CrowdFusion: Multi-Signal Fusion SLAM Positioning Leveraging Visible Light

Zan Li, Xiaohui Zhao, Zhongliang Zhao, Torsten Braun

***Abstract*—With the fast development of location-based services, an ubiquitous indoor positioning approach with high accuracy and low calibration has become increasingly important. In this work, we target on a crowdsourcing approach with zero calibration effort based on visible light, magnetic field and WiFi to achieve sub-meter accuracy. We propose a CrowdFusion Simultaneous Localization and Mapping (SLAM) comprised of coarse-grained and fine-grained trace merging respectively based on the Iterative Closest Point (ICP) SLAM and GraphSLAM. ICP SLAM is proposed to correct the relative locations and directions of crowdsourcing traces and GraphSLAM is further adopted for fine-grained pose optimization. In CrowdFusion SLAM, visible light is used to accurately detect loop closures and magnetic field to extend the coverage. According to the merged traces, we construct a radio map with visible light and WiFi fingerprints. An enhanced particle filter fusing inertial sensors, visible light, WiFi and floor plan is designed, in which visible light fingerprinting is used to improve the accuracy and increase the resampling/rebooting efficiency. We evaluate CrowdFusion based on comprehensive experiments. The evaluation results show a mean accuracy of $0.67m$ for the merged traces and $0.77m$ for positioning, merely replying on crowdsourcing traces without professional calibration.**

***Index Terms*—Crowdsourcing, Visible Light, SLAM**

I. Introduction

In recent years, location-based services become increasingly important and widely used due to the fast development of Internet services related to locations of people. To support those location-based services, diverse positioning techniques have been proposed in last decade. For indoor environments, since the Global Positioning System (GPS) is not available, wireless signals are widely adopted for positioning, such as Radio Frequency Identification (RFID), Ultra Wide Band (UWB), WiFi and Bluetooth. RFID and UWB can achieve high positioning accuracy [1], but the limited deployment of RFID readers and UWB base stations reduces their ubiquitousness [2]. Fingerprinting approaches based on Received Signal Strength Indicator (RSSI) of WiFi and Bluetooth are the most ubiquitous and mature solutions, but their accuracy is limited on meter level due to severe multi-path propagation [3], [4]. To achieve ubiquitous positioning with high accuracy, many solutions relying on ubiquitous signals besides WiFi have been proposed, such as visible light and magnetic field. A traditional visible light positioning approach normally relies on range estimation. In visible light positioning, the Received Signal Strength (RSS) of visible light is converted to the propagation distance based on a propagation model. Because the RSS of visible light is less affected by multi-path propagation, the positioning accuracy based on visible light can achieve sub-meter level. However, locations of Light Emitting Diode (LEDs) are required and propagation models need to be calibrated, which limit the deployment of a visible light positioning system. In our previous work [5], we propose a fingerprinting approach based on visible light, which does not require LEDs with known locations and correspondingly reduces the deployment efforts compared with the range-based approaches.

Although fingerprinting approaches with either WiFi or visible light have achieved acceptable accuracy, the calibration effort of a radio map is extremely time consuming and labor intensive. To reduce calibration efforts, GraphSLAM techniques have been proposed to merge moving traces based on ubiquitous signals, such as WiFi, Bluetooth or visible light. Under the constraints of loop closures detected by ubiquitous signals and inertial sensors, poses are optimized following an assumption of the Gaussian distributed errors. Hence, these GraphSLAM approaches require long traces with loop closures. Moreover, in GraphSLAM, the errors of initial relative locations and directions among the obtained traces should be small to guarantee the convergence of optimization, which limits its application.

Recently, to reduce the calibration effort of fingerprinting, crowdsourcing approaches for automatic radio map construction have attracted more attention. Crowdsourcing positioning generates a radio map by labeling RSSI values with locations on merged crowdsourcing traces without participation of professionals. Walkie-Markie [6] and PiLoc [7] are two pioneering works on merging user traces based on WiFi signals as signal-marks. Both assume small rotation errors of the collected traces to guarantee convergence of the algorithms. In our previous work [5], we adopt a visible light GraphSLAM to merge a long user trace with small relative direction and location errors. However, in a real scenario, rotation errors of crowdsourcing traces are normally large and relative locations are unknown, resulting in failures of these algorithms. Hence, *it is challenging to accurately merge those crowdsourcing traces with unknown relative directions and locations.*

In this work, we propose a system named CrowdFusion to leverage crowdsourcing traces from multiple users to construct radio maps by fusing visible light and magnetic field. Based on the radio map, CrowdFusion system locates the users with

This work is supported by the National Natural Science Foundation of China (62171197) and the Science and Technology Development Program of Jilin Province (20210508059RQ). (Corresponding author: Zan Li)

Zan Li and Xiaohui Zhao are with College of Communication Engineering, Jilin University, China (e-mail: zanli@jlu.edu.cn; xhzhao@jlu.edu.cn);

Zhongliang Zhao is with the School of Electronic and Information Engineering, Beihang University, China (e-mail: zhaozl@buaa.edu.cn);

Torsten Braun is with the Institute of Computer Science, University of Bern, Switzerland (e-mail: braun@inf.unibe.ch).

sub-meter accuracy and without professional calibration. The contributions of this work are from three aspects, i.e., an enhanced signal SLAM approach for merging crowdsourcing traces, a robust loop closure detection approach leveraging visible light and magnetic field, and an improved particle filter for sensor fusion positioning. The details of contributions are summarized as follows.

- **Enhanced Signal SLAM**: We propose a CrowdFusion SLAM with two step trace merging based on an enhanced ICP SLAM and GraphSLAM to merge the crowdsourcing traces. ICP SLAM is enhanced to coarse-grained merge the traces as rigid bodies by correcting the relative rotations and locations among them. Afterwards, GraphSLAM is adopted to fine-grained tune the poses in the merged traces of ICP SLAM. Based on this two step trace merging approach, the crowdsourcing traces can be fine-grained merged even with unknown relative rotations and locations.
- **Robust Loop Closures**: In CrowdFusion, we propose to leverage visible light to detect loop closures in both ICP SLAM and GraphSLAM. To solve the challenges of detecting loop closures based on visible light, a sequential pose matching and a set of validation are proposed to reduce the false positive matches. Besides visible light for the loop closure detection, GraphSLAM further takes the magnetic field data to detect the loop closures for compensating the limited coverage of visible light.
- **Improved Particle filter**: After merging the traces, we build radio maps with visible light and WiFi fingerprints. We design an enhanced particle filter in CrowdFusion by fusing inertial sensors, WiFi, visible light and floor plan. In CrowdFusion, visible light fingerprinting is adopted in the observation likelihood of particle filter and re-sampling/rebooting the particles to respectively improve the positioning accuracy and efficiently solve the sample degeneracy problem.

According to the evaluation results in a complex indoor environments, the proposed CrowdFusion SLAM can achieve a sub-meter level accuracy relying on multiple traces of unknown rotations and locations with zero-effort calibration. Based on the produced crowdsourcing radio map, the proposed particle filter achieves a mean accuracy of $0.77m$, which is significantly higher than that of the traditional fusion approach based on WiFi fingerprinting as observations.

This paper is organized as follows. The related works are reviewed in Section II. The detailed proposed CrowdFusion positioning system description is given in Section III. The system implementation and experiment are shown in Section IV and Section V. Finally, the conclusion is provided in Section VI.

II. Related Work

Over the past decades, indoor positioning shows growing popularity for the increasing demand of location-aware applications [8]. Intensive research has been conducted in the field of indoor positioning mainly from two directions, i.e., enhancing deployment efficiency and improving positioning accuracy. To enhance the deployment efficiency, crowdsoucing and signal SLAM approaches have been proposed in recent years. To improve the positioning accuracy, many ubiquitous signals have been investigated, such as visible light. We summarize the related works from three aspects, i.e., signal SLAM, crowdsourcing positioning and visible light positioning.

A. Signal SLAM

SLAM is a task to estimate a robot location and reconstruct an environment based on observation from LiDAR and other sensors, which is widely used in robotic applications [9], [10]. Many techniques such as Kalman filter [11], particle filter [12] and graph-based solutions [13] have been proposed. Besides the traditional SLAM, signal SLAM developed for light-weight mobile devices with WiFi, magnetic field or other signals has been investigated for indoor positioning, which can build signal maps of buildings [14]. C-SLAM-RF [15] applies RSSI from WiFi access points and the Pedestrian Dead Reckoning (PDR) from user's smartphone to implement WiFi SLAM. However, the low spatial insensitivity of WiFi fingerprints means that it can only provide a low positioning accuracy. GraphSLAM [16] uses magnetic signals to find loop closures to design signal SLAM, which may be limited by magnetic environment of a building due to their rapid change. To improve accuracy, some researches use fused multiple signals. The authors in [17] introduce WiFi similarity constraints to the magnetic sequence matching in a complicated indoor environments. The authors in [18] use the multi-path parameters extracted from radio signals as input measurements and develop a neural network to estimate the states of a mobile agent. The authors in [19] propose a DCOGI-SLAM positioning framework, in which multiple traces can be fused based on Bluetooth signals. The authors in [20] propose a cooperative localization method based on Laser-Visual-Inertial (LVI) and GPS sensors achieved by communicating optimization events between a LiDAR-Inertial-SLAM (LI-SLAM) and Visual-Inertial-SLAM (VI-SLAM). In our previous work [5], we propose a GraphSLAM approach based on visible light to detect the loop closures.

Although these approaches can reduce the efforts of radio map construction, they still require professionals to collect long traces with loop closures in targeted buildings.

B. Crowdsourcing Positioning

When indoor space becomes huge, to construct a radio map with the data collected by professionals is more labor-intensive. Different from the pre-measured solutions, crowdsourcing is an economical and efficient way to obtain location information from normal users [3]. Crowdsourcing positioning based on WiFi RSSI is the most widely used method. Zee [21] is an early crowdsourcing indoor positioning system to leverage the convergence of true locations with WiFi fingerprints. LIFS [22] maps a high dimensional fingerprint space formed by the fingerprints and the PDR traces to a floor plan to construct a 3-D radio map. PiLoc [23] divides a single trajectory into disjoint path segments by turn detection and merges the segments by path shape correlation and WiFi

similarity. MPiLoc [7] also leverages this WiFi similarity to merge the traces. KAILOS [24] provides a crowdsourcing platform and a toolkit for contributors to upload indoor maps and construct radio maps at target buildings. Other RSSI such as Bluetooth, magnetic field and multi-signal fusion methods can also be used in crowdsourcing positioning. FineLoc [25] adopts a reference provided by iBeacons nodes to merge the user traces which can profile detailed layout information of indoor space. The authors in [26] take big data techniques to obtain user traces and mark these traces with an accuracy indicator to improve positioning accuracy by integrating inertial, wireless and magnetic sensors, which requires more time to collect enough PDR traces. While SoiCP [27] proposes a seamless outdoor–indoor crowdsensing positioning which uses WiFi and gates as landmarks to merge user traces.

However, these crowdsourcing approaches with WiFi, Bluetooth and magnetic signals can only achieve meter-level accuracy due to low discrimination and severe multipath propagation.

C. Visible Light Positioning

To further improve positioning accuracy, visible light is investigated due to its high tolerance to multipath propagation. A Visible Light Positioning (VLP) system consists of transmitters and receivers. For receivers, there are Photo Detector (PD) sensors and imaging sensors. Using PD sensor is a low cost solution and provides high rate data communication, while imaging sensors are more costly and only suitable for low data rate transmission. After extracting position related parameters from receivers through some commonly used methods, such as Time Of Arrival (TOA), Time Differences Of Arrival (TDOA), Phase Difference Of Arrival (PDOA), Angle Of Arrival (AOA) or RSSI, position estimation can be performed. The primary requirement for TOA or TDOA information in VLP systems to estimate position is to precisely synchronize embedded clocks in LED transmitters and receivers, which will increase costs [28], [29], [30]. The authors in [31] propose an improved PDOA scheme, which avoids using local oscillators at the receiver. The information regarding AOA can be implemented using PDs [32]. Normally, the VLC receiver structures involving multiple PDs with different orientations can be employed to obtain AOA information from the received signals. The system using RSSI [33], [34] needs to know the locations of LED nodes as the anchors and trains a ranging model for each anchor, which is laborious.

Comparing with the VLP systems mentioned above, our proposed CrowdFusion system can construct a visible light map for positioning with high accuracy only based on the crowdsourcing traces from normal users.

III. CrowdFusion SLAM for Trace Merging

To leverage the crowdsourcing traces with unknown relative directions and locations, we design a CrowdFusion system to achieve sub-meter level positioning without professional calibration. As shown in Fig. 1, in this system, CrowdFusion SLAM is proposed to make use of visible light and magnetic field for a two-step trace merging algorithm, i.e., multi-trace merging based on visible light ICP SLAM and fine-grained trace merging based on GraphSLAM. After fusing the traces, the system builds a radio map with visible light and WiFi signals for sensor fusion positioning based on an improved particle filter.

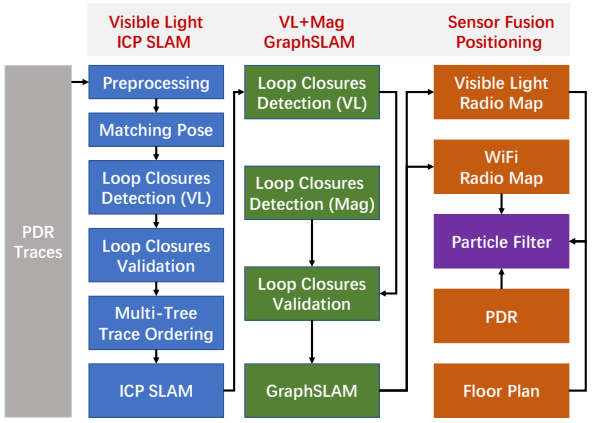

Fig. 1: CrowdFusion System

A. Multi-Trace Merging Based on Visible Light ICP SLAM

Crowdsourcing traces collected from normal users are with unknown relative directions and locations among different traces. Hence, it is challenging to directly apply pose optimization approaches on individual poses. In CrowdFusion, we first apply an ICP SLAM based on visible light data for loop closure detection to merge the large amount of crowdsourcing traces by correcting the relative directions and locations of the traces, in which each trace is considered as a rigid body and the relative locations of poses in one trace are unaltered.

1) Preprocessing of Visible Light RSS: In our system, a set of traces is obtained based on a PDR algorithm. On the i-th trace $\mathbf{T}^i$ with the length of L, each pose corresponds to an array of visible light RSS, i.e.,

$$\mathbf{T}^i = \left\{ \left(x_1^i, y_1^i\right), \left(x_2^i, y_2^i\right), ..., \left(x_L^i, y_L^i\right) \right\}; \quad (1)$$

$$\mathbf{X}_l^i = \left(x_l^i, y_l^i\right) \rightarrow \mathbf{RSS}_l^i = \left\{ \mathrm{RSS}_{l,1}^i, \mathrm{RSS}_{l,2}^i, ..., \mathrm{RSS}_{l,K}^i \right\} \quad (2)$$

where $\mathbf{X}_l^i$ is the l-th pose on the i-th trace, $\mathbf{RSS}_l^i$ is the visible light RSS vector attached with $\mathbf{X}_l^i$, $\mathrm{RSS}_{l,k}^i$ denotes RSS detected from the LED with the k-th luminous frequency.

It is redundant to search all the poses for loop closures based on the visible light RSS, which will result in massive false positive matches. Hence, to reduce these mismatches, we remove the invalid visible light RSS on the poses with low discrimination and sheltered by user body. First, according to the path loss model of visible light RSS, the discrimination of RSS far from the effective coverage of LEDs is low. Hence, we remove the visible light RSS smaller than a threshold RSS_{min}. Second, if the difference between two neighboring RSS in one trace is higher than a threshold $\Delta\mathrm{RSS}_{max}$, the light sensor is considered to be sheltered by user body. Thus, these RSS is also removed.

2) Trace Clustering: After preprocessing of the traces and their attached visible light RSS, we cluster the traces based on the detected frequency of visible light to prepare for trace merging. For each trace, we calculate the mean RSS of each LED frequency over all poses and mark the top four LED frequencies with the largest mean RSS as their identification. The traces with the same identification are clustered into the same cluster.

3) Loop closure Detection and Validation: In this work, we propose a sequential pose matching algorithm via the similarity of visible light RSS on a sequence of poses. To design the sequential pose matching algorithm, we first define *matching pose*, i.e., a matching pose is determined by the similarity of the visible light RSS vectors between two traces ($\mathbf{T}^i$ and $\mathbf{T}^j$) as

$$E(\mathbf{RSS}_m^i, \mathbf{RSS}_n^j) = \frac{1}{K}\sum_{k=1}^{K} |\mathrm{RSS}_{m,k}^i - \mathrm{RSS}_{n,k}^j| \tag{3}$$

where $\mathbf{RSS}_m^i$ represents the visible light RSS corresponding to the m-th pose in the i-th trace. A pair of matching poses is determined if the distance in Equation (3) is shorter than a threshold of $E_{\min}$ by

$$\begin{cases} \mathbf{RSS}_m^i = \mathbf{RSS}_n^j, & E(\mathbf{RSS}_m^i, \mathbf{RSS}_n^j) \leqslant E_{\min} \\ \mathbf{RSS}_m^i \neq \mathbf{RSS}_n^j, & E(\mathbf{RSS}_m^i, \mathbf{RSS}_n^j) > E_{\min} \end{cases} \tag{4}$$

where "$=$" and "$\neq$" denotes matching and mismatching.

Traditional approaches for detecting loop closures are normally based on the individual pose matching. However, due to the noisy visible light, the individual pose matching approach is prone to large amount of false positive matches. Hence, to reduce the false positive matches, a pair of matching segments is defined as the segments with more than three sequential matching poses.

In a crowdsourcing scenario, the absolute direction of a trace is normally unavailable. Therefore, the sequential matching poses in a pair of matching segments need to be searched by a two-round algorithm respectively for the same and opposite directions. In the two-round sequential pose matching algorithm, we fix one trace $\mathbf{T}^i$ and slide the other one $\mathbf{T}^j$ to find out the sequential matching poses. Then, $\mathbf{T}^j$ is reversed and calculated with the same principle as the first round. Afterwards, we compare the length of segments with more than three sequential matching poses in this two round and keep the longer one.

To further remove the false positive matching segments, the matching segments are validated according to the following rules.

- **Best matching poses:** For each pair of the matching poses in the matching segments, we find the top w sequential matching poses with (3) and (4). In this work, we set $w = 3$.
- **Consistence of trace shape:** We calculate the turning angle at the middle pose in trace $\mathbf{T}^i$, i.e., θ_{mid}. If the difference of the turning angles in two matching segments is lager than a threshold θ_t, i.e., $\Delta\theta_{mid} > \theta_t$, these two candidates are removed.
- **Consistence of RSSI vector:** We only reserve the candidate matching segments with the consistent order of RSS values from different LEDs.

4) Multi-Tree Structure for Trace Ordering: Since multiple traces are potential to be matched, the order of matching traces can affect the positioning accuracy. Therefore, we propose a multi-tree structure to decide the merging benchmark and the merging order. In this multi-tree structure, the connections between parent and child nodes represent merging relationships obtained by the algorithm in Subsection III-A3. A root node as the merging benchmark and the matching order from parent nodes to child nodes are decided based on three metrics, i.e., the number of matching traces, the number of matching poses, and trace length. In these metrics, the number of matching traces is with the highest priority. We observe that the trace with more matching traces carries more valid RSS information which is closer to the ground truth. If more than one candidate traces have the same amount of the matching traces, longer traces and the traces with more matching poses are preferred. We perform a breadth-first traversal on the multi-tree structure and repeat this process until all the candidate traces enter the tree structure.

5) Trace merging with ICP algorithm: After determining the loop closures and the merging order, an enhanced ICP SLAM algorithm is adopted to merge the traces with loop closures based on the merging order. In the proposed ICP SLAM, the traces are considered as the rigid bodies without shape adjustment. The traces are merged by rotation $\mathbf{R}$ and translation $\mathbf{T}$ based on the shared loop closures. ICP SLAM [35] algorithm is primarily used to solve the data registration problem of point cloud on free form surface.

The ICP SLAM for the trace merging algorithm is taken for this optimization by translation and rotation. Since ICP SLAM may fail due to random initialization of locations and directions of traces, the selected trace is first translated by matching its center coordinate to its parent node in the multi-tree structure for trace ordering. The coordinates of the target pose set in the trace of $\mathbf{T}_i$ are defined as

$$K = \left\{ \mathbf{k}_i \left| \mathbf{k}_i \in R^2, i = 1, 2, \cdots N_k \right. \right\}, \tag{5}$$

and the coordinates of the reference pose set $\mathbf{T}_j$ as

$$M = \left\{ \mathbf{m}_i \left| \mathbf{m}_i \in R^2, i = 1, 2, \cdots N_m \right. \right\}. \tag{6}$$

A rotation matrix is defined as $\mathbf{R}$ and a translation matrix as $\mathbf{T}$. The pose coordinates are optimized by minimizing an error function associated with $\mathbf{R}$ and $\mathbf{T}$ as

$$E(\mathbf{R}, \mathbf{T}) = \frac{1}{N_k}\sum_{i=1}^{N_k} \|\mathbf{R}\mathbf{k}_i + \mathbf{T} - \mathbf{m}_i\|^2 \tag{7}$$

$$\mathbf{R}^*, \mathbf{T}^* = \arg\min_{\mathbf{R},\mathbf{T}} E(\mathbf{R}, \mathbf{T}) \tag{8}$$

The iterative calculation is performed until the error function is minimized to find $\mathbf{R}^*$ and $\mathbf{T}^*$ for each pair of matching traces. Based on the aforementioned ICP SLAM approach, the crowdsourcing traces with unknown relative directions and locations are sequentially merged under the merging order determined in Subsection III-A4.

B. Fine-Grained Trace Merging based on GraphSLAM Fusing Visible Light and Magnetic Field

The above ICP SLAM based on visible light merges the crowdsourcing traces by rotating and translating the traces as rigid body. However, due to the unalterable shape of traces, the traces are only coarse-grained merged. As the unknown relative directions and locations of the crowdsourcing traces have been compensated based on the ICP SLAM, a GraphSLAM can be further adopted to further fine-grained tune the poses in the merged traces. In this work, we propose a GraphSLAM approach based on visible light and magnetic field.

1) Loop Closure Detection based on Visible Light and Magnetic Field: In the stage of GraphSLAM, the loop closures are detected and validated based on visible light and magnetic field data to build the constraints among the poses, i.e., edges.

The loop closures detected by visible light RSS as introduced in Section III-A are only between the traces. Hence, in GraphSLAM, we further detect the loop closures inside each trace based on visible light with the same method introduced in Section III-A3. The aforementioned loop closures detected by visible light are accurate but the number of loop closures are normally less especially in the area without visible light coverage.

In those areas without visible light coverage, we take the sequential variation of magnetic field as features to increase the number of detected loop closures. The sequential variation of magnetic field on the same ground truth path is very similar. We leverage Dynamic Time Warping (DTW) to search the matched poses in the same trace cluster based on the magnitude of triaxial magnetometer data. DTW can find the segments of trace $\mathbf{T}_i$ that best matches $\mathbf{T}_j$. Moreover, it will compress or stretch one sequence to find the best match, which helps to mitigate speed changes as the candidates walk.

To further remove the false positive loop closures, we validate all the detected ones according to the following rules.

- The spatial distance between the matched poses must be less than $10m$. According to our investigation of ICP SLAM based on visible light RSS, the maximum merging accuracy is smaller than $8m$.
- The orientation difference between the matched poses must be less than 30 degrees. When the same location from the opposite directions is visited, the light RSS observations will be affected by body shade. Hence, only the matched pose locations from the same orientation are expected.

2) Backend of GraphSLAM: A Dynamic Bayesian Graph (DBG) with undirected edges among poses is adopted to represent the pose graph SLAM problem, in which edges are composed of PDR and loop closure constraints. The pose of a pedestrian at each moment is expressed as a state vector $\mathbf{x}_i = (x_i, y_i)^{\mathrm{T}}$ with a 2-Dimensional position. In this work, GraphSLAM considers multiple traces as a set of poses $X = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N\}$ with the pose number of N. To optimize the poses based on the edges, a robust SLAM backend is designed, in which a new formulation for the cost function based on the types of constraints is proposed. This method aims to mitigate the constraints associated with the remaining false loop closures during the optimization process.

The PDR constraints connect two successive pose states $\mathbf{x}_i$ and $\mathbf{x}_{i+1}$ via a motion model $f(\cdot,\cdot)$. We assume that the successive poses are the normally distributed random variables as

$$\mathbf{x}_{i+1} \sim \mathcal{N}(f(\mathbf{x}_i, \mathbf{s}_i), \Sigma_i) \tag{9}$$

where $\mathbf{s}_i$ is obtained from a PDR algorithm to denote the pose change between the i^{th} and $i+1^{th}$ pose, and Σ_i is the covariance. Note that the relative locations and directions of raw PDR traces in the motion model $f(\cdot,\cdot)$ have been corrected based on the aforementioned ICP SLAM.

The loop closure constraints obtained based on visible light and magnetic field connecting two poses $\mathbf{x}_i$ and $\mathbf{x}_j$ between two different traces follow a Gaussian distribution

$$\mathbf{x}_j \sim \mathcal{N}(f(\mathbf{x}_i, \mathbf{s}_{ij}), \Lambda_{ij}) \tag{10}$$

where Λ_{ij} indicates the certainty degree of the loop closures. The motion model function $f(\cdot,\cdot)$ is reused, since a constraint contains a pseudo-odometry measurement $\mathbf{s}_{ij}$ to express a displacement between $\mathbf{x}_i$ and $\mathbf{x}_j$. Note that Gaussian distribution is assumed because that there is no priori knowledge on the distributions of pose variables and a Gaussian distribution can simplify the calculation.

The two constraints $S = \{\mathbf{s}_i, \mathbf{s}_{ij}\}$ defined through the observations from different sensors are assumed to be independent. Therefore, the conditional probability over all variables and constraints is expressed by [36]

$$P(X\,|S\,) \propto \prod_i P\left(\mathbf{x}_{i+1}\,|\mathbf{x}_i, \mathbf{s}_i\,\right) \prod_{ij} P\left(\mathbf{x}_j\,|\mathbf{x}_i, \mathbf{s}_{ij}\,\right). \tag{11}$$

where $P\left(\mathbf{x}_{i+1}\,|\mathbf{x}_i, \mathbf{s}_i\,\right)$ and $P\left(\mathbf{x}_j\,|\mathbf{x}_i, \mathbf{s}_{ij}\,\right)$ are called factors to encode probabilistic constraints over the corresponding nodes. Based on the Gaussian assumption on two constrains, the joint pose distribution is expressed as

$$\begin{aligned} P(X\,|S\,) \propto & \prod_i \exp\left(-\frac{1}{2}\|f(\mathbf{x}_i, \mathbf{s}_i) - \mathbf{x}_{i+1}\|^2 \Sigma_i\right) \\ & \cdot \prod_{ij} \exp\left(-\frac{1}{2}\|f(\mathbf{x}_i, \mathbf{s}_{ij}) - \mathbf{x}_j\|^2 \Lambda_{ij}\right). \end{aligned} \tag{12}$$

The final cost function is calculated based on a negative logarithm operation on the joint pose distribution as

$$\begin{aligned} & \mathrm{F}(\mathbf{x}) = \mathrm{F}_{pdr}(\mathbf{x}) + \mathrm{F}_{loop}(\mathbf{x}) \\ & = \sum_i \|f(\mathbf{x}_i, \mathbf{s}_i) - \mathbf{x}_{i+1}\|^2 \Sigma_i + \sum_{ij} \|f(\mathbf{x}_i, \mathbf{s}_{ij}) - \mathbf{x}_j\|^2 \Lambda_{ij} \\ & = \sum_{(i,j)\in\Phi} \mathbf{e}_{ij}^{pdr}(\mathbf{x})^{\mathrm{T}} \mathbf{\Omega}_{ij}^{pdr} \mathbf{e}_{ij}^{pdr}(\mathbf{x}) + \sum_{(i,j)\in\Gamma} \mathbf{e}_{ij}^{loop}(\mathbf{x})^2 \mathbf{\Omega}_{ij}^{loop} \end{aligned} \tag{13}$$

where $\mathbf{\Omega}_{ij}^{pdr}$ is the information matrix set to a second-order unit diagonal matrix, $\mathbf{\Omega}_{ij}^{loop}$ represents the uncertainty of the loop closure constraint, Φ is the set of the PDR constraints, and Γ is the set of the loop closure constraints between pose pairs. Note that the loop closure constraints in Equation (13) are detected by visible light and magnetic field.

IV. Implementation of CrowdFusion Positioning System

According to Section III, we implement the CrowdFusion SLAM to construct a radio map based on the crowdsourcing traces. Based on the radio map, an enhanced particle filter fusing PDR, WiFi, visible light and floor plan in the CrowdFusion system is designed to locate the users with high accuracy.

A. Data Collection

1) Configuration of LED emitters: Micro-controllers are used to generate pulse width modulation (PWM) signals to control the specified positioning frequencies of LEDs. We control 10W LEDs with metal radiator panels as light sources in 6 different frequencies of 714Hz, 909Hz, 1250Hz, 1428Hz, 1666Hz, and 2500Hz. A trigger switch module is used to amplify PWM signal power so that the smartphones with light sensors can capture the LED light signals.

2) Data Acquisition: In the data acquisition, users use smartphones to collect visible light signals and inertial sensor data. However, the low scanning frequency (100Hz) of the light sensor equipped in an original smartphone cannot meet the requirement to distinguish these high luminous frequencies. We use an external light sensor to collect visible light signals, whose analog output is converted to a digital signal by the Analog-to-Digital Converter (ADC) in the microphone of smartphone with a sample rate of 8000Hz. Since 8000Hz is larger than twice of the maximum frequency (2500Hz), the Nyquist Sampling Theory is satisfied. More detailed implementation is referred to [33]. The resulted digital signal, a two-byte short value, is converted to an RSSI value by Fast Fourier Transform (FFT) having a positive linear correlation with LED intensity. Additionally, the inertial sensor data is sampled with 50Hz from accelerometer, gyroscope and magnetometer in the smartphone. WiFi RSSI are also collected with the smartphone. All data are marked with their corresponding timestamps to synchronize all the data from different sensors.

B. Trace Generation

1) Pedestrian dead reckoning: The trace information is derived from the inertial sensor data where the orientation and step count determine the direction and the length of the trace. Step detection is calculated based on a pair of peaks and valleys in the accelerometer data because human walking motion patterns are periodic. A linear model between the step length and frequency is used to estimate the step length.

For heading estimation, we design two PDR approaches, respectively for offline trace generation and sensor fusion positioning.

- For offline trace generation, we pay more attention to the accuracy of relative direction change between two adjacent poses instead of its absolute direction, because the absolute direction errors of traces can be further corrected in the following procedure of trace merging. Hence, for offline trace generation, heading direction estimation is obtained by integrating the angular velocity from gyroscope based on an initial angle as a reference. An X-AHRS filter is used to fuse data from gyroscope and accelerometer to estimate the Euler angles [37], in which data of magnetometer is not fused to protect the trace from magnetometer noise.
- For PDR in sensor fusion positioning, data of magnetometer is required to obtain absolute direction. To mitigate the influence of magnetometer noise, we leverage an adaptive β X-AHRS filter to fuse data from gyroscope, accelerometer, and magnetometer to estimate the Euler angles [37]. In the adaptive β X-AHRS filter, β is automatically tuned according to the surround magnetic environments. We refer to our previous work [37] for more details.

C. CrowdFusion Trace Merging

The crowdsourcing traces are merged by the proposed two-step trace merging procedures in Section III. According to [5], to anchor the merged traces in a floor plan and reduce the calibration cost, the locations of doors are used as anchors for mapping. To detect the doors, we find that the magnetic signals collected by the IMU have a high peak over space due to steel shells in most large buildings when a user passes a door. Thus we can detect the peaks of the magnetic signals at door locations. Finally, we can adjust the trace according to the ground-truth locations of these doors in the floor plan.

D. Radio Map Generation

The locations of the poses on the merged traces are used to label the measured signals to build up radio maps. In the radio map, we store two kinds of signals including visible light RSS and WiFi RSS.

- In those stored signals, the poses with the largest visible light RSS larger than a predefined threshold are stored as visible light fingerprints. The threshold is defined because the low visible light RSS at the edge of LED coverage is with low discrimination.
- Moreover, all the poses are attached with a vector of WiFi RSSIs as WiFi fingerprints.

E. Sensor Fusion Positioning based on Particle Filter

To further improve the positioning accuracy and solve the problem of limited coverage of a radio map, we take a particle filter to fuse a real-time PDR with the above fingerprinting for the final locations of users.

In a particle filter, a target moves according to a non-linear function, i.e., a system model $\mathbf{x}_k = \mathbf{f}_k(\mathbf{x}_{k-1}, \mathbf{v}_k)$. The measurement system observes the target according to another non-linear function, i.e., an observation model $\mathbf{z}_k = \mathbf{h}_k(\mathbf{x}_k, \mathbf{u}_k)$, where $\mathbf{v}_k$ and $\mathbf{u}_k$ are the system and measurement noise, respectively.

From Bayesian perspective, the goal is to calculate the posterior Probability Density Function (PDF) $p(\mathbf{x}_k|\mathbf{z}_{1:k})$, which is estimated by the following delta function

$$p(\mathbf{x}_k|\mathbf{z}_{1:k}) \approx \sum_{i=1}^{N_s} w_k^i \delta(\mathbf{x}_k - \mathbf{x}_k^i), \tag{14}$$

where $\mathbf{x}_k^i$ is the i-th particle and w_k^i is the *associated weight*. N_s is the total number of particles. For Bootstrap Particle Filter (BPF) [38], which is commonly used and efficiently implementable, the associated weights are calculated by

$$w_k^i \propto w_{k-1}^i \cdot p(\mathbf{z}_k|\mathbf{x}_k^i), \tag{15}$$

in which the associated weights are only determined by the *likelihood* function of $p(\mathbf{z}_k|\mathbf{x}_k^i)$.

1) Movement Constraint based on Floor Plan: Leveraging the knowledge of floor plan can remove the particles with unreasonable movements. In this work, we represent the environment with a grid of square cells, in which the size of a cell is $l = 0.25cm$. We label the cells with two markers, i.e., allowed areas and obstacles. The particles are only allowed to be located and move in the cells marked with allowed areas.

2) System Model based on Pedestrian Dead Reckoning: In the system, state $\mathbf{x}$ at step k is defined as the location of the target user as

$$\mathbf{x}_k = [x_k, y_k]. \tag{16}$$

In this work, the particles are updated based on the system model once a step is detected in the PDR algorithm. The system model is defined based on the step length l and the heading direction θ estimated in the PDR algorithm as

$$\mathbf{x}_k = \mathbf{F}\mathbf{x}_{k-1} + \mathbf{\Phi} + \boldsymbol{n} \tag{17}$$

where $\mathbf{x}_k$ is the center location in one cell defined in Subsection IV-E1, $\boldsymbol{n}$ is the Gaussian noise vector on the latitude and longitude, and

$$\mathbf{F} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \mathbf{\Phi} = \begin{bmatrix} l\cos\theta & 0 \\ 0 & l\sin\theta \end{bmatrix}.$$

Note that the particles only move when the connections between two cells for one step is not blocked by obstacles. Otherwise, the particles would stay at the original cells.

3) Observation Model based on Visible Light: The user locations estimated by visible light are used in the observations of particle filter. If the intensity of visible light is larger than a predefined threshold and the detected frequencies match the ones in radio map, the system leverages visible light fingerprinting as the observations.

We adopt a coordinate pair $(x^{\text{VL}}, y^{\text{VL}})$ to design $p(\mathbf{z}_k|\mathbf{x}_k^i)$ in Equation (15) for updating the associated weights in the particle filter, which is calculated by

$$p(\mathbf{z}_k|\mathbf{x}_k^i) = p(x_k^{\text{VL}}|x_k^i) \cdot p(y_k^{\text{VL}}|y_k^i) \tag{18}$$

where $p(x_k^{\text{VL}}|x_k^i)$ and $p(y_k^{\text{VL}}|y_k^i)$ follow the Gaussian distribution as

$$\begin{cases} p(x_k^{\text{VL}}|x_k) \sim \mathcal{N}(0, \sigma^2) \\ p(y_k^{\text{VL}}|y_k) \sim \mathcal{N}(0, \sigma^2) \end{cases} \tag{19}$$

where σ is the standard deviation of the Gaussian noise.

4) Observation Model Fusing WiFi Fingerprints: Because the coverage of visible light fingerprints is limited, if the intensity of visible light is smaller than a predefined threshold or the detected frequencies do not match any one in the radio map, the WiFi fingerprinting is additionally adopted as the observations in particle filter. Then,

$$p(\mathbf{z}_k|\mathbf{x}_k^i) = p(x_k^{\text{WiFi}}|x_k^i) \cdot p(y_k^{\text{WiFi}}|y_k^i) \tag{20}$$

where $p(x_k^{\text{WiFi}}|x_k^i)$ and $p(y_k^{\text{WiFi}}|y_k^i)$ follow the Gaussian distribution as

$$\begin{cases} p(x_k^{\text{WiFi}}|x_k) \sim \mathcal{N}(0, \sigma'^2) \\ p(y_k^{\text{WiFi}}|y_k) \sim \mathcal{N}(0, \sigma'^2) \end{cases} \tag{21}$$

where σ' is the standard deviation of the Gaussian noise.

5) Resampling and Reboot: Because the movement of particles is limited based on floor plan, the particle filter may face severe sample degeneracy problem, which may result in serious performance degradation. To deal with this problem, resampling is a typical solution. A suitable measure of degeneracy is the effective sample size $N_{\text{eff}} = 1/\sum_{i=1}^{N_s}(w_k^i)^2$ where N_s is the number of particles and w_k is the weight of the k-th particle . As soon as N_{eff} is smaller than half of N_s, the degeneracy is considered to be serious and a suitable resampling method should be adopted. A systematic resampling method is adopted in our work, because of its high accuracy and efficient implementation.

Moreover, if the particle cloud diverges too much from the visible light observations, resampling can not effectively update the particles to the correct locations. Hence, in this case, the system will directly reboot the particle filter from the locations obtained from the visible light fingerprinting.

V. Experiment and Evaluation

A. Experiment Setup

We conduct comprehensive experiments to evaluate the proposed trace merging and positioning approaches in the CrowdFusion system. The testing environment is a space of a building at Jilin University. LED lighting infrastructures are deployed in a 600m^2 space with 4 sub-regions and 16 LEDs shown in Fig. 2. Fig. 3 shows the deployment of LEDs in one sub-region. Note that the locations of LEDs are not required in the CrowdFusion system.

We collect sensor data from 10 users and store them in a center server for further processing, in which all the sensor data are timestamped for sensor fusion. Based on the PDR approach in Section IV-B1, data from accelerometer and gyroscope are fused to generate the traces. In this experiment, There are 31 recorded traces containing $1,016$ steps with lengths varying from 18 steps to 182 steps. These traces are used as input to the proposed CrowdFusion SLAM to construct the walking paths, which further label the WiFi and visible light RSS to construct a radio map of the target building. Moreover, after constructing the radio map, we further evaluate the positioning accuracy of the proposed CrowdFusion system.

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 moving path. All the sensor information including WiFi and visible light signal strength lists are timestamped and mapped to each ground truth position based on the time when the user passes.

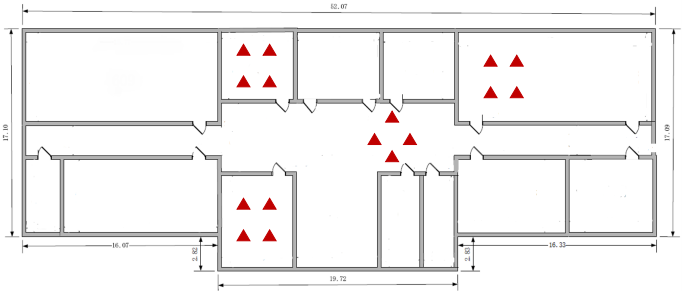

Fig. 2: Experiment Layout

Fig. 3: LED Setup in Office

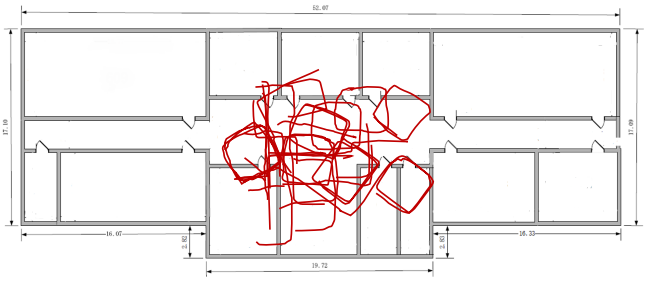

Fig. 4: Initial Traces

Fig. 5: ICP SLAM (Random Ordering)

Fig. 6: ICP SLAM

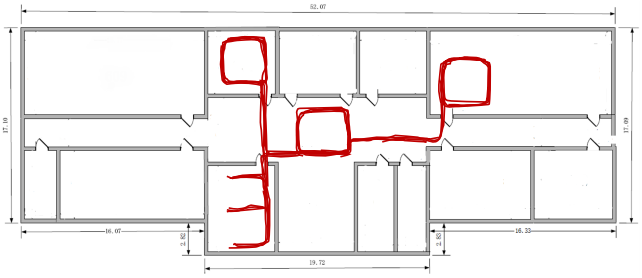

Fig. 7: ICP+GraphSLAM

B. Evaluation of CrowdFusion Trace Merging

1) Evaluation of Multi-Trace Merging based on ICP SLAM: Fig. 4 shows the original 31 traces with unknown relative locations and directions. A traditional GraphSLAM approach can not directly been applied to optimize the poses on these traces. Hence, to merge these traces, the ICP SLAM based on visible light is first applied to correct the unknown relative locations and directions among these traces.

TABLE I: Loop Closures based on Visible Light ICP SLAM

	Individual poses	Sequential poses	Matching poses after validation
No. of Poses	422	232	176
Precision	64%	89%	98%

Loop closure detection and validation: To merge the original traces, loop closures are detected based on visible light RSS. Table I summarizes the amount of loop closures based on individual matching poses, sequential matching poses and matching poses after validation, which are introduced in Subsection III-A3. The evaluation results in Table I shows that the quantity of loop closures detected based on the individual matching poses is large and a large number of the false positive ones are included. These false positive loop closures would severely degrade the performance of ICP SLAM. The loop closure detecting method based on sequential matching poses significantly reduces the amount of loop closures, especially the false positive ones. Moreover, the validation approaches in Subsection III-A3 further reduce the the false positive loop closures and the amount of remaining loop closures is 176, which is 246 less than that of the individual matching poses. Because false positive matching poses significantly affect the merging accuracy, we require high precision ($precision = TP/(TP+FP)$), where TP is True Positive and FP is False Positive. The precision of matching poses is given in Table I. The precision of sequential poses achieves 89% which is significantly higher than the individual ones with 64%. By conducting the validation approaches in Subsection III-A3, the precision further increases to 98%. Moreover, we evaluate the influence of two parameters (ω and θ_t) introduced in Section III-A3 on the precision of loop closure and mean accuracy of ICP SLAM based on visible light via Taguchi methods. The evaluation results in Table II show that the best combination of ω and θ_t are $\omega = 3$ and $\theta_t = 15°$. With larger ω which is the number of best matching poses, the precision gets lower and correspondingly the merging accuracy

TABLE II: Evaluation of Different Parameters in Matching Poses Validation

Best Matching Poses	Consistence of Shape	Precision	Merging Accuracy
$\omega = 3$	$\theta_t = 15^\circ$	98%	$0.84m$
$\omega = 4$	$\theta_t = 15^\circ$	97%	$0.92m$
$\omega = 5$	$\theta_t = 15^\circ$	97%	$0.93m$
$\omega = 3$	$\theta_t = 5^\circ$	98%	$1.01m$
$\omega = 3$	$\theta_t = 25^\circ$	96%	$0.95m$

degrades. If θ_t indicating the shape of matching poses is too low such as 5°, the precision is still high (98%) but many true positive matching poses are removed. Hence, with $\theta_t = 5^\circ$, the merging accuracy degrades to $1.01m$. With larger θ_t such as 25°, the precision gets lower and correspondingly the merging accuracy degrades. Hence, $\omega = 3$ and $\theta_t = 15^\circ$ are selected in the following evaluation.

Trace ordering: The trace order in ICP SLAM affects the merging accuracy. In Subsection III-A4, we propose a multi-tree structure trace ordering approach based on three metrics, i.e., the number of matching traces, the number of matching poses, and the trace length. Fig. 5 and 6 respectively represent the merged traces based on a random order and the multi-tree structure trace order, showing that the proposed multi-tree structure trace ordering approach can effectively improve the merging accuracy. In the procedure of trace merging, we find that if the parent traces introduced in Section III-A4, especially the benchmark one, are wrongly selected, the child traces would be difficult to accurately merged. Especially, if a short trace is selected as the benchmark (root) trace, the performance will significantly deteriorate. Hence, the proposed multi-tree structure trace ordering approach is fundamental for trace merging.

TABLE III: Accuracy of Trace Merging

	Mean	Median	90%
ICP+GraphSLAM (VL+Mag)	0.67m	0.65m	1.11m
ICP+GraphSLAM (VL)	0.74m	0.64m	1.26m
ICP SLAM (Trace Ordering)	0.84m	0.72m	1.61m
ICP SLAM (Random Ordering)	1.04m	0.76m	1.92m

Fig. 8 and Table III summarize the Cumulative Density Function (CDF) and statistics of trace merging errors based on different approaches. The evaluation results show that the trace merging based on ICP SLAM with multi-tree structure trace ordering achieves a mean accuracy of $0.84m$ outperforming the one with the random ordering ($1.04m$) by 16.3%.

ICP SLAM merging: As shown in Fig. 6, the original traces are merged by using the proposed ICP SLAM, in which the relative locations and directions among traces are effectively corrected. However, because each trace is considered as a rigid body, the relative locations of the poses inside each trace can not be optimized by ICP SLAM.

2) Evaluation of Fine-Grained Trace Merging based on GraphSLAM: Since the relative locations and directions among traces have been corrected by ICP SLAM, the poses inside traces are further fine-grained optimized by GraphSLAM based on visible light and magnetic field. Because the traces are merged by rotation and translation in ICP SLAM as rigid bodies, more than three loop closures on one trace will be enough for merging. In GraphSLAM, each pose is fine-grained tuned based on the loop closures. Hence, visible light with limited coverage may not be enough to obtain sufficient loop closures. In this evaluation, we compare the trace merging based on GraphSLAM with and without fusing magnetic field.

Trace merging merely based on visible light: According to the evaluation results in Fig. 8 and Table III, the fine-grained traces based on GraphSLAM are more accurate than those coarse-grained ones only with ICP SLAM, because the relative locations of the poses in each trace are optimized under the constraints of the loop closures and PDR. Considering 90% accuracy, the combination of ICP SLAM and GraphSLAM based on visible light achieves $1.26m$, which is 21.7% higher than that of only using ICP SLAM ($1.61m$). However, visible light only provides a limited coverage in the defined four sub-regions covered by LEDs. Hence, the optimized poses by GraphSLAM still do not well behave in the area without LED covering.

Trace merging based on visible light and magnetic field: To further optimize the poses outside the coverage of LEDs in GraphSLAM, we fuse the loop closures detected by magnetic field based on DTW matching introduced in Section III-B1. Fig. 7 shows the final merged traces through the combination of ICP SLAM and GraphSLAM based on visible light and magnetic field. According to the evaluation results in Fig. 8 and Table III, the CrowdFusion approach achieves a 90% accuracy of $1.11m$, which outperforms the one with only taking the visible light by 12%. Hence, by fusing magnetic field data into the visible light data, the accuracy of the merged traces are further improved, especially in the areas without coverage of LEDs.

In CrowdFusion, the merged traces after applying the proposed two-step trace merging approach with ICP SLAM and GraphSLAM by fusing visible light and magnetic field achieve a mean accuracy of $0.67m$ and a median accuracy of $0.65m$, which is fundamental for achieving a sub-meter accuracy of crowdsourcing positioning system.

TABLE IV: Accuracy of CrowdFusion Positioning

	Mean	Median	90%	Success rate
Light	0.71m	0.67m	1.31m	46%
PDR+light	0.98m	0.76m	1.81m	100%
PDR+light+WiFi	0.88m	0.73m	1.62m	100%
WiFi	2.0m	1.62m	3.81m	100%
PDR+light+ WiFi+FloorPlan	0.77m	0.71m	1.39m	100%
PDR+WiFi+ FloorPlan	1.73m	1.35m	2.94m	100%

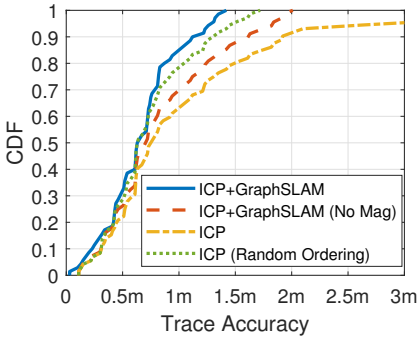

Fig. 8: Accuracy of Merged Traces

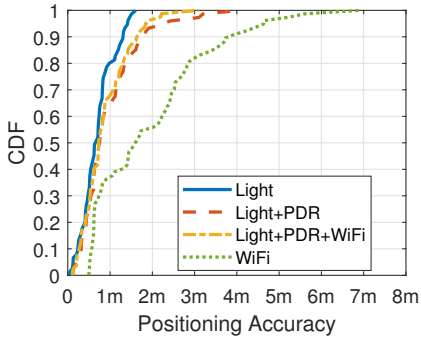

Fig. 9: Positioning without Floor Plan

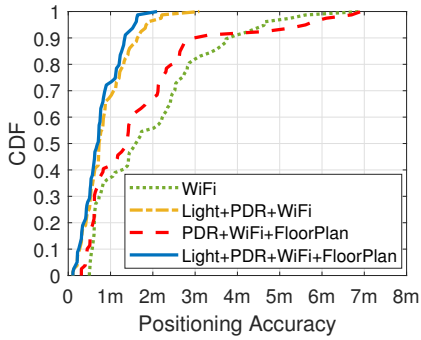

Fig. 10: Positioning with Floor Plan

C. Accuracy of CrowdFusion Positioning

1) Comparison of WiFi and Visible Light Fingerprinting: Based on the merged traces in CrowdFusion as the location labels, a radio map with visible light and WiFi fingerprints is constructed. Based on this radio map, we compare the accuracy of the visible light and WiFi fingerprinting. From Fig. 9 and Table IV, we find that the accuracy of the visible light fingerprinting significantly outperforms WiFi fingerprinting. The mean accuracy of the visible light fingerprinting achieves $0.71m$ and 90% positioning errors are lower than $1.31m$. In contrast, the WiFi fingerprinting only achieves a mean accuracy of $2.0m$ and a 90% accuracy of $3.81m$. Although the accuracy of the visible light fingerprinting is significantly higher than that of WiFi fingerprinting, the positioning coverage by visible light is only 46% because visible light can not penetrate walls and the discrimination of the visible light fingerprints is low at the edge of visible light coverage.

2) Comparison of PDR, Visible Light Fusion with and without Fusing WiFi Fingerprinting: To increase the coverage of the visible light fingerprinting, we evaluate two fusion approaches with particle filters, i.e., with and without fusing WiFi. In Fig. 9 and Table IV, the coverage of the fusion approaches achieves 100%. However, the mean accuracy of the fusion approaches is lower than that of visible light fingerprinting, because only PDR and WiFi positioning are used outside the coverage of visible light fingerprinting. In these two fusion approaches, the mean accuracy of the particle filter fusing visible light, WiFi and PDR achieves $0.88m$, which is better than the one only fusing visible light and PDR with a mean accuracy of $0.98m$, because the WiFi fingerprinting can correct the accumulated errors of PDR outside the coverage of visible light.

3) Comparison of Different Fusion Approaches with Floor Plan: By fusing the floor plan, the particle movement is limited by the floor plan and the positioning accuracy is improved. We compare the fusing approaches with and without visible light when using the particle filter to fuse the floor plan. From Fig. 10 and Table IV, it is shown that the fusion approach of PDR, WiFi, visible light and floor plan achieves a mean accuracy of $0.77m$, which outperforms the fusion approach without the floor plan by $0.11m$.

From the evaluation results in Fig. 10 and Table IV, we find that the approach (with floor plan) fusing the visible light fingerprinting with a mean accuracy of $0.77m$ is much higher than the one without the visible light of $1.73m$. On one hand, the accuracy of the visible light fingerprinting is much higher than that of the WiFi fingerprinting. Hence, fusing visible light as the observation in the particle filter of Equation (19) can improve the accuracy. On the other hand, a large number of particles in the particle filter fusing WiFi, PDR and floor plan are often locked by the walls when users pass doors. It takes a long time for the particle filter only fusing the WiFi data as observation to detect the sample degeneracy problem, which degrades the accuracy. In CrowdFusion positioning, since a resampling and rebooting approach with the visible light fingerprinting is adopted, the particle filter is more frequently resampled and rebooted when the visible light fingerprints are detected. Hence, the sample degeneracy problem is more efficiently solved which results in a higher accuracy.

VI. Conclusion

In this work, we propose a crowdsourcing system, namely CrowdFusion, with sub-meter positioning accuracy and zero-effort calibration. In this system, we develop a two-step trace merging approach on the basis of ICP SLAM and GraphSLAM by fusing visible light and magnetic field data. In the CrowdFusion trace merging approach, ICP SLAM is proposed to correct the relative directions and locations among traces. In ICP SLAM, we give a sequential pose matching approach to reduce the false positive loop closures and a multi-tree structure trace ordering approach to improve the merging accuracy. After ICP SLAM, the proposed GraphSLAM optimizes the poses inside the merged traces. Based on these merged traces, an enhanced particle filter fusing PDR, WiFi, visible light and floor plan is designed in CrowdFusion system to achieve a sub-meter positioning accuracy. According to a set of comprehensive experiments, the proposed CrowdFusion system achieves a mean positioning accuracy of $0.67m$ for the merged traces. Based on the constructed radio map labeled by the merged traces, the CrowdFusion system finally reaches a positioning accuracy of $0.77m$.

References

[1] Yuan, Xu, Choon, Ki, Ahn, Yuriy, S. Shmaliy, Xiyuan, Chen, and Lili, "Indoor ins/uwb-based human localization with missing data utilizing predictive ufir filtering," *IEEE/CAA Journal of Automatica Sinica*, vol. v.6, no. 04, pp. 91–99, 2019.

[2] W. Li, Z. Chen, X. Gao, W. Liu, and J. Wang, "Multimodel framework for indoor localization under mobile edge computing environment," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 4844–4853, 2019.

[3] Y. Zhang, S. Zhang, R. Li, D. Guo, Y. Wei, and Y. Sun, "Wifi fingerprint positioning based on clustering in mobile crowdsourcing system," in *2017 12th International Conference on Computer Science and Education (ICCSE)*, pp. 252–256, 2017.

[4] X. Chen, A. Peng, and B. Tang, "Automatic radio map adaptation for wifi fingerprint positioning systems," in *2020 5th International Conference on Communication, Image and Signal Processing (CCISP)*, pp. 64–69, 2020.

[5] Y. Yue, X. Zhao, and Z. Li, "Enhanced and facilitated indoor positioning by visible-light graphslam technique," *IEEE Internet of Things Journal*, vol. PP, no. 99, 2020.

[6] G. Shen, C. Zhuo, P. Zhang, T. Moscibroda, and Y. Zhang, "Walkie-markie: indoor pathway mapping made easy," in *Proceedings of the 10th USENIX conference on Networked Systems Design and Implementation*, 2013.

[7] 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, 2018.

[8] A. Yassin, Y. Nasser, M. Awad, A. Aldubai, 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. 99, pp. 1327–1346, 2017.

[9] H. Wang, C. Wang, and L. Xie, "Intensity-slam: Intensity assisted localization and mapping for large scale environment," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 1715–1721, 2021.

[10] H. Zhang, L. Jin, and C. Ye, "An rgb-d camera based visual positioning system for assistive navigation by a robotic navigation aid," *IEEE/CAA Journal of Automatica Sinica*, vol. 8, no. 8, pp. 1389–1400, 2021.

[11] M. Brossard, S. Bonnabel, and A. Barrau, "Invariant kalman filtering for visual inertial slam," in *2018 21st International Conference on Information Fusion (FUSION)*, pp. 2021–2028, 2018.

[12] D. Zhu, X. Sun, L. Wang, B. Liu, and K. Ji, "Mobile robot slam algorithm based on improved firefly particle filter," in *2019 International Conference on Robots Intelligent System (ICRIS)*, pp. 35–38, 2019.

[13] H. Gao, X. Zhang, J. Wen, J. Yuan, and Y. Fang, "Autonomous indoor exploration via polygon map construction and graph-based slam using directional endpoint features," *IEEE Transactions on Automation Science and Engineering*, vol. 16, no. 4, pp. 1531–1542, 2019.

[14] C. Laoudias, A. Moreira, S. Kim, S. Lee, L. Wirola, and C. Fischione, "A survey of enabling technologies for network localization, tracking, and navigation," *IEEE Communications Surveys Tutorials*, vol. 20, no. 4, pp. 3607–3644, 2018.

[15] R. Liu, S. H. Marakkalage, M. Padmal, T. Shaganan, C. Yuen, Y. L. Guan, and U. Tan, "Collaborative slam based on wifi fingerprint similarity and motion information," *IEEE Internet of Things Journal*, vol. 7, no. 3, pp. 1826–1840, 2020.

[16] S. Wang, H. Wen, R. Clark, and N. Trigoni, "Keyframe based large-scale indoor localisation using geomagnetic field and motion pattern," in *IEEE/RSJ International Conference on Intelligent Robots & Systems*, 2016.

[17] Q. Liang and M. Liu, "An automatic site survey approach for indoor localization using a smartphone," *IEEE Transactions on Automation Science and Engineering*, vol. 17, no. 1, pp. 191–206, 2020.

[18] E. Leitinger, F. Meyer, F. Hlawatsch, K. Witrisal, F. Tufvesson, and M. Z. Win, "A belief propagation algorithm for multipath-based slam," *IEEE Transactions on Wireless Communications*, vol. 18, no. 12, pp. 5613–5629, 2019.

[19] Y. Ding, Z. Xiong, J. Xiong, Z. Cao, and W. Li, "Distributed collaborative pedestrian inertial slam with unknown initial relative poses," *IEEE Internet of Things Journal*, vol. 9, no. 21, pp. 21632–21647, 2022.

[20] V. Kachurka, B. Rault, F. I. Ireta Muñoz, D. Roussel, F. Bonardi, J.-Y. Didier, H. Hadj-Abdelkader, S. Bouchafa, P. Alliez, and M. Robin, "Weco-slam: Wearable cooperative slam system for real-time indoor localization under challenging conditions," *IEEE Sensors Journal*, vol. 22, no. 6, pp. 5122–5132, 2022.

[21] A. Rai, K. K. Chintalapudi, V. N. Padmanabhan, and R. Sen, "Zee: Zero-effort crowdsourcing for indoor localization," pp. 293–304, 2012.

[22] Z. Yang, C. Wu, and Y. Liu, "Locating in fingerprint space: Wireless indoor localization with little human intervention," pp. 269–280, 2012.

[23] C. Luo, H. Hong, and M. C. Chan, "Piloc: A self-calibrating participatory indoor localization system," in *IPSN-14 Proceedings of the 13th International Symposium on Information Processing in Sensor Networks*, 2014.

[24] D. Han, S. Lee, and S. Kim, "Kailos: Kaist indoor locating system," in *International Conference on Indoor Positioning & Indoor Navigation*, 2015.

[25] X. Tong, K. Liu, X. Tian, L. Fu, and X. Wang, "Fineloc: A fine-grained self-calibrating wireless indoor localization system," *IEEE Transactions on Mobile Computing*, vol. PP, pp. 1–1, 2018.

[26] 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 of Things Journal*, vol. 6, no. 2, pp. 3585–3600, 2019.

[27] Z. Li, X. Zhao, F. Hu, Z. Zhao, J. L. C. Villacres, and T. Braun, "Soicp: A seamless outdoor–indoor crowdsensing positioning system," *Internet of Things Journal, IEEE*, vol. 6, no. 5, pp. 8626–8644, 2019.

[28] M. F. Keskin, S. Gezici, and O. Arikan, "Direct and two-step positioning in visible light systems," *IEEE Transactions on Communications*, vol. 66, no. 1, pp. 239–254, 2018.

[29] M. F. Keskin and S. Gezici, "Comparative theoretical analysis of distance estimation in visible light positioning systems," *Journal of Lightwave Technology*, vol. 34, no. 3, pp. 854–865, 2016.

[30] P. Wang and Y. J. Morton, "Multipath estimating delay lock loop for lte signal toa estimation in indoor and urban environments," *IEEE Transactions on Wireless Communications*, vol. 19, no. 8, pp. 5518–5530, 2020.

[31] Z. Sheng, W.-D. Zhong, D. Pengfei, C. Chen, and D. Wu, "Pdoa based indoor visible light positioning system without local oscillators in receiver," in *2017 Conference on Lasers and Electro-Optics Pacific Rim (CLEO-PR)*, pp. 1–3, 2017.

[32] Z. Zhu, B. Zhu, Y. Wang, Z. Shi, and Y. Jiang, "Angle-of-arrival estimator for light signals based on optimized photodiode array," in *2019 International Conference on Computing, Networking and Communications (ICNC)*, pp. 62–66, 2019.

[33] L. Li, P. Hu, C. Peng, G. Shen, and F. Zhao, "Epsilon: A visible light based positioning system," *11th USENIX Symposium on Networked Systems Design and Implementation (NSDI 14)*, pp. 331–343, 01 2014.

[34] Z. Wang, Z. Yang, Q. Huang, L. Yang, and Q. Zhang, "Als-p: Light weight visible light positioning via ambient light sensor," in *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications*, 2019.

[35] 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, 1992.

[36] N. Sünderhauf and P. Protzel, "Towards a robust back-end for pose graph slam," in *2012 IEEE International Conference on Robotics and Automation*, pp. 1254–1261, IEEE, 2012.

[37] S. O. H. Madgwick, A. J. L. Harrison, and R. Vaidyanathan, "Estimation of imu and marg orientation using a gradient descent algorithm," in *2011 IEEE International Conference on Rehabilitation Robotics*, pp. 1–7, 2011.

[38] Z. Li and T. Braun, "Passively track wifi users with an enhanced particle filter using power-based ranging," *IEEE Transactions on Wireless Communications*, vol. 16, no. 11, pp. 7305–7318, 2017.