ISPRS Journal of Photogrammetry and Remote Sensing 177 (2021) 131–146

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Crowdsourcing-based indoor mapping using smartphones: A survey

Baoding Zhou [a,b], Wei Ma [c], Qingquan Li [a,d,e,f,*], Naser El-Sheimy [g], Qingzhou Mao [h], You Li [i], Fuqiang Gu [j], Lian Huang [k], Jiasong Zhu [a,b]

[a] College of Civil and Transportation Engineering, Shenzhen University, Shenzhen 518060, China
[b] Institute of Urban Smart Transportation & Safety Maintenance, Shenzhen University, Shenzhen 518060, China
[c] School of Civil Engineering, Chongqing Jiaotong University, Chongqing 400074, China
[d] Guangdong Key Laboratory of Urban Informatics, Shenzhen University, Shenzhen 518060, China
[e] MNR Key Laboratory for Geo-Environmental Monitoring of Great Bay Area, Shenzhen University, Shenzhen 518060, China
[f] Shenzhen Key Laboratory of Spatial Smart Sensing and Services, Shenzhen University, Shenzhen 518060, China
[g] Department of Geomatics Engineering, University of Calgary, Calgary, AB T2N 1N4, Canada
[h] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430072, China
[i] State Key Laboratory of Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430072, China
[j] College of Computer Science, Chongqing University, Chongqing 400044, China
[k] Shenzhen Escope Intelligence Technology Co. LTD, Shenzhen, China

ARTICLE INFO

Keywords:
Indoor mapping
Crowdsourcing
Smartphone
Survey

ABSTRACT

Indoor map is a fundamental element of indoor location-based services (ILBS). However, traditional indoor mapping techniques are labor-intensive and time-consuming. The advancement of smartphones offers great opportunities for crowdsourcing-based indoor mapping, which is one of the most promising applications due to its low cost and flexibility. Over the last decade, many crowdsourcing-based indoor mapping solutions using smartphones have been proposed. This article provides a systematic review of these works. Different from former surveys, we classify the indoor mapping process by the stage of map construction. In particular, we highlight the two key steps, geospatial-element acquisition, and indoor-map construction, and provide state-of-the-art techniques on these topics. Then, we systematically review the crowdsourcing-based indoor mapping solutions under grid-based, landmark-based, and semantic maps. In addition to covering the principles, benefits, and challenges, these systems are compared in terms of sensors, participation, output, experimental environment, and reported accuracy. Besides these existing performance criteria, we extract quantitative performance criteria that are suitable to evaluate crowdsourcing-based indoor mapping solutions. Finally, we present open issues and future research directions.

1. Introduction

Indoor location-based service (ILBS) is a hot research topic in the geographic information science arena and has attracted the attention of both industry and academia (Kang et al., 2020; Liu et al., 2021). It has many prospective applications in smart cities, intelligent transportation, and logistics management fields (Ma et al., 2020). The ILBS market value is estimated to be $10 billion in 2020 (Connolly and Boone, 2013) . Indoor maps play an important role in ILBS applications because an indoor navigation system needs a map to show the navigation path and user location. Traditional indoor map construction relies on manual operation according to geometric information of blueprints in computer-aided design (CAD) (Gilliéron and Merminod, 2003; Han et al., 2014), industry foundation classes (IFCs) (Lin et al., 2013; Liu et al., 2014), and building information modeling (BIM) (Isikdag et al.,2013; Volk et al., 2014) formats. However, the manual process is labor-intensive, time-consuming, and expensive. In particular, it requires a substantial effort to keep the maps up-to-date since the manual process has to be repeated to capture environmental changes (Pipelidis et al., 2017). Moreover, building blueprints are usually inaccessible due to privacy issues.

In recent years, many automatic indoor mapping techniques based on special sensors (e.g., laser-based (Karam et al., 2019; Surmann et al., 2003; Turner et al., 2015), depth camera-based (Endres et al., 2014; Khoshelham and Elberink, 2012; Li et al., 2019), sonar-based (Ismail and Balachandran, 2015; Tardós et al., 2002), and multi-sensor fusion-based (Luo and Lai, 2012)) have been proposed to produce high-quality

* Corresponding author at: College of Civil and Transportation Engineering, Shenzhen University, Shenzhen 518060, China.
E-mail address: liqq@szu.edu.cn (Q. Li).

https://doi.org/10.1016/j.isprsjprs.2021.05.006
Received 2 September 2020; Received in revised form 20 April 2021; Accepted 8 May 2021

indoor maps. These sensors are equipped on a robot (Luo and Lai, 2012) or a backpack (Lauterbach et al., 2015; Wen et al., 2016)platform. The robot platform can save manpower but may encounter difficulties while running in complicated multi-floor indoor environments. In contrast, the backpack platform is more flexible; however, its surveying process is labor-intensive and time-consuming. In general, indoor mapping platforms are costly because of the requirement for professional hardware and expert handlers.

With the rapid development of microelectromechanical systems (MEMS) and mobile communications, smartphones become powerful in sensing, computing, and communication (Lane et al., 2010; Wang et al., 2016a; Yürür et al., 2014). In particular, today's smartphones are packed with many sensors (e.g., accelerometers, gyroscopes, magnetometers, cameras, WiFi, Bluetooth, and microphone) for environmental perception. The sensed information can reflect the geometric structure of the indoor map because the users' trajectories are restricted by obstacles (e.g., walls and furniture) indoors (Alzantot et al., 2012). For example, WiFi signals can be used to identify landmark similarities (Shen et al., 2013), while image and acoustic signals can be used to measure distances (Gao et al., 2014; Zhou et al., 2017a). The powerful sensing capacity of smartphones makes indoor-map construction possible. Furthermore, smartphone-based indoor mapping has two advantages: (i) it is low-cost since it does not require any additional equipment, and (ii) it is straightforward to implement because of the high penetration of smartphones.

The advancement of smartphones facilitates the development of crowdsourcing-based applications, such as remote sensing (Toth and Jozkow, 2016; Zhang et al., 2020), indoor localization (Lashkari et al., 2019; Rai et al., 2012; Wu et al., 2015; Zhou et al., 2017b), urban traffic management (Wang et al., 2016b; Zang et al., 2018), road navigation (Fan et al., 2017), and digital map updating (Peng et al., 2018; Tang et al., 2016; Wang et al., 2013; Zhou et al., 2021). Crowdsourcing has emerged as a hot topic for collecting and sharing sensing data (Ganti et al., 2011; Guo et al., 2015; Guo et al., 2017; Li et al., 2017a; Ma et al., 2014). In particular, crowdsourcing-based indoor-map construction is promising because many smartphone-based indoor sensing data can be generated when people spend the majority of their time indoors (Zhou et al., 2015a).

As a new sensing paradigm, crowdsourcing has gained increasing attraction. A large number of papers have been published. With these publications, there are several survey papers on the mobile crowdsourcing technique. The survey in (Heipke, 2010) introduces crowdsourcing geospatial data and reviews the developments. The survey in (Guo et al., 2015) characterizes the unique features, existing application areas, and reference framework of mobile crowdsourcing and computing systems, while the review (Hosseini et al., 2015) characterizes a systematic mapping study for crowdsourcing methods. Meanwhile, there are survey papers on specific crowdsourcing applications. The survey in (Lashkari et al., 2019) concerns crowdsourcing and sensing for indoor localization, while the survey in (Guo et al., 2017) proposes the concepts, unique features, application areas, challenges, and key techniques of visual crowdsourcing. Meanwhile, the paper (Wang et al., 2016b) reviews crowdsourcing applications in intelligent transportation systems. These survey papers have provided comprehensive information about the crowdsourcing technique and its applications in indoor localization, visual navigation, and intelligent transportation. However, the existing surveys do not have a systematic review of crowdsourcing-based indoor mapping. A review of these technologies from crowdsourced smartphone sensor data is therefore warranted. In this survey, we summarize almost 8 years of research and development from 2012 to 2020 in the field of crowdsourcing-based indoor mapping using smartphones (see Table 1).

The contributions in this paper are as follows:

- This article outlines a systematic review of crowdsourcing-based indoor-map construction using smartphones. Different from former surveys, we classify the indoor mapping process by the stage of map construction, which is clearer to understand. From this article, readers can directly get the full knowledge of how to obtain indoor maps by using crowdsourced smartphone sensor data.
- In the summarized process, we highlight the two key steps: geospatial-element acquisition and indoor-map construction. The state-of-the-art techniques in trajectory tracking, range estimation, landmark detection, and the construction of grid-based, landmark-based, and semantic maps are covered.
- We systematically introduce and explain the existing crowdsourcing-based indoor mapping systems. Besides the principles, benefits, and challenges, these systems are compared in terms of sensors, participation, output, experimental environment, and reported accuracy. This part can guide how to design or select a crowdsourcing-based indoor mapping system.
- Based on the summarization of existing works, we extract performance criteria that are especially suitable to evaluate crowdsourcing-based indoor mapping solutions. Examples of these criteria are the amount of data, the position of feature points, the graph/shape discrepancy metric, the room aspect ratio, and the indoor semantic information. The quantitative calculation of these criteria is provided.
- We emphasize the open issues and future directions in crowdsourcing-based indoor mapping. In particular, we point out that the lack of uniform comparison criteria, datasets, and incentive mechanisms are the main factors that have limited the promotion of crowdsourcing-based indoor mapping. Such issues have not been covered by previous surveys.

Table 1
Comparison of different indoor map construction.

	Occupancy grid map	Landmark-based map	Semantic map
Techniques	position clustering	landmark recognition	semantic recognition
Algorithm	convex hull and alpha-shape	spring-relaxation and multidimensional scaling	Faster R-CNN, SSD, and R-FCN
Cost	low	high	high
Advantage	low-cost, no need for multiple sensors	map constructed completely	semantic information
Disadvantage	difficult to shape room and hallway	algorithm complicated, multi-sensor required	algorithm complicated, multi-sensor required, depends heavily on the visual conditions of the scene

This survey is organized as follows. Section 2 provides an overview of the crowdsourcing-based indoor mapping process. Meanwhile, it highlights two main steps: geospatial-element acquisition and indoor-map construction. The techniques for these steps are reviewed in Sections 3 and 4, respectively. Afterward, Section 5 compares the techniques, while Section 6 illustrates the open research issues and future directions. Finally, Section 7 concludes the paper.

2. Overview of crowdsourcing-based indoor mapping

Indoor mapping is the process that transforms sensor data into indoor maps. This section describes the fundamentals of indoor maps, followed by an overall workflow of the crowdsourcing-based indoor mapping process.

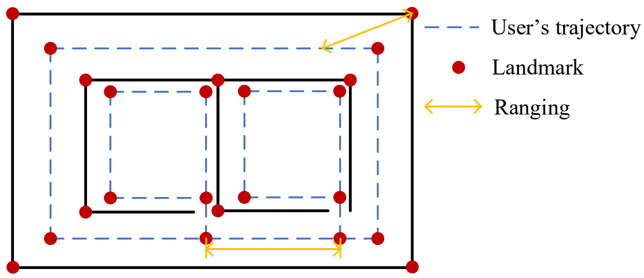

Fig. 1. Landmark-based map.

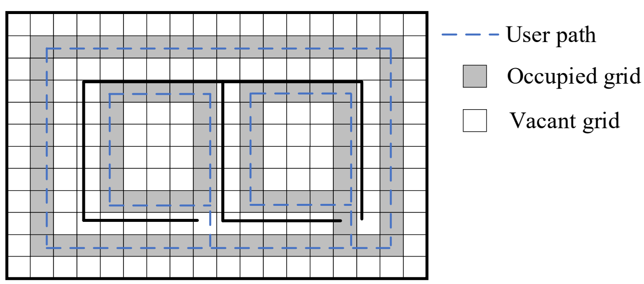

Fig. 2. Occupancy grid map.

2.1. *The data structure of indoor maps*

Generally, there are two data structures of the geometric map: vector maps and raster maps (Mao et al., 2015). For indoor maps, there are also two predominant paradigms: landmark-based maps and occupancy grid maps (Cadena et al., 2016; Yu and Amigoni, 2014). The former is a vector map, which models the environment as a sparse set of landmarks, and the latter is a raster map, which discretizes the environment in cells and assigns a probability of occupation to each cell.

Landmark-based maps model the environment as a sparse set of landmarks, as shown in Fig. 1. The solid lines represent walls in the map, and the dotted line is the user's trajectory. The locations of landmarks are estimated in the indoor mapping process as follows. First, the user's trajectories are inferred, and the ranging distances that between the path and the landmarks or between landmarks are estimated. Then, based on the user's trajectory and ranging distance, the locations of landmarks can be calculated.

To construct a landmark-based map, we obtain the location of the user path and the ranging distances between the path and the landmarks. In addition to constructing indoor maps, landmarks can also be used to assist path inference.

The occupancy grid map discretizes the environment in cells and assigns a probability of occupation to each cell. The probability of occupation is determined by occupancy detection. User path-based occupancy detection is a common method. The detection mechanism works as follows: if a user path passes the cell, it is detected as an occupied grid, as shown in Fig. 2. Besides, occupancy detection can be employed based on wireless signals (Qiu and Mutka, 2017). However, the detection effect of a wireless signal tends to decrease sharply when an obstacle exists between the wireless links.

Besides geometric construction, semantic information is the other important element for electronic maps. It offers the attributes (e.g., names and functionalities) of objects in an indoor environment. The semantic information acquisition is usually by extracting texts from images.

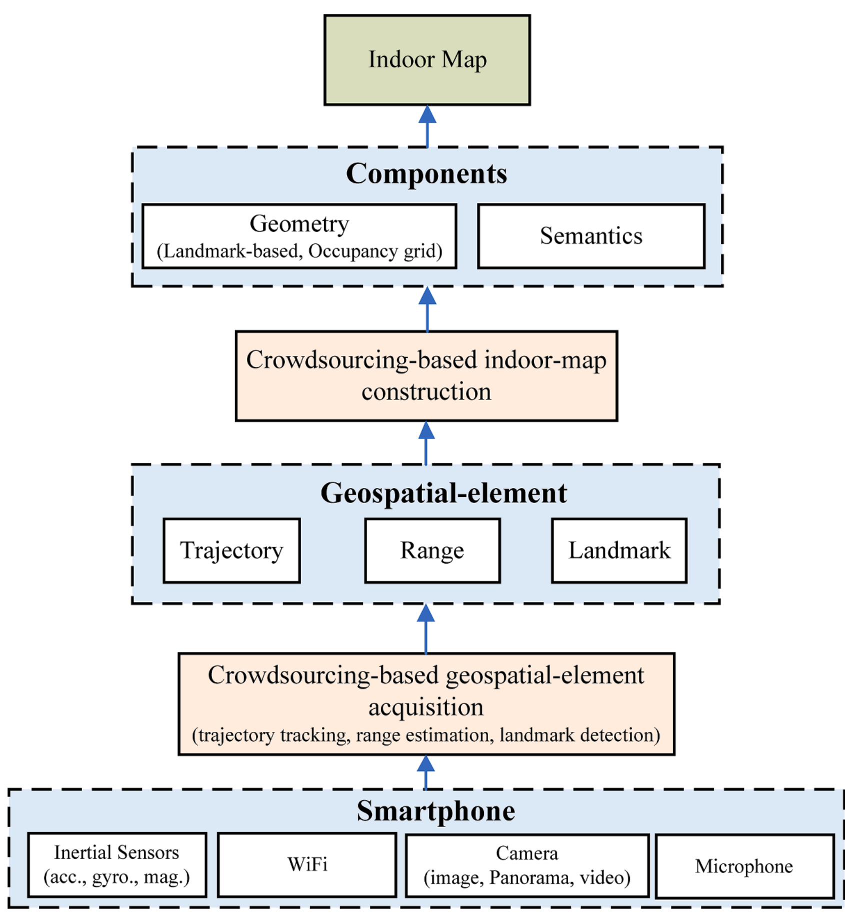

Fig. 3. Crowdsourcing-based indoor mapping process.

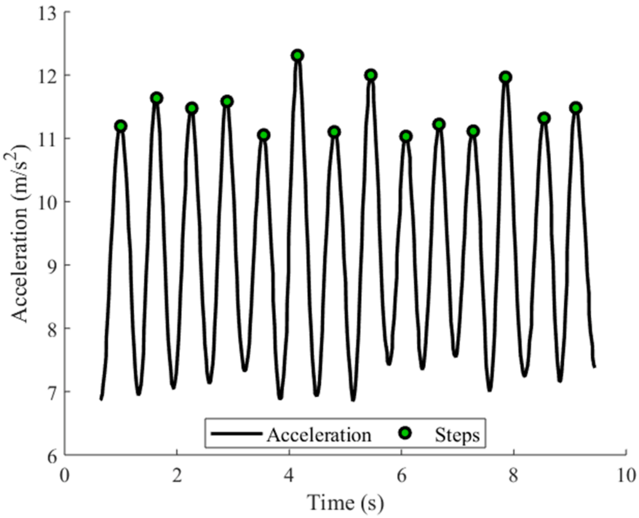

Fig. 4. Accelerometer signal results during normal walking and step detection.

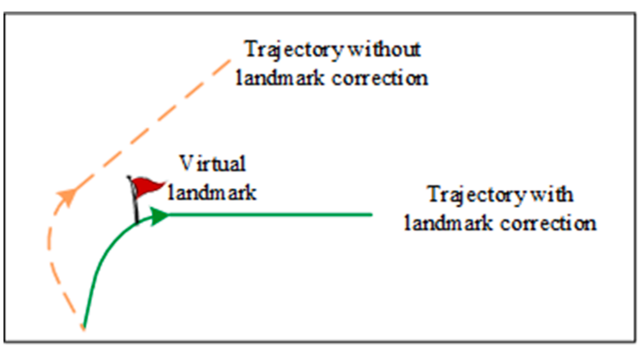

Fig. 5. Illustration of the use of virtual landmarks to calibrate a trajectory.

2.2. *Indoor mapping framework*

Based on the description of the indoor map, we summarized a technical framework of crowdsourcing-based indoor mapping. As shown in Fig. 3, the indoor mapping process consists of two main steps, geospatial-element acquisition, and indoor-map construction. These steps are demonstrated by the orange modules in the flow chat. Their functions are as follows.

- Geospatial-element acquisition. Basic geospatial elements (e.g., trajectories, ranges, and landmarks) are extracted from crowdsourced data from sensors (e.g., accelerometers, gyroscopes, magnetometers, WiFi, cameras, and microphones). The three key modules, trajectory tracking, range estimation, and landmark detection, are described in Sections 3.1–3.3, respectively.
- Indoor-map construction. The geospatial elements are processed to generate indoor maps. The approaches for generating grid-based, landmark-based, and semantic maps are reviewed in Sections 4.1–4.3, respectively.

3. Crowdsourcing-based geospatial-element acquisition

The crowdsourced data from the smartphone are disordered and lack spatial reference. These data need to be analyzed and processed to become useful geometric structures and basic elements of indoor maps. The fundamental techniques for this purpose mainly consist of trajectory tracking, ranging estimation, and landmark detection. The basic principles and recent advances of these techniques are introduced in this section.

3.1. *Trajectory tracking*

The crowdsourcing user trajectories are important data sources for extracting the basic route and structure of indoor maps. According to the method of data collection, smartphone-based trajectory tracking methods can mainly be divided into pedestrian dead reckoning (PDR)-based and simultaneous localization and mapping (SLAM)-based ones.

(1) Pedestrian Dead Reckoning (PDR)

PDR is a specific dead reckoning technique dedicated to pedestrians. Dead reckoning calculates the current position from a previously determined position by integrating angular and linear motion measurements. Almost all smartphones are now equipped with inertial sensors, making them able to track user trajectories by dead reckoning. Constrained by size, cost, and power consumption, smartphone inertial sensors usually have low accuracy. Therefore, the traditional dead reckoning method is not suitable for smartphones due to the rapid accumulation of errors (Davidson and Piche, 2017; Harle, 2013). To alleviate this issue, the cyclic pedestrian motions (e.g., the periodic steps in Fig. 4) can be applied to correct the DR solutions. This fact inspires the use of PDR (Judd, 1997; Levi and Judd, 1996).

A typical PDR system consists of three parts: step detection, step-length estimation, and walking-direction estimation. Step detection is realized by searching for repeating patterns in accelerometer or gyroscope data during walking. Thresholding and peak detection algorithms are often used together for step detection, where walking motion is detected by thresholds, and steps are counted by peak detection (Brajdic and Harle, 2013). In most situations, steps are detected using an accelerometer, and in some situations, a gyroscope is used (Davidson and Piche, 2017).

Step-length estimation is used to estimate the length of each step. Then, the distance can be estimated by summing all the step lengths.

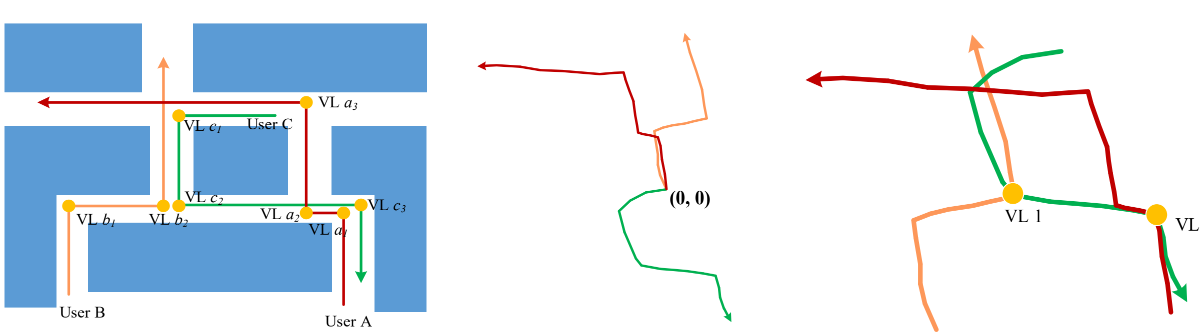

Fig. 6. Illumination of the use of virtual landmarks to align trajectories. The left figure shows the ground truth of three users; the middle figure shows the inferred trajectories by PDR; the right figure shows the aligned result. (VL: Virtual landmark).

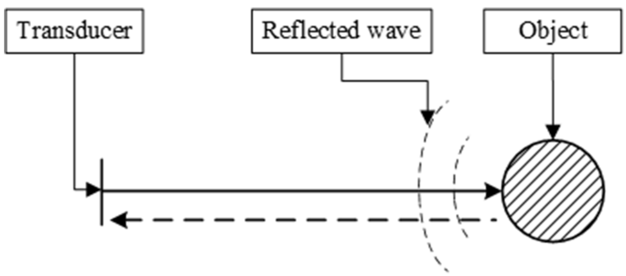

Fig. 7. Illustration of echo-based ranging (Graham et al., 2015).

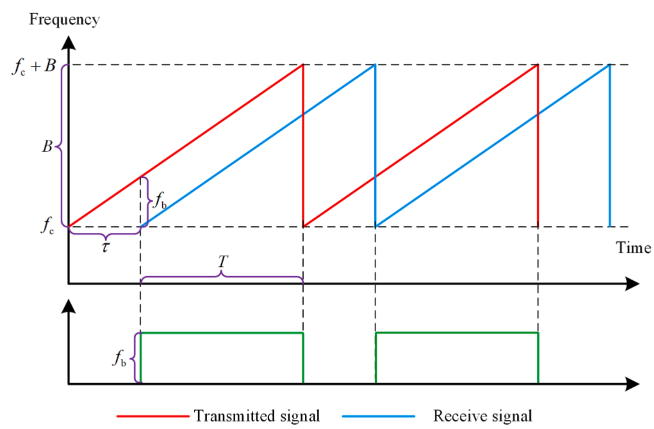

Fig. 8. Chirp signal in FMCW (Wang et al., 2018a).

Some research assumes that the step length of a user is a constant that is determined by user height (Pratama and Hidayat, 2012). There are several step-length models, including the step frequency-based model (Hilsenbeck et al., 2014) and step frequency and height-based model (Renaudin et al., 2012).

Walking-direction estimation can be obtained directly from the readings of the magnetic compass. The integration of gyroscope signals can provide estimations of direction changes. The accuracy of smartphone sensors is low; walking direction estimation based on a single sensor has large errors. The combination of the magnetic compass and gyroscope will give better results than using the sensors separately (Ladetto and Merminod, 2002).

(2) Virtual landmark

Due to the existence of inertial sensor errors, the estimated trajectory deviates from the ground truth. The deviation makes it difficult to construct an accurate indoor map using raw PDR trajectories. Moreover, the trajectory inferred by the PDR algorithm contains only relative coordinates. If the initial position is unknown, the crowdsourced trajectories are without spatial reference. To overcome these challenges, virtual landmark-based approaches have been proposed for PDR correction.

The virtual landmarks used for indoor mapping include inertial-based landmarks (e.g., elevators, stairs, escalators, and corners) and WiFi landmarks. Inertial-based landmarks are related to pedestrian activity, which can be detected by an activity-detection algorithm (Yang et al., 2018; Zhou et al., 2019b). With the advancement of smartphone technology, smartphone-based activity detection has drawn attention from many scholars (Jain and Kanhangad, 2017; Lara and Labrador, 2013; Ronao and Cho, 2016). For example, Crowd Inside (Alzantot et al., 2012) proposes a decision-tree-based classification algorithm for activity detection. The activity-related locations are used as virtual landmarks to recalibrate the crowdsourced trajectories. Refer to (Lara and Labrador, 2013) for smartphone-based human activity recognition.

The virtual landmarks are used for PDR correction through two techniques: trajectory recalibration and trajectory alignment. Trajectory recalibration resets crowdsourced trajectories based on virtual landmarks. Since the accuracy of the smartphone inertial sensors is low, the inferred trajectories deviate from the ground truth. If the locations of the virtual landmarks are known, the trajectories can be recalibrated when a virtual landmark is detected. Its principle is shown in Fig. 5. CrowdInside (Alzantot et al., 2012) and SenseWit (He et al., 2017) utilize this method to assist indoor mapping. This method requires a sufficient number of virtual landmarks to exist in the indoor environment. Moreover, the positions of the virtual landmarks are inferred by PDR. The initial points are obtained by other localization techniques, e.g., the loss of the GPS fix (Alzantot et al., 2012). In a crowdsourcing-based system, there is usually more than one trace containing the same virtual landmark. In this case, the position of the virtual landmark is estimated by averaging all the traces. Due to the low accuracy of inertial smartphone sensors, the position estimation results are always inaccurate, which limits the accuracy of the constructed maps.

Virtual landmark-based indoor map construction systems are based on the assumption that there are a sufficient number of virtual landmarks to calibrate traces. Moreover, to ensure the location estimation accuracy of the virtual landmarks, there needs to be enough traces passing through them. In reality, these assumptions do not always hold in all environments.

Besides trajectory recalibration, virtual landmarks can be used for trajectory alignment. Trajectory alignment is important because the

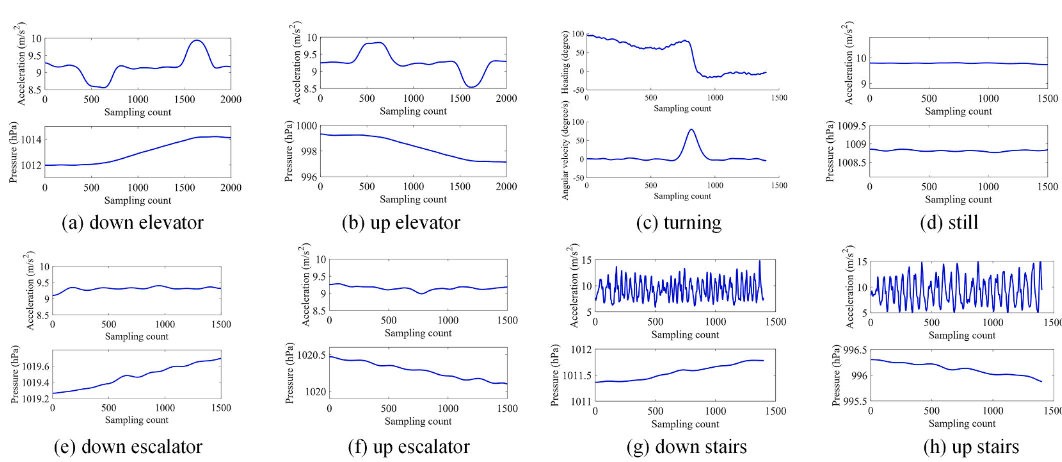

Fig. 9. Sensor signatures of different inertial-based landmarks.

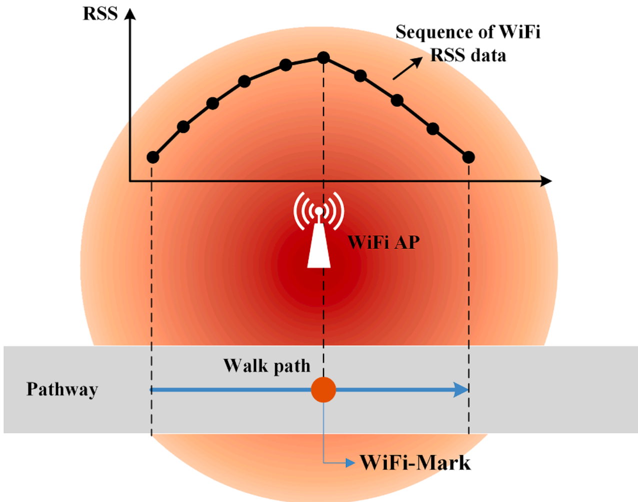

Fig. 10. Illustration of the WiFi-Mark (Shen et al., 2013).

crowdsourced trajectories are without spatial reference. The reason for this phenomenon is that the crowdsourced trajectories are collected by various users at different times; moreover, the trajectories inferred by PDR contain only relative coordinates. Trajectory alignment is based on the observation that some trajectories contain the same virtual landmarks. That is, different users may walk through the same virtual landmarks.

Fig. 6 shows the use of virtual landmarks to align crowdsourced trajectories. Three smartphone users walked in the indoor environment. There are three, two, and three turns in User A, B, and C's trajectories, respectively. These turns can be used as virtual landmarks. Some virtual landmarks are at the same position, which can be clustered based on the context information (e.g., WiF signals (Zhou et al., 2015b; Zhou et al., 2018)). For example, VL b_2 / VL c_2 and VL a_1 / VL c_3 are two clusters. Since PDR can only generate a relative user trajectory, the three user trajectories all start at the origin of a local coordinate system, as shown in the middle subfigure. Based on the virtual landmark clustering result, we can align the trajectories, as shown in the right subfigure.

Several indoor mapping approaches utilize virtual landmarks to align crowdsourced trajectories. Walkie-Markie leverages WiFi-Mark as the virtual landmark and applies a spring relaxation-based graph optimization algorithm for indoor map construction (Shen et al., 2013), while ALIMC utilizes activity landmarks as virtual landmarks and uses multidimensional scaling (MDS) to combine crowdsourced trajectories (Zhou et al., 2015b). Meanwhile, Zhou et al. propose a graph optimization method for crowdsourcing-based indoor map construction that leverages activity landmarks as loop positions (Zhou et al., 2018).

(3) Simultaneous Localization and Mapping (SLAM)

SLAM has been one of the most popular topics in the robotic field over the last few decades. It is designed to build a map of an unknown environment and simultaneously localize the robot on the map (Bailey and Durrant-Whyte, 2006; Grisetti et al., 2010).

A traditional SLAM approach usually employs odometry and a laser/video scanner to capture the physical space and constraints of the environment. In smartphone-based SLAM approaches, the step detection and direction estimation results are used as the rangefinder, while the images or wireless signals are used to observe the environment. SmartSLAM first utilizes the SLAM technique to construct indoor maps using smartphones (Shin et al., 2012). To improve the trajectory estimation accuracy, some researchers use WiFi observations as landmarks to perform SLAM. Within these works, WiFi SLAM was first used for localization (Grisetti et al., 2010; Zhou et al., 2018). It utilizes inertial data to estimate the mobility and WiFi data for the observation model.

3.2. Range estimation

Because of the complex geometric structure, rebuilding an entire indoor scene by relying only on trajectory tracking is not sufficient. For example, user trajectories cannot cover indoor open spaces such as rooms or lobbies. Hence, other means are needed to estimate the range (e.g., room size and corridor width) of the indoor scene. There are four main range-estimation methods: PDR, structure from motion (SfM), echo-based ranging, and frequency modulated continuous wave (FMCW)-based ranging. PDR was introduced in the previous section; thus, this section introduces the remaining three methods.

(1) Structure from motion (SfM)

SfM is a classical method that is used to determine camera and point geometry based on pixel correspondences (Snavely et al., 2010). It is usually used to model the 3D world from 2D images by feature detection and matching (Li et al., 2017b; Szeliski, 2010). The feature detection and matching are usually called correspondence estimation, which finds matching 2-D pixel sets in all the input images. Each set of matching pixels represents a 3-D point. Once the correspondences between images are estimated, SfM uses them to recover the 3-D camera poses and 3-D positions of the points in the feature matching sets. SfM can be considered an optimization problem to find the configuration of the camera poses and 3-D points that best agree with the correspondences based on the principle of triangulation. Developing methods to find matching 2-D pixels in an unorganized set of images is a research hotspot that has

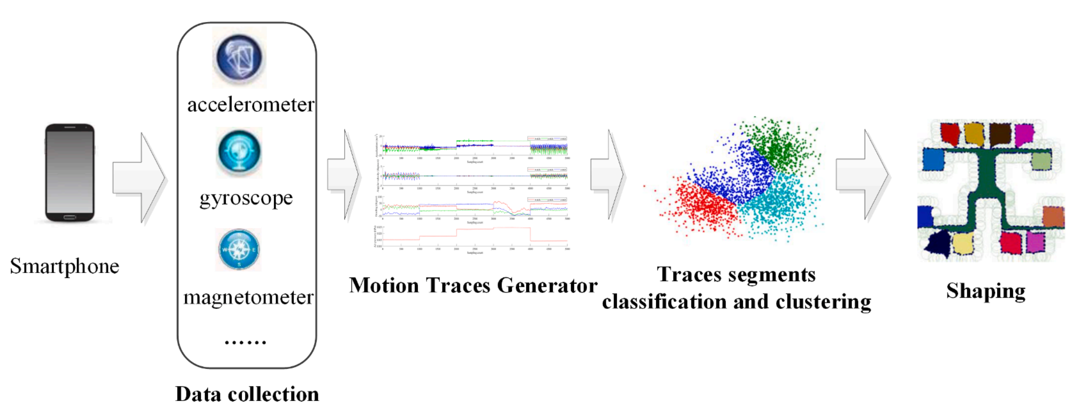

Fig. 11. The general process of occupancy grid map construction.

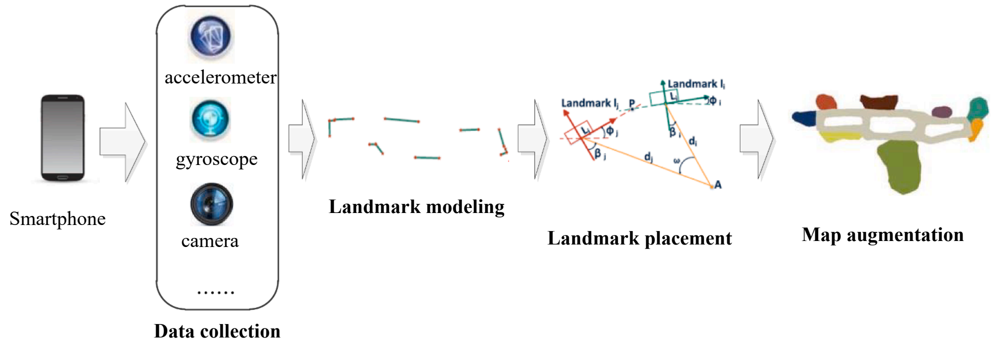

Fig. 12. The general process of landmark-based map construction.

attracted attention from the computer vision community in recent decades (Förstner, 1986; Lowe, 2004; Mikolajczyk, 2005). Various powerful feature extractors have been proposed (Lowe, 2004; Mikolajczyk, 2005), among which the scale-invariant feature transform (SIFT) (Lowe, 2004) is one of the most popular features.

(2) Echo-based ranging

Intuitively, the distance between the audio sender and the receiver can be calculated by the time of flight (ToF) of the acoustic signal. Precise time synchronization is required for ToF-based ranging. However, such precision cannot always be ensured in indoor environments.

To mitigate time-synchronization errors, the echo-based ranging technique is used (Graham et al., 2015; Peng et al., 2007), as shown in Fig. 7. First, the system emits an acoustic signal. The signal will be reflected, and an echo is generated if it encounters an object. The time difference between the initial emitted signal and the echo wave is used to calculate the distance between the sender and the object. The distance is calculated by multiplying the time difference by the sound velocity and dividing the result by two.

(3) Frequency modulated continuous wave (FMCW)-based ranging

FMCW is a classical ranging technique that was originally designed for radar systems (Stove, 1992). Different from the echo-based ranging technique, which estimates the time delay, FMCW estimates the propagation delay based on the frequency shift of the chirp signal.

Fig. 8 illustrates the process of FMCW (Wang et al., 2018a). The red line denotes the transmitted signal, the blue line denotes the received signal, and the green line is the mixed signal composed of the transmitted signal and received signal. According to the equations of signal processing (Wang et al., 2018a), the distance between the transceiver and target is calculated by $D = \frac{CT}{2B} f_b$, where D is the distance, C is the propagation speed, T is the sweep period, B is the sweep bandwidth, and f_b is the frequency of the mixed signal, which is obtained by multiplying the transmitted signal by the received signal.

3.3. Landmark detection

As introduced in Section 2, an indoor map can be described as the connection of a sparse set of landmarks. According to smartphone sensors, landmark-detection methods can be classified as inertial-based, WiFi-based, and visual-based ones.

(1) Inertial-based landmark detection

Inertial-based landmarks are based on the observation that built-in sensors generate unique sensing data signatures when people pass certain locations. These locations can be detected as landmarks. For example, when people take the elevator, there is an overweight or a weightlessness signature, which can be detected by the accelerometers in a smartphone. Fig. 9 shows the sensor signatures of different inertial-based landmarks in indoor environments.

(2) WiFi-based landmark detection

Walkie-Markie considers a WiFi signal sequence for landmark detection, which is called WiFi-Marks (Shen et al., 2013). Walkie-Markie utilized WiFi-Marks to combine crowdsourced trajectories. A WiFi-Mark is a specific position, where the Received Signal Strength (RSS) trend changes from increasing to decreasing, as shown in Fig. 10. The reason for using the RSS tread is that it is robust to device diversity.

(3) Visual-based landmark detection

Visual-based landmark detection is the process that applies computer vision techniques to extract geometric features (e.g., room layout, widths of entrances, and lengths/orientations of adjoining walls) from images. Jigsaw (Yürür et al., 2014) combined inertial data and images to construct indoor maps. This method extracted the sizes and coordinates of landmarks from the images using SfM and vanishing line detection algorithms. Also, CrowdMap (Chen et al., 2015b) detected the room layout by selecting line segments along the vanishing line direction based on the generated panorama.

4. Crowdsourcing-based indoor-map construction

After extract the basic elements and structures of indoor maps from crowdsourcing data, there is a need to connect and integrate these trajectories, structures, and landmarks to form the indoor map. In this section, we introduce the different indoor map construction technologies and the details of each indoor mapping system. The indoor map construction technologies can be mainly divided into two ways, which are occupancy grid map construction and landmark-based map construction. On this basis, some researchers tried to add semantic information to maps.

4.1. Construction of occupancy grid map

The occupancy status of each cell can be determined by occupancy detection. To construct an indoor map, the boundary of the occupied cells is estimated. Boundary estimation is a computational geometry problem. For this purpose, convex, which is a concept in computational geometry that has been widely used for point cloud processes (Lipuš et al., 2019; Sampath and Shan, 2007), can be used. For example, CrowdInside used the alpha-shape algorithm to generate the overall floorplan shape (Alzantot et al., 2012; Elhamshary et al., 2018).

Fig. 11 shows the main process of occupancy grid map construction.

Each user step is represented by a point. Therefore, a point cloud is generated from the crowdsourcing trajectories. The goal of the convex hull algorithm is to estimate the best shape that represents the point cloud. The alpha-shape algorithm is implemented to generate the shape of the building (Edelsbrunner, et al., 1983). Compared to the building shape generated by the convex hull, the building shape generated by the alpha-shape is more accurate.

(1) Systems for crowdsourcing-based construction of indoor occupancy grid maps

- **CrowdInside**

CrowdInside is an indoor map automatic construction system based on crowdsourced trajectories collected by users moving in a building (Alzantot et al., 2012; Elhamshary et al., 2018). To generate accurate motion traces, it proposed an anchor point (virtual landmark)-based error resetting technique. Anchor points were points in the environment that can generate specific sensor features for trajectory calibration. CrowdInside consists of three modules: (a) the data-collection module, which collects crowdsourced data; (b) the trace-generation module, which builds accurate motion traces; and (c) the floorplan-estimation module, which generates the indoor map by separating the corridors from the rooms and detecting the room boundaries.

Experiments were conducted in a campus building with an area of approximately 448 m^2. Results show that the false-positive and false-negative rates of anchor point detection were 0.2% and 1.3%, respectively. The experiments demonstrate that CrowdInside can efficiently construct an indoor map using crowdsourcing trajectories. However, it is based on the assumption that there are sufficient numbers of anchor points to reset trace errors. Thus, it cannot be implemented in buildings with few anchor points. Moreover, CrowdInside requires that user steps must cover the whole area of the room, which would require much time to construct the indoor map because many users only move within part of a room.

- **SenseWit**

SenseWit is an efficient crowdsourcing-based indoor map construction application that uses only inertial sensors (Jain and Kanhangad, 2017). Similar to CrowdInside, SenseWit is based on the observation that people's activity offers useful information for location estimation.

The SenseWit experiment was conducted with 10 volunteers in two environments: an office of 24 m * 19.2 m and one floor in a university library. Experimental results show that the feature location recognition achieved an accuracy higher than 85% recall and more than 90% precision. For the labeling accuracy, approximately 90% of the feature locations had errors within 1 m. The room size error was 31.4%.

- **iFrame**

iFrame is an indoor floor plan construction system that can be constructed automatically by sharing information between smartphones (Qiu and Mutka, 2017). The moving trajectories of the users were estimated by the dead reckoning algorithm. To mitigate the drifts in dead reckoning solutions, iFrame used a Markov chain to improve trajectory estimation and employs Bluetooth and WiFi position updates. Moreover, it adopted curve fit fusion to improve the output matrices and generates an initial shadow map. Finally, it employed the anchor points proposed by CrowdInside (Alzantot et al., 2012) to rebuild the floor plan.

To evaluate iFrame, an experiment was conducted in an academic lab at Michigan State University. The experimental results show that iFrame could effectively generate a skeleton map. The layouts of 12 rooms were reconstructed within 5–10 min. Also, the changes in the layout of the indoor map could be detected.

- **SmartSLAM**

SmartSLAM is a crowdsourcing-based indoor-map construction system that combines inertial and WiFi data for crowdsourced trajectory estimation (Shin et al., 2012). It used the WiFi signal strength as the observation value and the inertial data as a mobility model. Based on the observation and mobility model, it employed the FastSLAM algorithm (Dong et al., 2018) to track the pedestrian and localize WiFi APs simultaneously. The pedestrian positions were used for floorplan construction.

The SmartSLAM experiments were implemented in the Yonsei Engineering and Research Center. The localization errors were less than 4 m on average. It was also performed in multiple buildings. The experimental results demonstrate that it can build corridor paths for buildings with different layouts.

- **ShopProfiler**

ShopProfiler is a crowdsourcing-based system that can automatically profile shop type and brand name (Guo et al., 2014). It used customer movement patterns as the feature for shop categorizing since the movement patterns are different in different shops. It consists of four modules. The data-collection module was used to collect data from inertial sensors, microphones, and WiFi. The movement pattern-capturing module was used to investigate movement patterns in different shops within a mall. The differentiating basic unit module is used to differentiate a shop and corridor based on the unique features of movement. Moreover, a gradient-based room boundary detection algorithm was applied to improve system performance. Based on the collected data and movement patterns, ShopProfiler utilized (Support Vector Machine, SVM) to categorize shops and labeled the shop brand names using (Service Set Identifier, SSID) mining.

To evaluate the ShopProfiler, experiments were conducted on three different scenarios. The experimental results show that it achieved over 80% prediction accuracy of the shop category.

- **Jigsaw**

Jigsaw is a crowdsourcing-based floor plan reconstruction system that can generate indoor maps with hallway connectivity, room sizes, and shapes. It combined inertial data and images to construct indoor maps (Gao et al., 2014). The position, size, and orientation information were extracted from the images. The Jigsaw system consists of three parts: landmark modeling, landmark placement, and map augmentation.

The experimental results of Jigsaw shown that the landmark position and orientation estimation errors were respectively 1–2 m and 5°–9° in the 90th percentile and the hallway connectivity accuracy was 100%. Jigsaw can generate the floorplan of a single floor. Recently, R. Gao et al. extended Jigsaw by adding connection area detection, which can be used to reconstruct a multistory indoor floorplan (Gao et al., 2016). The connection areas in a multistory building include stairs, elevators, and escalators. The connection area detection is based on the distinctive patterns of the inertial and wireless signals when a user passes through such areas. The idea is similar to CrowdInside. The connection areas are detected by an unsupervised classification algorithm. By connection area detection, a topological map can be generated, which can be used for indoor navigation.

(2) Summary

The advantages of CrowdInside and SenseWit are that they use only inertial sensors to construct indoor maps. To improve the accuracy of crowdsourcing trajectories, these systems apply the virtual landmark-based trajectory recalibration method, which achieves a good result. However, they are based on the assumption that there is a sufficient number of anchor points for trajectory recalibration and that the user trajectories will cover the whole area of the building. iFrame assumes that Bluetooth Received Signal Strength (RSS)is related to distance and needs to evaluate the RSS-distance mapping relations in advance. The assumption is unreliable due to the complex indoor environment, which has a multipath effect. Also, the pre-trained mapping relations for one type of smartphone are not suitable for all types. Moreover, iFrame assumes that the abrupt changes in the WiFi signal are related to obstacle detection. This assumption is also unreliable because many factors can cause WiFi signal mutations, such as multipath effects and human occlusion. SmartSLAM is the first SLAM-based indoor map construction system that uses smartphones and can achieve good mapping performance. However, it can only generate corridor paths for buildings. Jigsaw requires participants to take images by two designed modes, Click-Rotate-Click (CRC) and Click-Walk-Click (CWC), which may cause inconvenience to the users. It utilizes SfM to obtain the sizes and coordinates of landmarks. However, this method tends to fail in a featureless environment.

The key technique of occupancy grid map construction is indoor positioning. However, the sensor-based indoor positioning effect is often influenced by the indoor environment. Only one type of sensor can't apply all the situations. Besides, the visual-based method like SfM may construct point insufficiently accurate when images are inadequate. Random variations such as moving customer flows may constitute a disturbance to the appearance of landmarks. Moreover, this data collection has high energy consumption and privacy issues. Although the authors claimed that the storeowners welcome such exposure for advertisement, the security guard of the mall may prevent the participants from taking a large number of pictures.

4.2. Construction of the landmark-based map

Compare to the occupancy grid map construction that generally using user's trajectories clustering, the landmark-based map construction tend to base on landmark recognition. As shown in Fig. 12, the main processes of landmark-based map construction are generally landmark modeling, landmark placement, and map augmentation. The locations of the landmarks can be estimated by the range-estimation method introduced in Section 3. However, there are ranging errors, which degrades the accuracy in landmark location estimation. Thus, the landmark locations tend to be assigned as a classical node-embedding problem and solved by optimization-based methods. The optimization algorithm assigns optimal coordinates to the landmarks, which can be divided into the spring-relaxation and multidimensional scaling (MDS) techniques.

- **Spring-relaxation**

The spring-relaxation technique has been widely used for cooperative localization (Priyantha et al., 2003; Seet et al., 2012; Zhang et al., 2010). Cooperative localization is used to solve the following problem: given a set of nodes with unknown locations and the distances between one node to a few neighboring nodes, determine the locations of every node by node-to-node communication (Priyantha et al., 2003).

Spring-relaxation-based landmark location optimization uses the concept to simulate the movements of the sensor under spring forces and find the final resting location of the sensor, which is the estimated location.

- **Multidimensional Scaling (MDS)**

MDS (Borg and Groenen, 2003) is a data analysis technique that is often used in information visualization for exploring similarities or dissimilarities in data. It is a dimensionality reduction method that displays data graphically to make it easier to understand (Saeed et al., 2019). The input of an MDS algorithm is a matrix of item-item dissimilarities (Borg and Groenen, 2003; Bronstein et al., 2006; Yang et al., 2012). The inter distance is usually used as a metric of dissimilarity. Many network localization approaches adopt MDS to estimate the locations of wireless devices (Costa et al., 2006; Shang and Ruml, 2004). Some researchers use MDS to estimate the locations of WiFi APs (Koo and Cha, 2012). They extract the dissimilarities between different WiFi APs based on the scanned RSS. Meanwhile, researchers use MDS to automatically label WiFi fingerprints for indoor localization (Yang et al., 2012).

(1) Systems for crowdsourcing-based construction of indoor landmark-based maps

- **Walkie-Markie**

Walkie-Markie is a crowdsourcing-based indoor mapping system that can reconstruct pathway maps by fusing the user trajectories and special landmarks called WiFi-Marks(Shen et al., 2013). Walkie-Markie consists of a client and a backend. The client applied a walking state detection engine to periodically detect a user's walking state. With the clustered WiFi-Marks and connecting user trajectories, the backend service generated the indoor pathway maps by an expansion-shrinking process.

To evaluate the performance of Walkie-Markie, experiments were conducted on an office floor with an area of 3600 m^2 and an internal pathway length of 260 m and a shopping mall with an internal pathway length is approximately 310 m. The experiments were shown that the maximum error of the anchor nodes and path segment estimation were within 3 m and 2.8 m, respectively.

- **ALIMC**

ALIMC is an activity landmark-based indoor mapping system that constructs indoor maps by using activity landmarks to merge crowdsourced trajectories (Zhou et al., 2015b). Activity landmarks are virtual landmarks where people engage in special activities, such as elevators, corners, and stairs.

ALIMC was implemented on two floors of an office building, with a 52.5 m*52.5 m floorplan. The experimental results demonstrated that ALIMC achieved a good result for indoor mapping, with an 80th-percentile mapping error of 0.8–1.5 m. However, ALIMC utilizes the MDS technique to merge crowdsourced trajectories, which is unsuitable for buildings with circular structures. Moreover, ALIMC is based on the assumption that there are enough activity landmarks in the environment, which is occasionally unrealistic. ALIMC can only generate a topology map for the indoor environment.

- **G2OMap**

G2OMap is a crowdsourcing-based indoor map construction method that utilizes graph optimization techniques to align crowdsourcing trajectories (Zhou et al., 2018). Similar to ALIMC, the initial innovation of G2OMap is that it used an activity landmark as the loop position point to realize the loop closure in the SLAM framework.

G2OMap was implemented in an office building with a 52.5 m*52.5 m floorplan and a shopping mall with a 100 m*70 m floorplan. The 80th-percentile mapping error was approximately 1.7–3.5 m.

- **Hallway**

Hallway[1] leveraged the WiFi signal and motion information to construct an indoor map (Jiang et al., 2013). It is composed of three parts: room adjacency graph construction, which constructed a room adjacency graph; hallway layout learning, which determined the rooms and their orders along the hallway; and force-directed dilation, which optimized the overall floorplan accuracy.

Hallway was implemented in five buildings with different floorplan

[1] We call this solution Hallway for convenience

structures. The experimental results show that the average room position accuracy was 91%, the room area estimation error was 33% and the room geometric aspect ratio error was 24%.

- **PiLoc**

PiLoc proposes an indoor map construction method by combining motion and WiFi information (Luo et al., 2014). It consists of three components. The clustering component was used to divide all collected trajectories into disjoint clusters that cover different indoor environments based on the WiFi signal strength and moving vector. The correlation matching component was used to find the overlapping trajectory segments of the different clusters based on AP signals and movement vectors. The overlap segments were further used to construct the floorplan and radio map simultaneously. Finally, the constructed floorplan and radio map were leveraged for indoor localization.

PiLoc was conducted on two different floors, with sizes of 900 mˆ2 and 120 mˆ2, to evaluate the floorplan construction performance. The experimental results show that PiLoc achieved average step mapping errors of 1.27 m–1.65 m and 0.46 m–0.6 m on the 900 mˆ2 and 120 mˆ2 floors, respectively.

- **CISWS**[2]

Sankar and Seitz have proposed a novel smartphone application for capturing and reconstructing indoor scenes based on camera and inertial sensor data collected by a smartphone (Sankar and Seitz, 2012). To reconstruct indoor scenes, the participant first captured a video of an indoor environment following a few guidelines. The video was then indexed spatially with the camera pose information obtained by the inertial sensors. With the obtained video, CISWS generates a visual rendering of the indoor scene. The position of the camera is estimated based on this information, and a floorplan is generated.

The proposed system was tested in four indoor environments. The results show that CISWS reconstructed the dimensions and floorplan with an average error of 10.45%.

- **CrowdX**

CrowdX is a crowdsourcing-based indoor floorplan construction system that leverages opportunistic user encounters to reset the dead-reckoning error (Chen et al., 2018). The relative spatial relationship between mobile user trace segments was derived by audio ranging and dead reckoning technology. CrowdX constructed a floorplan based on the inertial, Bluetooth, and audio data. Tracing profiling was used to generate traces based on inertial sensor data and check whether the trace contains the segments inside the room. The segments inside the room are used to estimate the room area, while others in the hallway are used to estimate the landmark position and assemble the hallway.

Experiments in three shopping malls were conducted to evaluate CrowdX. The experimental results show that the average F-score was approximately 89.4%, and the average room area estimation error was approximately 20%.

- **BatMapper**

BatMapper is an acoustic sensing-based indoor floorplan construction system that combines smartphone acoustic and inertial data to generate indoor maps(Zhou et al., 2017a; Zhou et al., 2019a). It was based on the principle that the distance between a smartphone and an object can be estimated based on the time difference between the sound emission and echo reception. The input of BatMapper included data from three sensing modalities, i.e., acoustics, gyroscopes, and accelerometers. From these modalities, echo candidates and user traces were extracted and combined by mapping algorithms for indoor map construction. The output of BatMapper includes regular rooms, irregular rooms, and corridors.

BatMapper was tested in three buildings: a 40*60 m laboratory, a 50*60 m teaching building, and a 45*45 m office building. The experimental results show that the distance measurement accuracy was 1–2 cm at approximately 4 m. BatMapper could generate fine rough corridor shapes with 2–3 min of walking. The door detection and localization accuracies were 92% and 1–2 m at percentile, respectively. The room geometry estimation error was less than 0.3 m at the 80th percentile.

- **SAMS**[3]

SAMS is an acoustic-based system for indoor map construction that infers the structure of indoor space by analyzing the audio signals reflected from the environment with a smartphone (Pradhan et al., 2018). SAMS allowed a smartphone to emit audio signals and analyzed the reflected signals to estimate the distance between the smartphone and walls. Meanwhile, it inferred the user moving trajectory and combines it with the audio-based distance estimation to create the contour of the indoor space.

The experimental results show that the median errors of the distance measurement of a single wall and multiple walls are 1.5 cm and 6 cm, respectively. Moreover, the maximum indoor map contour estimation error of SAMS is 1.2 m.

(2) Summary

Walkie-Markie works well for normal indoor pathways that are usually narrow. However, for large open areas, the performance of Walkie-Markie may deteriorate when users walk arbitrarily. This is because the WiFi-Mark clustering process may cause errors in a large open area. ALIMC utilizes the MDS technique to merge crowdsourced trajectories, which is unsuitable for buildings with circular structures. Moreover, it is based on the assumption that there are enough activity landmarks in the environment, which is sometimes unrealistic. ALIMC can only generate a topology map for the indoor environment. Similar to ALIMC, G2OMap can only generate a topology network of an indoor map, which is not sufficient for ILBS applications. Hallway can estimate the room dimensions along the corridor based on the pedestrian trajectory. However, it cannot obtain the dimensions perpendicular to the corridor. Moreover, Hallway focuses on rectangular rooms, which is not suitable for more complicated layouts containing curves. For trajectory matching, PiLoc utilizes path and signal correlations as the metric. The matching algorithm works well in narrow paths. However, in open areas where people may walk arbitrarily, path and signal correlations may fail to match trajectory since the arbitrary walk path does not contain a sufficiently large distance. Due to this limitation, PiLoc can only construct the corridor path of an indoor environment. BatMapper can generate indoor maps by acoustic sensing using smartphones. However, it requires participants to walk around a whole building to collect data. Since the acoustic ranging distance is limited, BatMapper requires the participants to walk a full loop near the walls for buildings with large open spaces. Moreover, it requires the participants to operate the smartphone to collect useful data. The complicated rules designed in BatMapper make it unusable for an untrained user. SAMS can achieve good performance for indoor-map construction, which is better than that of BatMapper (2.6 m). However, SAMS is based on certain strict assumptions. First, a user is required to hold a phone in a straight line, keeping the phone microphone facing the walls. Second, SAMS assumes that a user makes only 90° turns and straight movements between the two turns to avoid heading direction errors.

[2] We use the acronyms for the title of the cited paper for easy reading.

[3] We use the acronyms for the title of the cited paper for easy reading.

Table 2
Summary of different indoor mapping solutions.

Solutions	Sensors	Participation	Output	Experiment environment	Reported accuracy
CrowdInside (Alzantot et al., 2012)	Inertial	Passive	Corridor, room	A shopping mall with plenty of virtual landmarks, and an office floor with the size of about 448 m^2	Room number estimation accuracy is 100%
SenseWit (He et al., 2017)	Inertial	Passive	Corridor, room	An office with the size of 24 m × 19.2 m, and one floor in a campus library with 464 m^2 area	Feature location recognition accuracy is more than 85% for recall, and 90% for precision; F-score of the hall shape is 78.7%; room size estimation error is 31.4%
Walkie-Markie (Shen et al., 2013)	Inertial, WiFi	Passive	Corridor	An office floor with a size of 3600 m^2, the total pathway length is 260 m; A shopping mall with an irregular layout and the pathway length is about 310 m.	The maximum error is 3 m and 2.8 m for the anchor nodes and path segments, respectively
ALIMC (Zhou et al., 2015b)	Inertial, WiFi	Passive	link-node model	Two floors of an office building, with the size of 52.5 m × 52.5 m	The 80% error is about 0.8–1.5 m
G2OMap (Zhou et al., 2018)	Inertial, WiFi	Passive	link-node model	An office building with the size of 52.5 m × 52.5 m, and one floor of a shopping mall with the size of 100 m × 70 m	The 80% error is about 1.7–3.5
iFrame (Qiu and Mutka, 2017)	Inertial, WiFi, Bluetooth	Proactive	Corridor, room	eLANs Lab of Michigan State University	The error of block value is 0.041
Hallway (Li et al., 2017a)	Inertial, WiFi	Passive	Corridor, room	A classroom building, a research lab building, a shopping mall, and an office building. The sizes are unknown	The position accuracy is 91%; the room area error is 33%; the average aspect ratio error is 24%
SmartSLAM (Shin et al., 2012)	Inertial, WiFi	Passive	Corridor	an office building, the size is unknown	The average error is 3 m
Piloc (Luo et al., 2014)	Inertial, WiFi	Passive	Corridor	Four different areas cover 5528 m^2 in total, the sizes of these four areas range from 120 m^2 to 3000 m^2	The average SME is 1.27 m, and 0.54 m for mid-size office area and research lab, respectively
Jigsaw (Gao et al., 2014)	Inertial, image	Proactive	Corridor, room	Two stories of a 150 m × 70 m shopping mall of irregular shape, and one story of a 140 m × 40 m long and narrow mall	The average RMSE of floor plans is 1.01 m and 1.32 m for landmarks and intersections, respectively. The average F-score of hallway shape estimation is 83.67%. The average room size error is 27.6%
SISE (Teng et al., 2018)	Inertial, image	Proactive	Entity semantic and location	An office building with the size of 4000 m^2	The precision and recall of semantic updating are 81.1% and 79.8%, respectively. The 90-percentile location error of the changed entities is less than 1.5 m
SemSense (Elhamshary and Youssef, 2015)	Inertial, images, WiFi, LBSN	Proactive	Corridor, room	An office building with the size of 24 m × 19.2 m, and one floor in a campus library with the size of 464 m^2	The average F-score of hallway shape estimation is 78.7%, the average room size error is 31.4%
SnapTask (Sankar and Seitz, 2012)	Image	Proactive	3D model	A library at Aalto University	The reconstruction ratio is 100% for the library walls and 98.12% for the objects and traversal areas within the library
IndoorCrowd2D (Chen et al., 2015a)	Inertial, image	Proactive	Indoor panoramic image, hallway skeleton	A teaching building and a GYM	The F-score of the hallway skeleton estimation is around 95%
CrowdMap (Chen et al., 2015b)	Inertial, video	Proactive	Corridor, room	Two college laboratories and a GYM, the sizes are unknown	The average F-score of hallway shape estimation is 90%, the average room size error is 9.8%, the average room aspect ratio is 6.5%, the average room location error is 1.3 m
CISWS (Sankar and Seitz, 2012)	video	Proactive (with user interaction)	2D and 3D indoor map	N/A	The average dimensions and map reconstruction error is 10.45%
CrowdX (Chen et al., 2018)	Inertial, acoustic, Bluetooth	Passive	Corridor, room	Three shopping malls, the sizes are 13000 m^2, 5200 m^2, and 1800 m^2, respectively	The landmarks location error is 1 m, the average F-score of the hallway shape estimation is 88.9%, the average error of room area estimation is 22%
BatMapper (Zhou et al., 2017a)	Inertial, acoustics	Proactive	Corridor, room	A laboratory with the size of 40 m × 60 m, a teaching building with the size of 50 m × 60 m, and an office building with the size of 45 m × 45 m	Door detection precision is 92%, and the 90% location error is 1–2 m; the room geometry estimation error is less than 0.3 m at 80%
SAMS (Pradhan et al., 2018)	Inertial, acoustics	Proactive	Corridor, room	N/A	The indoor map contour estimation error is 1.2 m

So far, the crowdsourcing-based construction for indoor landmark-based maps focuses on landmark recognition and connection. Most of the methods rely on the Manhattan world (MW) assumption, which can only serve for the regular indoor environment. The way of crowdsourcing data acquisition is another limitation. For connecting the whole landmark of the indoor scene, the data collectors often need to follow certain established rules. It is increased the difficulty of data acquisition and the artificial burden.

4.3. Semantic map construction

In addition to geometry, semantic information is also an important element for indoor maps. Indoor semantics represent the attributes (e.g., shop names and functionalities) of objects (called entity (Teng et al., 2018)) in an indoor environment. Recently, numerous methods have been proposed for indoor semantic labeling. Deep learning techniques are usually used to recognize entities from images. The commonly-used entity-recognition algorithms include Faster R-CNN (Ren et al., 2015), SSD (Liu et al., 2016), and R-FCN (Dai et al., 2016). These algorithms use neural networks to extract and localize the entity in an image. After entity recognition, POI information (e.g., shop name and room number) extraction is also important for indoor maps. To extract texts from images, optical character recognition (OCR) (Smith, 2007; Ye and Doermann, 2015) techniques are the common approach. ViNav (Dong et al., 2018) uses photos as input and applies OCR techniques to extract texts from images.

In addition to images, semantic information can also be extracted from other data, such as those from location-based service networks (LBSNs) (Elhamshary and Youssef, 2015), WiFi SSID, and microphones (Guo et al., 2014). These semantic information extraction methods are not universal. We will introduce them in the following section, which details each system.

(1) ***Systems for crowdsourcing-based construction of indoor semantic maps***

- **SemSense**

SemSense is a semantic indoor floorplan construction system that can automatically label indoor maps with semantic labels (Elhamshary and Youssef, 2015). In addition to smartphone sensor data, it also used check-in data from location-based social networks. SemSense consists of two parts: client and server. A SemSense client was used to collect sensor data and interact with the could server. The collected sensor data include locations, WiFi signals, and images.

To evaluate SemSense, experiments were conducted in four shopping malls including 711 stores. It achieved a semantic labeling accuracy of 87%. Also, its coverage ratio was over 27% greater than that of the current location-based social networks.

- **SnapTask**

SnapTask is a crowdsourcing-based indoor map construction system that is based on visual data (Sankar and Seitz, 2012). The innovation of SnapTask is that it proposed an efficient manner to guide participants to collect visual data with high quality-of-information (QoI) (Noreikis et al., 2018). SnapTask proposed several methods to overcome specific challenges. First, it applied model coverage analysis to determine the areas lacking data. Then, it generated a data-collection strategy that determined which data to collect. Moreover, it developed an online tool for participants to mark the bounds of featureless areas to overcome the issue of featureless surfaces.

SnapTask was used in a university library. The reconstruction ratio was 100% for walls and 98.12% for obstacles and traversable areas. Moreover, its model coverage was 20.72% and 34.45% better than that of the unguided participatory and opportunistic visual crowdsourcing methods, respectively. When reconstructing featureless surfaces, the precision and F-score of SnapTask were 98.14% and 90.23%, respectively.

- **IndoorCrowd2D**

IndoorCrowd2D is an indoor scene reconstruction system that can generate building interactive panoramic maps at a large scale with untrained users (Chen et al., 2015a). The interactive panoramic map consists of indoor panoramic images and building hallway skeletons. The former was used for indoor interview visualization, while the latter was used for interactive navigation. IndoorCrowd2D consists of two parts: a mobile data acquisition client and a computing backend. The mobile data acquisition part was used for collecting crowdsourced images and sensor data. In contrast, the cloud-computing backend was used for crowdsourcing data processing.

IndoorCrowd2D was evaluated in two college buildings by untrained and uncorrelated volunteers. The volunteers used the data-acquisition application to capture indoor scenes. 55,453 images from 1151 datasets were collected by 25 users. The experimental results show that the precision, recall, and F-score of IndoorCrowd2D were 85%, 100%, and 95%, respectively.

- **CrowdMap**

CrowdMap is a crowdsourcing system that utilizes inertial sensors and video data to reconstruct indoor floorplans (Chen et al., 2015b). It was first tracked user movements based on these data and then uses the inferred user trajectories and image context to generate a floorplan.

CrowdMap was evaluated in three different college buildings. The precision, recall, and F-score of the hall-shape estimation results were approximately 88%, 93%, and 90%, respectively. The room-area estimation error was 9.8%, and the room aspect ratio error was 6.5%. Similar to IndoorCrowd2D, CrowdMap is also based on the RMW assumption, which assumes that each room has a rectangular shape.

- **SISE**

SISE is a mobile crowdsourcing system that can automatically and continuously update indoor semantic floorplans for general entities in dynamic indoor environments (Teng et al., 2018). It focused on the problem of indoor semantic floorplan updating. SISE consists of two components: a mobile application and an updating engine. The mobile application was used to collect and upload crowdsourced data, including images and inertial data. In contrast, the updating engine updated indoor semantic floorplans based on the collected crowdsourced data.

To evaluate the SISE, experiments were implemented on one floor of an office building with a size of 100 m*40 m. The precision and recall of entity recognition were approximately 81.1% and 79.8%, respectively. The localization error was within 1.5 m for 90% of the changed entities.

(2) ***Summary***

SnapTask constructs 3D models of indoor environments by SfM techniques and then converts them into indoor maps. On the one hand, the SfM always suffers from featureless surfaces, resulting in a lack of data in certain areas. On the other hand, redundant data in the hotspots generate extra processing costs. IndoorCrowd2D and CrowdMap are based on the RMW assumption, which may not be suitable for non-rectangular buildings.

The semantic information extraction is mainly based on image recognition techniques like Optical Character Recognition (OCR). The technique shows good performance when the character type is single. But, when there are more character types in the scene, the accuracy of the algorithm will be low. Besides, the technique of 3D Indoor Scene Understanding (Hedau et al., 2009) can't apply in the nonrectangular buildings. For improving the accuracy of semantic map construction, the crowdsourcing system may consider designing a friendly human–computer interaction by using advanced visualization methods, such as Augmented Reality (AR).

4.4. Solutions comparison

The three types of indoor mapping solutions are compared from the criteria of the techniques, the algorithm, the cost, the advantage, and the disadvantage.

5. Comparison of techniques

Table 2 provides a summary of different indoor mapping solutions. The solutions are compared in terms of the following criteria: the sensors used, the method of participation, the output of the solutions, the experimental environment, and the reported accuracy.

Inertial sensors are the most commonly used smartphone sensors for crowdsourcing-based indoor mapping solutions since inertial sensors can generate user trajectories, which are the most important elements for indoor-map construction. One of the greatest advantages of an inertial sensor-based method is that it is independent of external infrastructure; thus, it can provide crowdsourcing data without user proactive effort. By contrast, the image-based solutions require users to proactively take photos, which may be inconvenient. The advantage of image-based solutions is that they can obtain accurate geometry information of the indoor environment. Also, semantic information can be extracted from images. Moreover, by shooting videos, the smartphone can provide an interactive interface for a user to correct the mapping errors. WiFi-based solutions require the support of WiFi infrastructures. Moreover, due to the complex indoor environment, the WiFi RSS measurement is unstable, which may cause large errors in indoor mapping. The acoustic-based solutions exploit the acoustic-based ranging technique to estimate the distance between the sound sources and object. These solutions require smartphones to repeatedly emit and record sound signals, which may affect the normal use of smartphones.

We reviewed the performance criteria proposed in the literature to evaluate indoor mapping methods. These performance criteria can be divided into two main categories, namely, quantitative criteria and qualitative criteria. The qualitative criterion includes sensors used, participation, and output, which are listed in Table 2. The quantitative criteria are elaborated as follows.

- **Amount of crowdsourcing data**

The amount of data needed for constructing an indoor map is a useful parameter for the performance of the crowdsourcing-based method. Some studies evaluate the performance of inferred maps with different amounts of crowdsourcing data (Alzantot et al., 2012; Zhou et al., 2015a; Zhou et al., 2018).

Generally, the indoor mapping performance is enhanced as the amount of crowdsourcing data increases until it reaches a certain threshold. Thus, determining the amount of data needed is an interesting problem. However, it is difficult to obtain a common metric, for example, x-minutes of data per square meter, because indoor environments are complex and varied. Moreover, the quality of crowdsourced data may vary for different users because the map elements may be different in various types of environments. For example, the map elements of an office building are more complicated than those of a lobby.

- **Hallway shape**

The hallway path skeleton is an important component of an indoor map. Although some navigational indoor maps may display corridors as single lines, these kinds of the map are limited in many location services, such as spatial analysis and indoor positioning expression. The map with hallway skeleton can offer more spatial information and display a more accurate map. The hallway reconstruction is usually based on the occupancy grid map building, which is a dominant paradigm for environment modeling in the smartphone. The cues of hallway construction generally include the boundary of the hallway, positions of cameras, and motion traces in the hallway.

To evaluate how close the shapes of constructed hallways resemble respective ground truth, normally overlaying the reconstructed hallway onto its ground truth to achieve maximum overlap by aligning both the center point and the orientation. Precision is the ratio of the size of the overlap area to the whole reconstructed hallway.

The hallway shape can be used to evaluate the similarity between the generated hallway path skeleton and the ground truth. To calculate this metric, the generated hallway path skeleton is first overlaid onto the ground truth. Then, the center point of the generated indoor path skeleton is moved and rotated to achieve maximum overlap. After that, the parts belonging to the room are cut off. To evaluate the hallway shape estimation performance, the metrics below are used:

$$P = \frac{|S_{gen} \cap S_{true}|}{|S_{gen}|}$$

$$R = \frac{|S_{gen} \cap S_{true}|}{|S_{true}|}$$

$$F = 2 \times \frac{P \times R}{P + R}$$

where P, R, and F are the precision, recall, and F-score of the hallway shape, respectively. P is defined as the ratio of the overlapped area to the generated area. R is defined as the ratio of the overlapped area to the ground truth.

- **Room size**

The room construction is often used as the occupancy grid map, whose cues are segments of landmark models and motion traces inside the room. The accurate shape of a room is an essential part of an indoor map, which makes the map expression more complete and clear. The room size is the difference between the size of the generated room and that of the ground truth divided by the size of the ground truth.

$$RS_{error} = \frac{|RS_{gen} - RS_{true}|}{RS_{true}}$$

Besides, the room aspect ratio (*RAR*) denotes the shape of a room, which is defined as the room length divided by the room width:

$$RAR = \frac{L_{room}}{W_{room}}$$

where L_{room} stands for the room length and W_{room} is the room width. The RAR error is defined as follows:

$$RAR_{error} = \frac{|RAR_{gen} - RAR_{true}|}{RAR_{true}}$$

- **Position of feature points**

The positions of feature points are evaluated by the root mean square error (RMSE) of the feature points in the inferred floorplan and their corresponding ground truth. Given n feature points on an indoor map with 2D coordinates $X_i^{map} = (x_i^{map}, y_i^{map})$ and the corresponding ground truth $X_i^{gt} = (x_i^{gt}, y_i^{gt})$, $i = 1, 2, \cdots, n$, the RMSE is calculated by

$$\boldsymbol{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (\boldsymbol{X}_i^{map} - \boldsymbol{X}_i^{gt})^2}{\boldsymbol{n}}}$$

The RMSE of the feature points is also called the graph discrepancy metric (GDM) in other publications (Shen et al., 2013; Zhou et al., 2015a; Zhou et al., 2018).

- **Graph/Shape Discrepancy Metric**
- As above, the graph discrepancy metric (GDM) reflects the difference between the feature points of the constructed map and that of the ground truth. The RMSE of the feature points is GDM. The shape discrepancy metric (SDM) reflects the differences between the shapes of the constructed paths and ground truth. To calculate the SDM, the link segments between feature points are uniformly sampled to obtain a series of sample points. The RMSE of the sample points is SDM. **Indoor semantic information**

For indoor semantic information extraction, the performance criteria are similar to the classification method. The most commonly used criteria are precision, recall, and F-score. The F-score is defined as F = (2 × P × R)=(P + R), where P refers to the precision and R refers to the recall.

6. Open issues and future research directions

Crowdsourcing-based indoor mapping using smartphones is one of the most promising applications for indoor map construction and updating. The techniques have both academic and industry values. On the other hand, there are still challenges in applying the current research results to real-life applications. This section discusses the open research issues for crowdsourcing-based indoor mapping using smartphones and possible future research directions.

6.1. Lack of uniform performance comparison criteria and datasets

To evaluate crowdsourcing-based indoor mapping systems, many performance comparison criteria have been proposed, as reviewed in Section 4. However, some evaluation criteria are subjective and unquantifiable. Also, there are currently no uniform performance comparison criteria for indoor mapping. To alleviate this issue, it is worthwhile to learn from other fields. For example, in the indoor-localization area, there are several competitions, such as EvAAL[4], Microsoft Indoor Localization Competition[5], and PerfLoc[6]. In these competitions, the committee gives some uniform performance evaluation metrics, which promote the development of the corresponding techniques. Thus, determining the uniform performance comparison criteria for indoor mapping is a promising research direction.

Moreover, there are few benchmarking datasets for crowdsourcing-based indoor mapping. Evaluation experiments are usually performed in various indoor environments, which makes it difficult to compare different solutions. This issue also exists in the indoor-localization field. Recently, for indoor localization, researchers have published several datasets, which bring great benefits for related research (Lohan et al., 2017; Mendoza-Silva et al., 2018; Mendoza-Silva et al., 2019). Also, benchmarking datasets in other research fields, such as Lenna images in the image-processing field, have achieved great success. Similarly, publishing crowdsourcing data for indoor mapping is important.

6.2. Multi-sensors data fusion

As illustrated in Section 4, although a single sensor-based method can implement indoor mapping, there are some limitations and differences. For example, the inertial sensors equipped in smartphones are used for positioning, which is not accurate enough as well known. Also, the positioning error accumulates with time. As well known, the smartphone has been equipped with many sensors (e.g., accelerometers, gyroscopes, magnetometers, cameras, Wi-Fi, Bluetooth, and microphone). Take these advantages of different sensors can effectively improve the quality of indoor map construction.

Considering that each type of sensor data has different characteristics, multiple data processing methods are usually required to build a complete map. Such as semantic labeling, landmark recognition from an image, and boundary extraction based on the alpha-shape. Therefore, one of the most important challenges of crowdsourcing systems is fusing the data from multi-sensors.

6.3. Incentive and standardization mechanism for data collection

As it's investigated in this paper, complex movement patterns of smartphones are carried by users that make it difficult for the underlying system to address the heading estimation issue. Besides, data collection consumes considerable energy from smartphones. For crowdsourcing indoor mapping solutions, challenges are quite different since these schemes need to apply satisfying incentive mechanisms to encourage users for data collection and contribution. Multiple studies have proposed incentive mechanisms for crowdsourcing-based applications (Jaimes et al., 2015; Li et al., 2018; Nie et al., 2019; Tian et al., 2017; Wang et al., 2018b; Zhang et al., 2016). For indoor localization, Li et al. proposed an incentive mechanism for crowdsourcing-based WiFi fingerprinting (Li et al., 2018). In addition to device positions, these systems should struggle with the human body position as well to handle location jumping and map rotation. Improving the data quality and design quality-based incentive rules for crowdsourcing-based indoor mapping systems will be an important research direction.

Recently, interactive mode-based indoor-map-construction approaches have been proposed (Chen et al, 2015a; MagicPlan, 2019; Sankar and Seitz, 2012). IndoorCrowd2D creates an interactive interface that allows users to input building floor information (Chen et al, 2015a), while Sankar and Seitz propose an interactive photogrammetric modeling-based smartphone application for indoor-map construction (Sankar and Seitz, 2012). Also, MagicPlan is a commercial indoor-map-generation application that can estimate the dimensions of the room and generate a corresponding map by marking the room corners via an augmented reality interface (MagicPlan, 2019). The data collected by these interactive mode-based applications are of high quality. Thus, designing a quick start guide and interactive interface for crowdsourced data collection will be helpful to improve the indoor mapping data quality.

7. Conclusions

With the development of ILBS, indoor-map construction has attracted enormous interest in both academic and industrial communities. The latest smartphones are equipped with various sensors, such as inertial sensors, WiFi, cameras, and microphones. The existence of such sensors makes smartphones a low-cost and up-to-date spatial data source, which is especially suitable for crowdsourcing-based indoor mapping. Numerous studies have been proposed during the last decade; however, a systematic review is lacked. In this paper, we survey the state-of-the-art crowdsourcing-based indoor mapping techniques via smartphones. We investigate the general process of crowdsourcing-based indoor mapping and highlight the key steps. Within these steps, we discuss and compare the functionality, advantages, and drawbacks of existing systems. Furthermore, we discussed the performance evaluation, open issues, and future research directions. Finally, we expect that this study will provide a useful perspective for recent crowdsourcing-based indoor mapping techniques using smartphones and promote its future development.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments:

This work was supported in part by National Key R&D Program of China (2016YFB0502203); Shenzhen Scientific Research and Development Funding Program (JCYJ20190808113603556, KQTD20180412181337494); Guangdong Basic and Applied Basic Research Foundation (2019A1515011910); National Natural Science Foundation of China (42001324).

[4] http://evaal.aaloa.org/

[5] https://www.microsoft.com/en-us/research/event/microsoft-indoor-localization-competition-ipsn-2017/

[6] https://perfloc.nist.gov/

References

Alzantot, M., Youssef, M., 2012. Crowdinside: automatic construction of indoor floorplans. In: Proceedings of the 20th International Conference on Advances in Geographic Information Systems. ACM, pp. 99–108.

Bailey, T., Durrant-Whyte, H., 2006. Simultaneous localization and mapping (SLAM): Part II. IEEE Robotics Automation Magazine. 13 (3), 108–117.

Borg, I., Groenen, P., 2003. Modern multidimensional scaling: Theory and applications. J. Educ. Meas. 40 (3), 277–280.

Brajdic, A., Harle, R., 2013. Walk detection and step counting on unconstrained smartphones. In: Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing. ACM, pp. 225–234.

Bronstein, A.M., Bronstein, M.M., Kimmel, R., 2006. Generalized multidimensional scaling: a framework for isometry-invariant partial surface matching. Proc. Natl. Acad. Sci. 103 (5), 1168–1172.

Cadena, C., Carlone, L., Carrillo, H., Latif, Y., Scaramuzzaet, D., Neira, J., Reid, I., Leonard, J.J., 2016. Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age. IEEE Trans. Rob. 32 (6), 1309–1332.

Chen, H., Li, F., Hei, X., Wang, Y.u., 2018. CrowdX: enhancing automatic construction of indoor floorplan with opportunistic encounters. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 2 (4), 1–21. https://doi.org/10.1145/3287037.

Chen, S., Li, M., Ren, K., Fu, X., Qiao, C., 2015a. Rise of the indoor crowd: Reconstruction of building interior view via mobile crowdsourcing. In: Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems, ACM, pp. 59–71.

Chen, S., Li, M., Ren, K., Qiao, C., 2015b. Crowd map: Accurate reconstruction of indoor floor plans from crowdsourced sensor-rich videos. In: Distributed Computing Systems (ICDCS), 2015 IEEE 35th International Conference on, IEEE, pp. 1–10.

Connolly, P., Boone, D., 2013. Indoor location in retail: Where is the money. Business Models Analysis Report.

Costa, J.A., Patwari, N., Hero III, A.O., 2006. Distributed weighted-multidimensional scaling for node localization in sensor networks. ACM Trans. Sens. Netw. 2 (1), 39–64.

Zhou, B., Zheng, T., Huang, J., Zhang, Y., Tu, W., Li, Q., 2021. A pedestrian network construction system based on crowdsourced walking trajectories. IEEE Internet Things J. 8 (9), 7203–7213.

Dai, J., Li, Y., He, K., Sun, J., 2016. R-fcn: Object detection via region-based fully convolutional networks. In: Advances in neural information processing systems, pp. 379-387.

Davidson, P., Piche, R., 2017. A Survey of Selected Indoor Positioning Methods for Smartphones. IEEE Commun. Surv. Tutorials 19 (2), 1347–1370.

Dong, J., Noreikis, M., Xiao, Y., Yla-Jaaski, A., 2018. ViNav: A Vision-based Indoor Navigation System for Smartphones. IEEE Transactions on Mobile Computing. 18(6), 1461-1475.

Edelsbrunner, H., Kirkpatrick, D., Seidel, R., 1983. On the shape of a set of points in the plane. IEEE Trans. Inf. Theory 29 (4), 551–559.

Elhamshary, M.M., Alzantot, M.F., Youssef, M., 2018. JustWalk: A crowdsourcing approach for the automatic construction of indoor floorplans. IEEE Trans. Mob. Comput. 18 (10), 2358–2371.

Elhamshary, M., Youssef, M., 2015. SemSense: Automatic construction of semantic indoor floorplans. In: Indoor Positioning and Indoor Navigation (IPIN), 2015 International Conference on, IEEE, pp. 1–11.

Endres, F., Hess, J., Sturm, J., Cremers, D., Burgard, W., 2014. 3-D mapping with an RGB-D camera. IEEE Trans. Rob. 30 (1), 177–187.

Fan, X., Liu, J., Wang, Z., Jiang, Y., Liu, X., 2017. Crowdsourcing in ITS: The state of the work and the networking. 55(6), 126-128.

Förstner, W., 1986. A feature based correspondence algorithm for image matching. ISPRS Com III, Rovaniemi, pp. 150–166.

Ganti, R.K., Ye, F., Lei, H., 2011. Mobile crowdsensing: current state and future challenges. IEEE Commun. Mag. 49 (11), 32–39.

Gao, R., Zhao, M., Ye, T., Ye, F., Luo, G., Wang, Y., Bian, K., Wang, T., Li, X., 2016. Multi-story indoor floor plan reconstruction via mobile crowdsensing. IEEE Trans. Mob. Comput. 15 (6), 1427–1442.

Gao, R., Zhao, M., Ye, T., Ye, F., Wang, Y., Bian, L., Wang, T., Li, X., 2014. Jigsaw: Indoor floor plan reconstruction via mobile crowdsensing. In: In: Proceedings of the 20th annual international conference on Mobile computing and networking. ACM, pp. 249–260.

Gilliéron, P.-Y., Merminod, B., 2003. Personal navigation system for indoor applications. 24.

Graham, D., Simmons, G., Nguyen, D.T., Zhou, G., 2015. A software-based sonar ranging sensor for smart phones. IEEE Internet Things J. 2 (6), 479–489.

Grisetti, G., Kummerle, R., Stachniss, C., Burgard, W., 2010. A tutorial on graph-based SLAM. IEEE Intell. Transp. Syst. Mag. 2 (4), 31–43.

Guo, B., Han, Q., Chen, H., Shangguan, L., Zhou, Z., Yu, Z., 2017. The emergence of visual crowdsensing: challenges and opportunities. IEEE Commun. Surv. Tutorials 19 (4), 2526–2543.

Guo, B., Wang, Z., Yu, Z., Wang, Y., Yen, N.Y., Huang, R., Zhou, X., 2015. Mobile crowd sensing and computing: The review of an emerging human-powered sensing paradigm. ACM Comput. Surv. 48 (1), 7.

Guo, X., Chan, E.C., Liu, C., Wu, K., Liu, S., Ni, L.M., 2014. Shopprofiler: Profiling shops with crowdsourcing data. In INFOCOM, IEEE, pp. 1240–1248.

Han, D., Jung, S., Lee, M., Yoon, G., 2014. Building a practical Wi-Fi-based indoor navigation system. IEEE Pervasive Comput. 13 (2), 72–79.

Harle, R., 2013. A survey of indoor inertial positioning systems for pedestrians. IEEE Commun. Surv. Tutorials 15 (3), 1281–1293.

He, Y., Liang, J., Liu, Y., 2017. Pervasive floorplan generation based on only inertial sensing: feasibility, design, and implementation. IEEE J. Sel. Areas Commun. 35 (5), 1132–1140.

Hedau, V., Hoiem, D., Forsyth, D., 2009. Recovering the spatial layout of cluttered rooms. In: 2009 IEEE 12th International Conference on Computer Vision. Kyoto, Japan.

Heipke, C., 2010. Crowdsourcing geospatial data. ISPRS J. Photogramm. Remote Sens. 65 (6), 550–557.

Hilsenbeck, S., Bobkov, D., Schroth, G., Huitl, R., Steinbach, E., 2014. Graph-based data fusion of pedometer and WiFi measurements for mobile indoor positioning. In: Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing. ACM, pp. 147–158.

Hosseini, M., Shahri, A., Phalp, K., Taylor, J., Ali, R., 2015. Crowdsourcing: A taxonomy and systematic mapping study. Comput. Sci. Rev. 17, 43–69.

Isikdag, U., Zlatanova, S., Underwood, J., 2013. A BIM-Oriented Model for supporting indoor navigation requirements. Comput., Environ., Urban Syst. 41, 112–123.

Ismail, H., Balachandran, B., 2015. Algorithm fusion for feature extraction and map construction from SONAR data. IEEE Sens. J. 15 (11), 6460–6471.

Jaimes, L.G., Vergara-Laurens, I.J., Raij, A., 2015. A survey of incentive techniques for mobile crowd sensing. IEEE Internet Things J. 2 (5), 370–380.

Jain, A., Kanhangad, V., 2017. Human activity classification in smartphones using accelerometer and gyroscope sensors. IEEE Sens. J. 18 (3), 1169–1177.

Jiang, Y., Xiang, Y., Pan, X., Li, K., Lv, Q., Dick, R.P., Shang, L., 2013. Hallway based automatic indoor floorplan construction using room fingerprints. In: Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing, ACM, pp. 315–324.

Judd, T., 1997. A personal dead reckoning module. ION GPS. 97, 1–5.

Kang, Z., Yang, J., Yang, Z., Cheng, S., 2020. A review of techniques for 3D reconstruction of indoor environments. ISPRS Int. J. Geo-Inf. 9 (5), 1–31.

Karam, S., Vosselman, G., Peter, M., Hosseinyalamdary, S., Lehtola, V., 2019. Design, calibration, and evaluation of a backpack indoor mobile mapping system. Remote Sens. 11 (8), 905.

Khoshelham, K., Elberink, S.O., 2012. Accuracy and resolution of kinect depth data for indoor mapping applications. Sensors 12 (2), 1437–1454.

Koo, J., Cha, H., 2012. Unsupervised locating of WiFi access points using smartphones. IEEE Trans. Syst. Man, Cybernetics. Part C. 42 (6), 1341–1353.

Ladetto, Q., Merminod, B., 2002. Digital magnetic compass and gyroscope integration for pedestrian navigation. 9th international conference on integrated navigation systems, Stpetersburg, Citeseer.

Lane, N.D., Miluzzo, E., Lu, H., Peebles, D., Choudhury, T., Campbell, A.T., 2010. A survey of mobile phone sensing. IEEE Commun. Mag. 48 (9), 140–150.

Lara, O.D., Labrador, M.A., 2013. A survey on human activity recognition using wearable sensors. IEEE Commun. Surv. Tutorials 13 (5), 1192–1209.

Lashkari, B., Rezazadeh, J., Farahbakhsh, R., Sandrasegaran, K., 2019. Crowdsourcing and sensing for indoor localization in IoT - a review. IEEE Sens. J. 19 (7), 2408–2434.

Lauterbach, H.A., Borrmann, D., Heß, R., Eck, D., Schilling, K., Nüchter, A., 2015. Evaluation of a backpack-mounted 3D mobile scanning system. Remote Sens. 7 (10), 13753–13781.

Levi, R.W., Judd, T., 1996. Dead reckoning navigational system using accelerometer to measure foot impacts, ed: Google Patents.

Li, J., Zhu, Y., Hua, Y., Yu, J., 2017a. Crowdsourcing sensing to smartphones: A randomized auction approach. IEEE Trans. Mob. Comput. 16 (10), 2764–2777.

Li, W., Wang, C., Zai, D., Huang, P., Li, P., 2017b. An efficiently volumetric fusing method for structure-frome-motion and terrestrial point cloud. Geoscience & Remote Sensing Symposium.

Li, W., Zhang, C., Liu, Z., Tanaka, Y., 2018. Incentive mechanism design for crowdsourcing-based indoor localization. IEEE Access 6, 54042–54051.

Li, Y., Khoshelham, K., Sarvi, M., Haghani, M., 2019. Direct generation of level of service maps from images using convolutional and long short-term memory networks. J. Intelligent Transp. Syst. 1–9.

Lin, Y.-H., Liu, Y.-S., Gao, G., Han, X.-G., Lai, C.-Y., Gu, M., 2013. The IFC-based path planning for 3D indoor spaces. Adv. Eng. Inf. 27 (2), 189–205.

Lipuš, B., Žalik, B., 2019. 3D convex hull-based registration method for point cloud watermark extraction. Sensors 19 (15), 3268.

Liu, H., Shi, R., Zhu, L., Jing, C., 2014. Conversion of model file information from IFC to GML. In: Geoscience and Remote Sensing Symposium (IGARSS), pp. 3133–3136.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., Berg, A.C., 2016. SSD: Single shot multibox detector. In: European conference on computer vision, Springer, pp. 21–37.

Liu, X., Zhou, B., Huang, P., Xue, W., Li, Q., Zhu, J., Qiu, L., 2021. Kalman filter-based data fusion of Wi-Fi RTT and PDR for indoor localization. IEEE Sens. J. 21 (6), 8479–8490.

Lohan, E., Torres-Sospedra, J., Leppäkoski, H., Richter, P., Peng, Z., Huerta, J., 2017. Wi-Fi crowdsourced fingerprinting dataset for indoor positioning. Data. 2 (4), 32.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision 60 (2), 91–110.

Luo, R.C., Lai, C.C., 2012. Enriched indoor map construction based on multisensor fusion approach for intelligent service robot. IEEE Trans. Ind. Electron. 59 (8), 3135–3145.

Luo, C., Hong, H., Chan, M.C., 2014. PiLoc: A self-calibrating participatory indoor localization system. In: Information Processing in Sensor Networks, IPSN-14 Proceedings of the 13th International Symposium on, IEEE, pp. 143–153.

Ma, H., Zhao, D., Yuan, P., 2014. pportunities in mobile crowd sensing. IEEE Commun. Mag. 52 (8), 29–35.

Ma, W., Li, Q., Zhou, B., Xue, W., Huang, Z., 2020. Location and 3D visual awareness-based dynamic texture updating for indoor 3D model. IEEE Internet Things J. 7 (8), 7612–7624.
MagicPlan, 2019. https://www.magicplan.app/.
Mao, Q., Zhou, B., Zou, Q., Li, Q., 2015. Efficient and lossless compression of raster maps. Signal, Image, Video Processing. 9 (1), 133–145.
Mendoza-Silva, G.M., Matey-Sanz, M., Torres-Sospedra, J., Huerta, J., 2019. BLE RSS measurements dataset for research on accurate indoor positioning. Data 4 (1), 12.
Mendoza-Silva, G., Richter, P., Torres-Sospedra, J., Lohan, E., Huerta, J., 2018. Long-term WiFi fingerprinting dataset for research on robust indoor positioning. Data. 3 (1), 3.
Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Tuytelaars, T., Van Gool, L., A comparison of affine region detectors. Int. J. Comput. Vision. 65, 43-72.
Nie, J., Luo, J., Xiong, Z., Niyato, D., Wang, P., 2019. A stackelberg game approach toward socially-aware incentive mechanisms for mobile crowdsensing. IEEE Trans. Wireless Commun. 18 (1), 724–738.
Noreikis, M., Xiao, Y., Hu, J., Chen, Y., 2018. SnapTask: towards efficient visual crowdsourcing for indoor mapping. In: 2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS), IEEE, pp. 578–588.
Peng, P., Shen, G., Zhang, Y., Li, Y., Tan, K., 2007. Beepbeep: a high accuracy acoustic ranging system using cots mobile devices. In: Proceedings of the 5th international conference on Embedded networked sensor systems. ACM, pp. 1–14.
Peng, Z., Gao, S., Xiao, B., Guo, S., Yang, Y., 2018. CrowdGIS: Updating digital maps via mobile crowdsensing. IEEE Trans. Automation Sci. Eng. 15 (1), 369–380.
Pipelidis, G., Prehofer, C., Gerostathopoulos, I., 2017. Bootstrapping the Dynamic Generation of Indoor Maps with Crowdsourced Smartphone Sensor Data. In: International Conference on Geographical Information Systems Theory, Applications and Management. Springer, pp. 70-84.
Pradhan, S., Baig, G., Mao, W., Qiu, L., Chen, G., Yang, B., 2018. Smartphone-based Acoustic Indoor Space Mapping. Proceedings of the ACM on Interactive. Mobile. Wearable Ubiquitous Technologies. 2 (2), 75.
Pratama, A.R., Hidayat, R., 2012. Smartphone-based pedestrian dead reckoning as an indoor positioning system. In: System Engineering and Technology (ICSET), 2012 International Conference on, IEEE, pp. 1–6.
Priyantha, N.B., Balakrishnan, H., Demaine, E.D., Teller, S.J., 2003. Anchor-free distributed localization in sensor networks. SenSys. 3, 340–341.
Qiu, C., Mutka, M.W., 2017. iframe: Dynamic indoor map construction through automatic mobile sensing. Pervasive Mobile Computing. 38, 346–362.
Rai, A., Chintalapudi, K.K., Padmanabhan, V.N., Sen, R., 2012. Zee: Zero-effort crowdsourcing for indoor localization. In: Proceedings of the 18th annual international conference on Mobile computing and networking. ACM, pp. 293–304.
Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. In: Advances in neural information processing systems, pp. 91-99.
Renaudin, V., Susi, M., Lachapelle, G., 2012. Step length estimation using handheld inertial sensors. Sensors 12 (7), 8507–8525.
Ronao, C.A., Cho, S.B., 2016. Human activity recognition with smartphone sensors using deep learning neural networks. Expert Syst. Appl. 59, 235–244.
Saeed, N., Nam, H., Al-Naffouri, T.Y., Alouini, M.-S., 2019. A state-of-the-art survey on multidimensional scaling based localization techniques. IEEE Commun. Surv. Tutorials 21 (4), 3565–3583.
Sampath, A., Shan, J., 2007. Building boundary tracing and regularization from airborne LiDAR point clouds. Photogrammetric Eng. Remote Sens. 73 (7), 805–812.
Sankar, A., Seitz, S., 2012. Capturing indoor scenes with smartphones. In: Proceedings of the 25th annual ACM symposium on User interface software and technology, ACM, pp. 403–412.
Seet, B.-C., Zhang, Q., Foh, C.H., Fong, A.C.M., 2012. Hybrid RF Mapping and Kalman Filtered Spring Relaxation for Sensor Network Localization. IEEE Sens. J. 12 (5), 1427–1435.
Shang, Y., Ruml, W., 2004. Improved MDS-based localization. In: IEEE INFOCOM 2004, IEEE. 4, 2640-2651.
Shen, G., Chen, Z., Zhang, P., Moscibroda, T., Zhang, Y., 2013. Walkie-Markie: Indoor pathway mapping made easy. In: Proceedings of the 10th USENIX conference on Networked Systems Design and Implementation, USENIX Association, pp. 85-98.
Shin, H., Chon, Y., Cha, H., 2012. Unsupervised construction of an indoor floor plan using a smartphone. IEEE Trans. Systems, Man, Cybernetics, Part C. 42 (6), 889–898.
Smith, R., 2007. An overview of the Tesseract OCR engine. In: Ninth International Conference on Document Analysis and Recognition (ICDAR 2007), IEEE. 2, 629-633.
Snavely, N., Simon, I., Goesele, M., Szeliski, R., Seitz, S.M., 2010. Scene reconstruction and visualization from community photo collections. Proc. IEEE 98 (9), 1370–1390.
Stove, A.G., 1992. Linear FMCW radar techniques. IEE Proceedings F (Radar and Signal Processing) 139 (5), 343–350.
Surmann, H., Nüchter, A., Hertzberg, J., 2003. An autonomous mobile robot with a 3D laser range finder for 3D exploration and digitalization of indoor environments. Robotics Autonomous Syst. 45 (3–4), 181–198.
Szeliski, R., 2010. Computer vision: algorithms and applications. Springer Science & Business Media.
Tang, L., Yang, X., Dong, Z., Li, Q., 2016. CLRIC: collecting lane-based road information via crowdsourcing. IEEE Trans. Intell. Transp. Syst. 17 (9), 2552–2562.
Tardós, J.D., Neira, J., Newman, P.M., Leonard, J.J., 2002. Robust mapping and localization in indoor environments using sonar data. Int. J. Robotics Res. 21 (4), 311–330.
Teng, X., Guo, D., Guo, Y., Zhao, X., Liu, Z., 2018. SISE: self-updating of indoor semantic floorplans for general entities. IEEE Trans. Mob. Comput. 17 (11), 2646–2659.
Tian, F., Liu, B., Sun, X., Zhang, X., Cao, G., Gui, L., 2017. Movement-based incentive for crowdsourcing. IEEE Trans. Veh. Technol. 66 (8), 7223–7233.
Toth, C., Jozkow, G., 2016. Remote sensing platforms and sensors: A survey. ISPRS J. Photogramm. Remote Sens. 115, 22–36.
Turner, E., Cheng, P., Zakhor, A., 2015. Fast, automated, scalable generation of textured 3d models of indoor environments. IEEE J. Sel. Top. Signal Process. 9 (3), 409–421.
Volk, R., Stengel, J., Schultmann, F., 2014. Building Information Modeling (BIM) for existing buildings—Literature review and future needs. Autom. Constr. 38, 109–127.
Wang, L., Zhang, D., Wang, Y., Chen, C., Han, X., M'hamed, A., 2016a. Sparse mobile crowdsensing: challenges and opportunities. IEEE Commun. Magazine. 54(7), 161-167.
Wang, T., Zhang, D., Zheng, Y., Gu, T., Zhou, X., Dorizzi, B., 2018a. C-FMCW Based Contactless Respiration Detection Using Acoustic Signal. Proceedings of the ACM on Interactive, Mobile, Wearable Ubiquitous Technologies. 1 (4), 170.
Wang, X., Zheng, X., Zhang, Q., Wang, T., Shen, D., 2016b. Crowdsourcing in ITS: The state of the work and the networking. IEEE Trans. Intell. Transp. Syst. 17 (6), 1596–1605.
Wang, Y., Cai, Z., Tong, X., Gao, Y., Yin, G., 2018b. Truthful incentive mechanism with location privacy-preserving for mobile crowdsourcing systems. Comput. Netw. 135, 32–43.
Wang, Y., Liu, X., Wei, H., Forman, G., Chen, C., Zhu, Y., 2013. Crowdatlas: Self-updating maps for cloud and personal use. In: Proceeding of the 11th annual international conference on Mobile systems, applications, and services. ACM, pp. 27–40.
Wen, C., Pan, S., Wang, C., Li, J., 2016. An indoor backpack system for 2-D and 3-D mapping of building interiors. IEEE Geosci. Remote Sens. Lett. 13 (7), 992–996.
Wu, C., Yang, Z., Liu, Y., 2015. Smartphones based crowdsourcing for indoor localization. IEEE Trans. Mob. Comput. 14 (2), 444–457.
Yang, J., Cheng, K., Chen, J., Zhou, B., Li, Q., 2018. Smartphones based Online Activity Recognition for Indoor Localization using Deep Convolutional Neural Network. In: 2018 Ubiquitous Positioning. Indoor Navigation and Location-Based Services (UPINLBS), IEEE, pp. 1–7.
Yang, Z., Wu, C., Liu, Y., 2012. Locating in fingerprint space: wireless indoor localization with little human intervention. In: Proceedings of the 18th annual international conference on Mobile computing and networking, ACM, pp. 269–280.
Ye, Q., Doermann, D., 2015. Text detection and recognition in imagery: A survey. IEEE Trans. Pattern Anal. Machine Intelligence 37 (7), 1480–1500.
Yu, W., Amigoni, F., 2014. Standard for robot map data representation for navigation. In: IROS2014 (IEEE/RSJ International Conference on Intelligent Robots and Systems) Workshop on "Standardized Knowledge Representation and Ontologies for Robotics and Automation", pp. 3–4.
Yürür, Ö., Liu, C.H., Sheng, Z., Leung, V.C., Moreno, W., Leung, K.K., 2014. Context-awareness for mobile sensing: A survey and future directions. IEEE Commun. Surveys Tutorials 18 (1), 68–93.
Zang, K., Shen, J., Huang, H., Wan, M., Shi, J., 2018. Assessing and mapping of road surface roughness based on GPS and accelerometer sensors on bicycle-mounted smartphones. Sensors 18 (3), 914.
Zhang, D., Liang, F., Yang, B., Xu, Y., Zang, Y., Li, J., Wang, Y., Dai, W., Fan, H., Hyyppä, J., Stilla, H., 2020. Registration of large-scale terrestrial laser scanner point clouds: A review and benchmark. ISPRS J. Photogramm. Remote Sens. 163, 327–342.
Zhang, Q., Foh, C.H., Seet, B.C., Fong, A.C., 2010. Location estimation in wireless sensor networks using spring-relaxation technique. Sensors (Basel). 10 (5), 5171–5192.
Zhang, X., Yang, Z., Sun, W., Liu, Y., Tang, S., Xing, K., Mao, X., 2016. Incentives for mobile crowd sensing: A survey. IEEE Communications Surveys Tutorials. 18 (1), 54–67.
Zhou, B., Elbadry, M., Gao, R., Ye, F., 2017a. BatMapper: acoustic sensing based indoor floor plan construction using smartphones. In: Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services. ACM, pp. 42-55.
Zhou, B., Elbadry, M., Gao, R., Ye, F., 2019a. Towards scalable indoor map construction and refinement using acoustics on smartphones. IEEE Trans. Mob. Comput. 19 (1), 217–230.
Zhou, B., Li, Q., Mao, Q., Tu, W., 2017b. A robust crowdsourcing-based indoor localization system. Sensors. 17 (4), 864.
Zhou, B., Li, Q., Mao, Q., Tu, W., Zhang, X., 2015a. Activity sequence-based indoor pedestrian localization using smartphones. IEEE Trans. Hum.-Mach. Syst. 45 (5), 562–574.
Zhou, B., Li, Q., Mao, Q., Tu, W., Zhang, X., Chen, L., 2015b. ALIMC: Activity landmark-based indoor mapping via crowdsourcing. IEEE Trans. Intell. Transp. Syst. 16 (5), 2774–2785.
Zhou, B., Li, Q., Zhai, G., Mao, Q., Long, C., 2018. A graph optimization-based indoor map construction method via crowdsourcing. IEEE Access 6, 33692–33701.
Zhou, B., Yang, J., Li, Q., 2019b. Smartphone-based activity recognition for indoor localization using a convolutional neural network. Sensors 9 (3), 621.