, and Y. Liu, "Smartphones based crowdsourcing for indoor localization," *IEEE Transactions on Mobile Computing*, vol. 14, no. 2, pp. 444–457, February 2015.

[34] P. Zhang, R. Chen, Y. Li, X. Niu, L. Wang, M. Li, and Y. Pan, "A localization database establishment method based on crowdsourcing

[35] R. Faragher, C. Sarno, and M. Newman, "Opportunistic radio slam for indoor navigation using smartphone sensors," in *Proceedings of the 2012 IEEE/ION Position Location and Navigation Symposium (PLANS 2012)*, Myrtle Beach, South Carolina, USA, April 23-26 2012, pp. 120–128.

[36] V. Radu, L. Kriara, and M. K. Marina, "Pazl: A mobile crowdsensing based indoor wifi monitoring system," in *Proceedings of the 9th International Conference on Network and Service Management (CNSM 2013)*, Zürich, Switzerland, 2013, pp. 75–83.

[37] R. Gao, M. Zhao, T. Ye, F. Ye, Y. Wang, K. Bian, T. Wang, and X. Li, "Jigsaw: Indoor floor plan reconstruction via mobile crowdsensing," in *Proceedings of the 20th Annual International Conference on Mobile Computing and Networking*, Maui, Hawaii, USA, September 07-11 2014, pp. 249–260.

[38] M. Jirku, V. Kubelka, and M. Reinstein, "Wifi localization in 3d," in *Proceedings of the 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2016)*, Daejeon, Korea, October 9-14 2016, pp. 4551–4557.

[39] L. Chen, K. Yang, and X. Wang, "Robust cooperative wi-fi fingerprint-based indoor localization," *IEEE Internet of Things Journal*, vol. 3, no. 6, pp. 1406–1417, December 2016.

[40] V. S. Saurav Agarwal and S. Chakravorty, "Rfm-slam: Exploiting relative feature measurements to separate orientation and position estimation in slam," in *Proc. of the 2017 IEEE International Conference on Robotics and Automation (ICRA)*, Singapore, May 29–June 3 2017.

[41] S. Agarwal, K. Mierle, and Others, "Ceres solver," http://ceres-solver. org.

[42] A. Sujiwo, T. Ando, E. Takeuchi, Y. Ninomiya, and M. Edahiro, "Monocular vision-based localization using ORB-SLAM with lidar-aided mapping in real-world robot challenge," *Journal of Robotics and Mechatronics*, vol. 28, no. 4, pp. 479–490, 2016.

[43] H. Xiong and D. Tao, "A diversified generative latent variable model for wifi-slam," in *The Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17)*, California, USA, February 4-9 2017.

[44] R. Liu, A. Koch, and A. Zell, "Path following with passive UHF RFID received signal strength in unknown environments," in *Proc. of the 2012 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS 2012)*, Vilamoura, Algarve, Portugal, October 07–12 2012, pp. 2250–2255.

[45] P. Vorst and A. Zell, "A comparison of similarity measures for localization with passive RFID fingerprints," in *Proc. of the joint Conf. of 41st Int. Symposium on Robotics and 6th German Conf. on Robotics*, Munich, Germany, June 07–09 2010, pp. 354–361.

[46] P. Vorst, A. Koch, and A. Zell, "Efficient self-adjusting, similarity-based location fingerprinting with passive UHF RFID," in *Proc. of IEEE Int. Conf. on RFID-Technology and Applications (RFID-TA 2011)*, Spain, September 15–16 2011, pp. 160–167.

[47] A. Rai, K. K. Chintalapudi, V. N. Padmanabhan, and R. Sen, "Zee: Zero-effort crowdsourcing for indoor localization," in *Proceedings of the 18th Annual International Conference on Mobile Computing and Networking*, ser. Mobicom '12, August 22–26 2012, pp. 293–304.

[48] T. N. Do, R. Liu, C. Yuen, M. Zhang, and U-X. Tan, "Personal dead reckoning using imu mounted on upper torso and inverted pendulum model," *IEEE Sensors Journal*, vol. 16, no. 21, pp. 7600–7608, November 1 2016.

[49] J. Zhang and S. Singh, "Visual-lidar odometry and mapping: low-drift, robust, and fast," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*, Seattle, USA, May 26-30 2015, pp. 2174–2181.

[50] J. Nilsson, A. K. Gupta, and P. Händel, "Foot-mounted inertial navigation made easy," in *2014 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Busan, South Korea, October 2014, pp. 24–29.

[51] R. Dubé, A. Gawel, H. Sommer, J. Nieto, R. Siegwart, and C. Cadena, "An online multi-robot slam system for 3d lidars," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Sep. 2017, pp. 1004–1011.

[52] Z. Jiang, J. Zhu, Y. Li, J. Wang, Z. Li, and H. Lu, "Simultaneous merging multiple grid maps using the robust motion averaging," *Journal of Intelligent and Robotic Systems*, vol. 94, no. 3-4, pp. 655–668, June 2019.

[53] S. G. Shahbandi, M. Magnusson, and K. Iagnemma, "Nonlinear optimization of multimodal two-dimensional map alignment with application to prior knowledge transfer," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 2040–2047, July 2018.

[54] 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 Transactions on Mobile Computing*, vol. 17, no. 1, pp. 141–154, Jan 2018.

[55] B. Zhou, Q. Li, G. Zhai, Q. Mao, J. Yang, W. Tu, W. Xue, and L. Chen, "A graph optimization-based indoor map construction method via crowdsourcing," *IEEE Access*, vol. 6, pp. 33 692–33 701, 2018.

[56] G. Shen, Z. Chen, P. Zhang, T. Moscibroda, and Y. Zhang, "Walkie-markie: Indoor pathway mapping made easy," in *Presented as part of the 10th USENIX Symposium on Networked Systems Design and Implementation (NSDI 13)*, Lombard, IL, 2013, pp. 85–98.

[57] Z. Li, X. Zhao, Z. Zhao, F. Hu, H. Liang, and T. Braun, "Crowdsensing indoor walking paths with massive noisy crowdsourcing user traces," in *2018 IEEE Global Communications Conference (GLOBECOM)*, Dec 2018, pp. 1–6.

[58] P. J. Besl and N. D. McKay, "A method for registration of 3-d shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, Feb 1992.

[59] Y. Sun and B. Wang, "Indoor corner recognition from crowdsourced trajectories using smartphone sensors," *Expert Systems with Applications*, vol. 82, no. 2017, pp. 266–277, 2017.

[60] D. Galvez-Lopez and J. D. Tardos, "Real-time loop detection with bags of binary words," in *Proceedings of the 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, San Francisco, CA, USA, September 25–30 2011, pp. 51 –58.

[61] R. B. Rusu and S. Cousins, "3d is here: Point cloud library (pcl)," in *Proceedings of the 2011 IEEE International Conference on Robotics and Automation (ICRA)*, Shanghai, China, May 9–13 2011, pp. 1–4.

[62] F. Tombari, S. Salti, and L. D. Stefano, "Unique signatures of histograms for local surface description," in *the 11th European conference on computer vision (ECCV)*, Crete, Greece, September 05–11 2010, pp. 356–369.

[63] X. Liu, J. Cen, Y. Zhan, and C. Tang, "An adaptive fingerprint database updating method for room localization," *IEEE Access*, vol. 7, pp. 42 626–42 638, 2019.

[64] C. Wu, Z. Yang, C. Xiao, C. Yang, Y. Liu, and M. Liu, "Static power of mobile devices: Self-updating radio maps for wireless indoor localization," in *2015 IEEE Conference on Computer Communications (INFOCOM)*, April 2015, pp. 2497–2505.

[65] S. He and K. G. Shin, "Steering crowdsourced signal map construction via bayesian compressive sensing," in *2018 IEEE Conference on Computer Communications (INFOCOM)*, April 2018, pp. 1016–1024.

[66] S. Jung and D. Han, "Automated construction and maintenance of wi-fi radio maps for crowdsourcing-based indoor positioning systems," *IEEE Access*, vol. 6, pp. 1764–1777, 2018.

[67] C. Chu and S. Yang, "A particle filter based reference fingerprinting map recalibration method," *IEEE Access*, vol. 7, pp. 111 813–111 827, 2019.

[68] J. Wang, N. Tan, J. Luo, and S. J. Pan, "Woloc: Wifi-only outdoor localization using crowdsensed hotspot labels," in *2017 IEEE Conference on Computer Communications (INFOCOM)*, May 2017, pp. 1–9.

[69] J. Wang, J. Luo, S. J. Pan, and A. Sun, "Learning-based outdoor localization exploiting crowd-labeled wifi hotspots," *IEEE Transactions on Mobile Computing*, vol. 18, no. 4, pp. 896–909, April 2019.

[70] M. Zhou, H. Yuan, L. Xie, W. Tan, and Z. Tian, "Mobile and redundant access point reduction for indoor unmanned aerial vehicle positioning using wlan crowdsourcing fingerprints," *Physical Communication*, vol. 36, p. 100770, 2019.

[71] K. Konolige and J. Bowman, "Towards lifelong visual maps," in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Oct 2009, pp. 1156–1163.

[72] A. Walcott-Bryant, M. Kaess, H. Johannsson, and J. J. Leonard, "Dynamic pose graph slam: Long-term mapping in low dynamic environments," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Oct 2012, pp. 1871–1878.