

VWR-SLAM: Tightly Coupled SLAM System Based on Visible Light Positioning Landmark, Wheel Odometer, and RGB-D Camera

Junlin Huang^{ID}, Shangsheng Wen^{ID}, Wanlin Liang^{ID}, and Weipeng Guan^{ID}

Abstract—Visible light positioning (VLP) is a promising technology since it can provide high-accuracy indoor localization based on the existing lighting infrastructure. Most VLP systems require a prior light-emitting diode (LED) location map, termed a VLP-landmark map in this article, for which manual surveys are costly in practical deployment at scale. What is more, the existing approaches also require dense LED deployments. In this work, we proposed a multisensor fusion framework, termed VWR-simultaneous localization and mapping (SLAM), which tightly fused the VLP, wheel odometer, and red green blue-depth map (RGB-D) camera to achieve SLAM. Our VWR-SLAM can provide accurate and robust robot localization and navigation in LED shortage/outage situations, meanwhile, constructing the 3-D sparse environment map and the 3-D VLP-landmark map without tedious manual measurements. The experimental results show that our proposed scheme can provide an average robot positioning accuracy of 1.81 cm and an LED mapping accuracy of 3.01 cm.

Index Terms—Geometry-based method, robotics, sensor fusion, visible light positioning (VLP), visual simultaneous localization and mapping (SLAM).

I. INTRODUCTION

LOCATION is the most fundamental topic in the field of robotics, such as simultaneous localization and mapping (SLAM)/visual odometer (VO), navigation, path-planning, drone control, and autonomous driving. Also, indoor positioning is an especially challenging problem, where localization cannot be achieved by global positioning system (GPS) due to the satellite signal being greatly attenuated. Compared with the traditional radio-based indoor positioning technologies, such as Bluetooth, Wi-Fi, radio frequency Identification

Manuscript received 5 October 2022; revised 19 November 2022; accepted 6 December 2022. Date of publication 29 December 2022; date of current version 16 January 2023. This work was supported in part by the Guangdong Science and Technology Project under Grant 2017B010114001; in part by the National Undergraduate Innovative and Entrepreneurial Training Program under Grant 202010561158, Grant 202010561155, Grant 202110561162, Grant 202110561165, and Grant 202110561163; and in part by the Guangdong Provincial Training Program of Innovation and Entrepreneurship for Undergraduates under Grant S202010561272. The Associate Editor coordinating the review process was Dr. Jochen Lang. (Corresponding authors: Shangsheng Wen; Weipeng Guan.)

Junlin Huang is with the School of Materials Science and Engineering and the School of Mathematics, South China University of Technology, Guangzhou 510641, China.

Shangsheng Wen is with the School of Materials Science and Engineering, South China University of Technology, Guangzhou 510641, China (e-mail: shshwen@scut.edu.cn).

Wanlin Liang and Weipeng Guan are with the School of Automation Science and Engineering, South China University of Technology, Guangzhou 510641, China (e-mail: augwpsc@scut.edu.cn).

Digital Object Identifier 10.1109/TIM.2022.3231332

(RFID), and ultra-wideband (UWB), visible light positioning (VLP) has the advantages of high positioning accuracy, no electromagnetic interference, fewer additional modules, good security, and the possibility of combining communication with lighting. Thus, it has aroused the attention of many experts and scholars around the world. In VLP, light-emitting diodes (LED) act as artificial beacons in the environment, and each allows known data association using its identity (ID). The modulated LED broadcasts its unique ID via switching at a high frequency imperceptible to the human eye, but which can be recognized by a photodiode (PD) or camera sensor. The LED-ID can be mapped/modulated once and for all since they are normally fixed. Hence, the “Last Mile Problem” of localization is solved via VLP and the prebuilt VLP-landmark map. Some state-of-the-art (SOTA) camera-based VLP systems can offer centimeter-level accuracy on commodity smartphones [1], [2] or mobile robots [3], [4].

A. Motivation

Most VLP systems normally require multiple LED observations at a time for successful positioning through triangulation. This is because normal LED lights offer less usable point features due to the lack of distinguishable appearance, e.g., one feature for a circular LED [5], [6]. This might lead to the requirement for LED density deployment. To address this problem, different VLP-aided inertial localization methods have been proposed in Liang and Liu [5], Liang et al. [7], and Guan et al. [8] to provide pose estimation by combining VLP and the inertial measurement unit (IMU). To further eliminate the cumulative error caused by the IMU, Guan et al. [4] proposed the VLP and LiDAR fusion system to realize accurate and robust robot localization and navigation in LED shortage/outage situations.

However, all these works require a prior map composed of global LED locations in the environment alongside their identifiers for data association. In general, the locations of each LED are recorded manually, which is acceptable in small-scale environments and is the de facto way in many research works. However, for large-scale real-world applications, manual LED location recording is difficult and is prone to human errors due to the increase in LEDs and coverage. This causes bottlenecks in the development of the VLP. What is more, sensors are imperfect, and their measurements are prone to errors. Through multisensors’ fusion, we can compensate for the deficiencies of stand-alone sensors and provide more reliable and robust estimations.

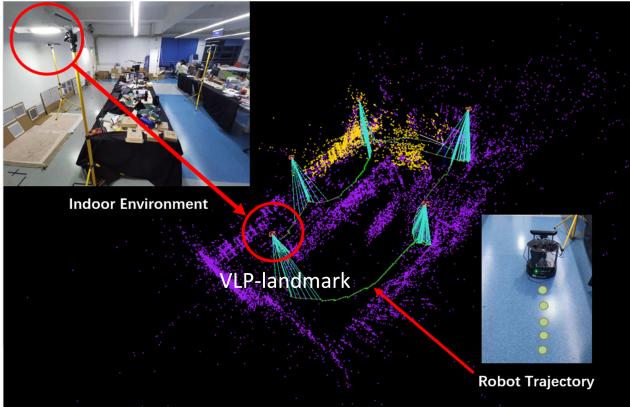


Fig. 1. Our proposed VWR-SLAM system can achieve highly accurate indoor localization and build the environment map for both VLP-Landmark and sparse point clouds.

To address the problems mentioned above, we propose a VWR-SLAM system, a tightly coupled multisensors' fusion scheme for VLP and visual SLAM. Integrating multisensors, including a wheel odometer, industrial camera, and red green blue-depth map (RGB-D) camera, enables a robot to obtain sufficient measurements, making it promising for LED mapping and robust indoor localization. To be more specific, this scheme entails a surveyor who controls a robot to wander around the indoor environment through a remote workstation. The robot equipped with a ceiling-facing industrial camera, termed a VLP-Cam, detects nearby light sources and their ID information. While an RGB-D camera, termed an RGBD-Cam, collects the images of the surrounding environment to perform the ORB-SLAM2 algorithm [9]. The inputs from the RGBD-Cam, the VLP-Cam, and the wheel odometer are integrated to be VLP-keyframe, which is used to calculate the positions of the LED and robot. As the position calculation scheme is based on a geometric method, the LED-ID can be wrongly decoded in our system, which is reckoned to be one of the contributions. The proposed system can build a 3-D sparse map of the environment, which can visualize the location of the LED and the high-accuracy real-time trajectory of the robot concurrently (an example is shown in Fig. 1). It is worth noting that the LED is termed VLP-landmark in this article, for its position information is unknown.

B. Contribution

Our contributions are summarized as follows.

- 1) *Autonomous VLP-Landmark Mapping:* A position calculation method is proposed to automatically obtain the localization of the VLP-landmark in the environment instead of a manual record. Unlike most VLP systems that require a prebuilt LED map, our scheme overcomes the shortages of inaccurate manual measurement and relieves the extensive human work at scale.
- 2) *Map Construction:* A map construction algorithm is proposed to build the VLP-landmark map and the 3-D sparse environment map by our VWR-SLAM framework. A video of the construction process has been posted online.¹

¹<https://b23.tv/ouLa8Zp>

- 3) *Sensor Fusion:* A tightly coupled multisensor fusion of camera, wheel odometer, and VLP using least-square optimization based on a geometric method is proposed to achieve high-accuracy localization of the VLP-landmark and robot, which can heavily relieve the strict requirement of accurate LED-ID decoding and enhance the robustness of the VLP system.

C. Organization

The remainder of the article is organized as follows: Section II introduces the related works. Section III introduces the principle of our proposed method. Section IV presents the experiments and results. Finally, the conclusion is given in Section V.

II. RELATED WORK

A. Camera-Based VLP Under LED Shortage Condition

The camera-based VLP uses modulated LED light to achieve localization by imaging geometry. It is favored in real applications for high-accuracy positioning performance and good compatibility with other devices, such as mobile robots and smartphones. However, the shortage/outage of LEDs can severely deteriorate the performance of the camera-based VLP in reality [4]. Considering the inferior performance of vision-only VLP under this poor condition, Liang and Liu [5] and Liang et al. [7] coupled VLP with the IMU, thus “compensating” the missing information due to the lack of LED beacons. Compared with the vision-only VLP systems, which require multiple LED observations at a time, Liang et al. [7] can reduce the number of LED observations to one. Using both the modulated and unmodulated LEDs, Liang et al. [10] proposed two types of blob features tightly integrated with the multistate constraint Kalman filter. This inertial-aided VLP system can provide satisfactory pose estimations in real-time. Although the above methods can achieve robot positioning in an LED shortage environment, they still need a prebuilt LED map surveyed manually, which is a shortcoming we aim to overcome by fusing multisensors into visual SLAM system.

Furthermore, the LED-ID decoding of VLP may encounter failure due to some unexpected situations, such as the unstable transmission channel and the asynchronous air conveying channel [11]. This case will also result in an LED shortage environment for the camera-based VLP systems. Different threshold schemes have been proposed for the decoding of the LED-ID [5], [12], but they still fail in adapting to different transmission channel conditions. To deal with the unstable LED-ID decoding, our scheme attempts to break away from the strict matching between accurately decoded LED-ID and the hardware facilities by a geometry-based method. For the readers' interest, we also refer to our website ² to demonstrate the works from our group for high-accuracy VLP localization and camera-based LED-ID decoding or optical camera communication.

²<https://b23.tv/FWQWPsg>

B. Autonomous LED Mapping

The manual site survey is the most common way to acquire an LED map with the help of measuring devices if LEDs reside on a precise frame. However, human measurements tend to cause errors for real applications at scale. At present, there are some schemes for nonhuman LED mapping. In [13] and [14], a calibration method for the camera-based VLP is proposed. To be more specific, a robot localized by the LiDAR-SLAM algorithm in [15] is controlled to approach the LED center so that the region of interest (ROI) can remain in the image center for a while, and the LED-ID can be decoded [14]. Although the 2-D positioning of the LED in this work reached centimeter-level accuracy, the height of the LED still needs a manual survey. Similarly, in [16], a PD-based light registration method has been proposed to measure the 2-D LED location, whereas the height of the LED still needs human measurements. Zhang et al. [17] use angle-of-arrival (AOA) estimators equipped with PD as the receiver to achieve centimeter-level LED mapping accuracy. However, these receivers may be disturbed by the thermal noise [17] and the estimators are fixed on a specific position and cannot be moved, resulting that the area range of LED distribution being limited. For solutions to large-scale manual work, Liang et al. [18] proposed a novel LED mapping system for modulated LEDs using visual-inertial (VI) sensors to solve a full-SLAM problem within a factor graph, which leads to much reduced human efforts. But it still needs a few known LED points for VLP alignment. Therefore, it is not completely automated.

To address the shortcomings of the above schemes, in this article, we propose the VWR-SLAM system which can conduct LED mapping work autonomously without any prior LED information. Compared with Amsters et al. [13], [14], and Zhang and Zhang [16], there is no need for our system to survey the height of the LED manually. The distribution of LED can be randomly sparse for VWR-SLAM, which is an advantage superior to Zhang et al. [17]. Moreover, the VWR-SLAM system outperforms the LedMapper system in Liang et al. [18], for VWR-SLAM needs no known LED points as beacons.

III. METHODOLOGY

A. Notation and Assumption

To facilitate discussion, we clarify some notations and treat LED as VLP-landmark defined before. The symbols shown in Fig. 2 will be widely adopted in the following description. A robot equipped with an RGB-D camera (RGBD-Cam) and an industrial camera (VLP-Cam) will be controlled to pass through different VLP-landmark illumination ranges. When a luminaire is observed, an ROI will form on the pixel plane of the VLP-Cam. The geometrical relationship between the VLP-Cam lens and the ROI is illustrated. We reckon the robot's position to be VLP-Cam's position, so the robot's trajectory corresponds to VLP-Cam's trajectory.

B. Overall Structure of VWR-SLAM System

Fig. 3 shows the architecture of the proposed VWR-SLAM system, which entails three main modules: front-end,

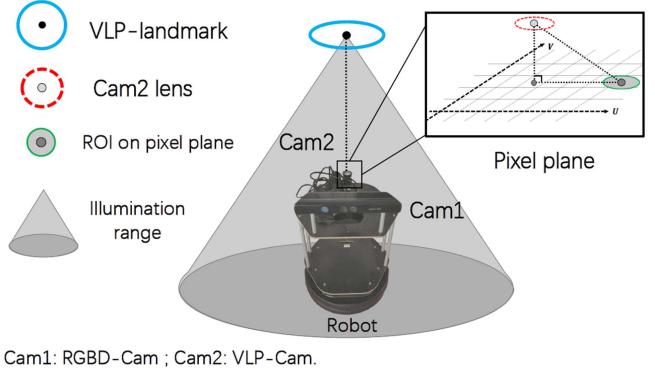


Fig. 2. VLP sensors suite and data collection.

optimization, and map construction. When the robot starts wandering indoor and collecting inputs, the VLP-Cam locates LED-ROI in the captured image for subsequent LED-ID decoding while the RGBD-Cam enables environment perception. The front-end module fuses LED-ROI and robot azimuth from the wheel odometer, providing VLP-feature, i.e., the integration block. The position of the VLP-Cam converted from RGBD-Cam's position and VLP-feature constitute VLP-keyframe jointly. Using VLP-keyframe as input, the optimization module will complete VLP-landmark mapping with the aid of a geometry-based method. The detected VLP-landmark position can be used for VLP-Cam's position correction by solving a least-squares problem. To acquire the robot's trajectory, the first step is to achieve loop closing by VLP-features. Next, the full optimization block uses the same scheme as the optimization section to improve the accuracy of all the positions of the robot and VLP-landmarks at once. After the full optimization section, the robot's trajectory from the VLP-keyframe and the position of the VLP-landmark will be updated. In the middle green block, the VLP-keyframes, VLP-landmarks' position, and MapPoints from ORB-SLAM2 constitute a map collectively, which mainly shows the distribution of VLP-landmarks and the trajectory of the robot in the 3-D real world. The VLP-feature and VLP-keyframe are two concatenations of input parameters, which will be introduced in Section III-C. To sum up, the proposed multisensor fusion for VWR-SLAM includes localization fusion and map fusion.

C. Front-End

In the following discussion, we use superscript $.^{fea}$ to denote the parameter belonging to VLP-feature. The front-end is mainly designed to obtain VLP-keyframe, based on the inputs from cameras and the wheel odometer. The VLP-keyframe is made up of VLP-feature and VLP-Cam position.

1) *Acquire VLP-Feature*: A novel data fusion scheme is proposed to obtain VLP-feature, as simply shown in Fig. 4. The VLP-Cam can capture barcode-like strip patterns from the high-frequency flashing VLP-landmark lightwave. We are interested in the image regions that carry visible light communication (VLC) information, termed as ROI. The feature distance d^{fea} , which describes the 2-D pixel distance from the VLP-Cam lens center to the ROI center, and the LED-ID

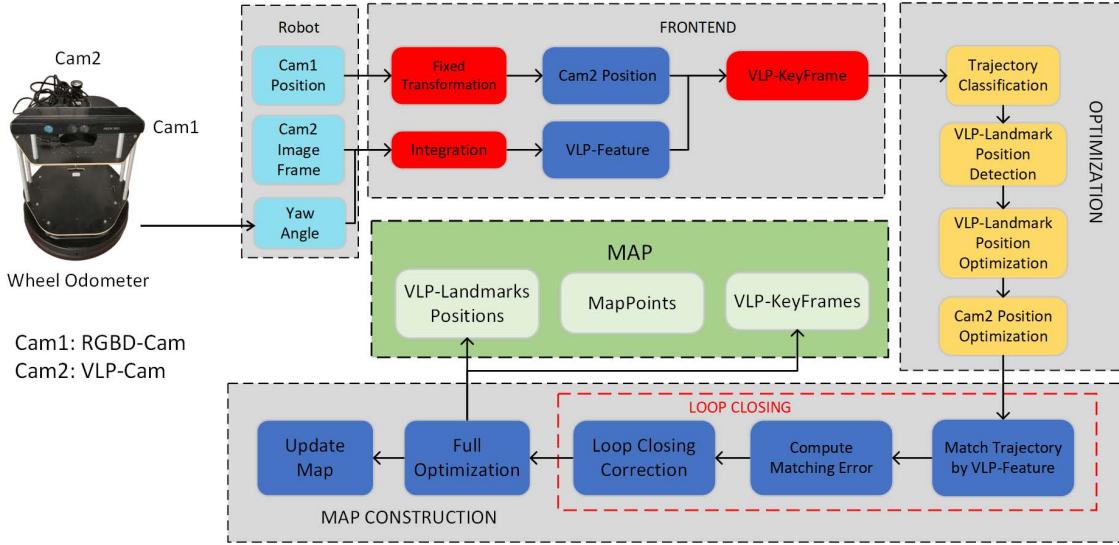
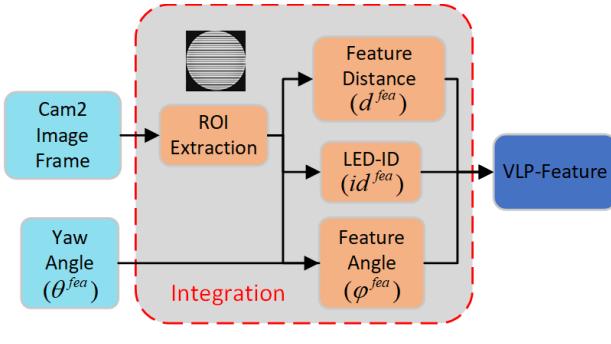


Fig. 3. Overall structure of the VWR-SLAM system.



Cam2: VLP-Cam

Fig. 4. Integration step in the VWR-SLAM system.

id^{fea} can be obtained from ROI. The scheme for decoding refers to previous work [11]. Coupled with the yaw angle θ^{fea} from the wheel odometer, the VLP-feature is defined as follows:

$$\begin{aligned} \text{fea}^{VLP} &\triangleq \begin{pmatrix} \sqrt{(u^{fea} - c_x)^2 + (v^{fea} - c_y)^2} \\ \arctan(u^{fea} - c_x, v^{fea} - c_y) - \theta^{fea} \\ id^{fea} \end{pmatrix} \\ &= \begin{pmatrix} d^{fea} \\ \varphi^{fea} \\ id^{fea} \end{pmatrix} \end{aligned} \quad (1)$$

where φ^{fea} is named as the feature angle, (c_x, c_y) represents the coordinate of the VLP-Cam's lens center projected on pixel plane, which is an intrinsic parameter, and (u^{fea}, v^{fea}) is the ROI center coordinate on VLP-Cam's pixel plane. Fig. 5 illustrates the parameters defined in VLP-feature. The significance of d^{fea} and φ^{fea} will manifest in Section III-E in terms of guaranteeing the uniqueness of the robot's position. Otherwise, if VLP-Cam cannot catch ROI, indicating that VLP-Cam is out of the VLP-landmark illumination range, VLP-feature is set to be a constant, which is invalid for the localization of

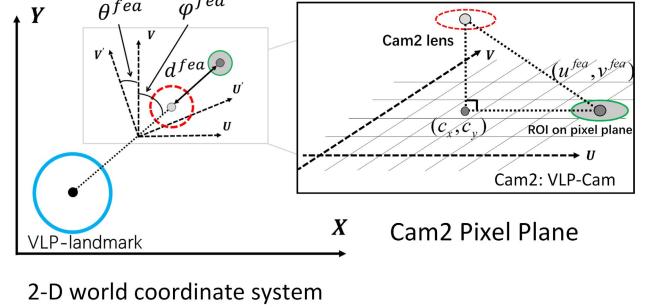


Fig. 5. Parameters defined in VLP-feature.

VLP-landmark and robot in our implementation. In this case, robot positioning merely depends on the ORB-SLAM2 system.

2) *Compute VLP-Cam Position*: First, the position of the VLP-Cam is defined as follows:

$$p \triangleq (x^{cam}, y^{cam}, H^{L1}) \quad (2)$$

where H^{L1} denotes the height of the VLP-Cam's lens. As H^{L1} can be manually measured beforehand, we only need to estimate the 2-D position of the VLP-Cam. For the sake of simplicity, we define the position of the VLP-Cam on the 2-D world coordinate system plane as follows:

$$\mathbb{X} \triangleq (x^{cam}, y^{cam}). \quad (3)$$

Because the observed position of RGBD-Cam can be obtained from ORB-SLAM2 directly, the preestimated position of VLP-Cam denoted by

$$\begin{aligned} \tilde{p} &\triangleq (\tilde{x}^{cam}, \tilde{y}^{cam}, H^{L1}) \\ \tilde{\mathbb{X}} &\triangleq (\tilde{x}^{cam}, \tilde{y}^{cam}) \end{aligned} \quad (4)$$

can be acquired by a fixed coordinate transformation between RGBD-Cam and VLP-Cam (Fig. 6). Each position of the

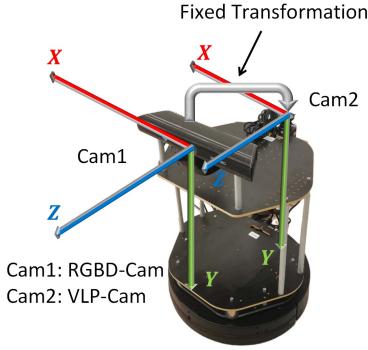
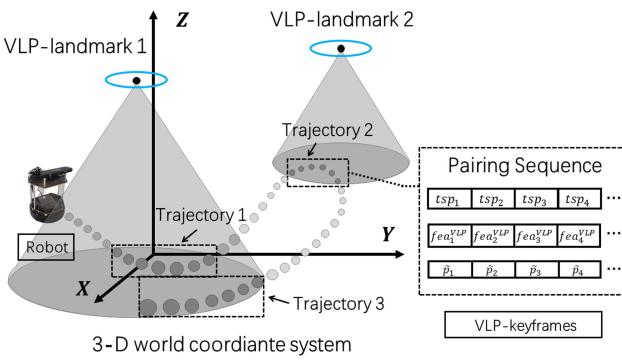


Fig. 6. Transformation between VLP-Cam and RGBD-Cam.

Fig. 7. Process of generating VLP-keyframe. tsp_i represents the i th timestamp.

VLP-Cam will be matched to the corresponding VLP-feature according to the same timestamp, which forms a pairing sequence (Fig. 7). Each element of the sequence is a VLP-keyframe key^{VLP} defined as follows:

$$\text{key}^{\text{VLP}} \triangleq \begin{pmatrix} \text{fea}^{\text{VLP}} \\ \tilde{p}^T \end{pmatrix} \quad (5)$$

namely, the output of the front-end.

D. Optimization

In this section, we leverage VLP-keyframe gained from the front-end to acquire and optimize VLP-landmark's position together with the VLP-Cam's position.

1) *Trajectory Classification*: When the robot passes through different irradiation ranges, each trajectory will be characterized by the unique LED-ID from VLP-landmark (Fig. 7). We assume that all the LED-IDs are different so that the classification step is feasible. In the following discussion, we will use subscript \cdot_i to denote the parameter belonging to the i th trajectory, and \cdot_{i_k} stands for the parameter belonging to the k th observation of the i th trajectory, in short, the i_k th observation. The set of VLP-keyframes in the i th trajectory is defined as follows:

$$\text{Traj}_i \triangleq \{\text{key}_{i_k}^{\text{VLP}} | k = 1, 2, \dots\} \quad (6)$$

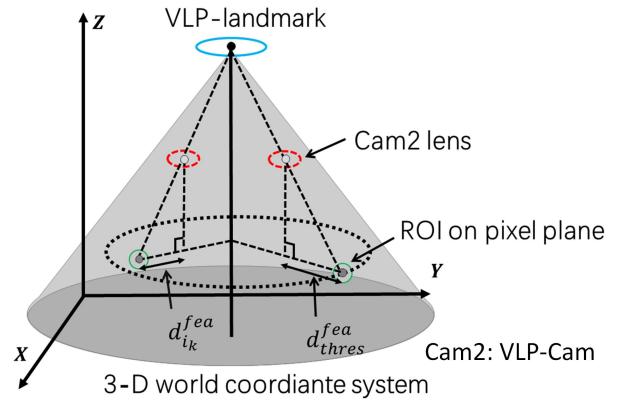


Fig. 8. Threshold method for VLP-landmark position detection.

where

$$\begin{aligned} \text{key}_{i_k}^{\text{VLP}} &= \begin{pmatrix} \text{fea}_{i_k}^{\text{VLP}} \\ \tilde{p}_{i_k} \end{pmatrix} \\ \text{fea}_{i_k}^{\text{VLP}} &= \begin{pmatrix} \sqrt{(u_{i_k}^{\text{fea}} - c_x)^2 + (v_{i_k}^{\text{fea}} - c_y)^2} \\ \arctan(u_{i_k}^{\text{fea}} - c_x, v_{i_k}^{\text{fea}} - c_y) - \theta_{i_k}^{\text{fea}} \\ id_{i_k}^{\text{fea}} \end{pmatrix} \\ &= \begin{pmatrix} d_{i_k}^{\text{fea}} \\ \varphi_{i_k}^{\text{fea}} \\ id_{i_k}^{\text{fea}} \end{pmatrix} \\ \tilde{p}_{i_k} &= (\tilde{x}_{i_k}^{\text{cam}}, \tilde{y}_{i_k}^{\text{cam}}, H^{L1}) \\ \tilde{X}_{i_k} &= (\tilde{x}_{i_k}^{\text{cam}}, \tilde{y}_{i_k}^{\text{cam}}). \end{aligned} \quad (7)$$

2) *VLP-Landmark Position Detection*: A threshold method is applied to VLP-landmark position detection. The 2-D coordinate of the VLP-landmark on the world coordinate system is denoted by $\mathbb{P} = (x, y)$. Then, the VLP-Cam position from the i_k th observation $\tilde{X}_{i_k} = (\tilde{x}_{i_k}^{\text{cam}}, \tilde{y}_{i_k}^{\text{cam}})$ is considered to be an alternative estimated position of the VLP-landmark $\mathbb{P}_{\text{alt}} = (x_{i_k}^{\text{alt}}, y_{i_k}^{\text{alt}})$ when $d_{i_k}^{\text{fea}}$ is smaller than an artificial threshold $d_{\text{thres}}^{\text{fea}}$. As shown in Fig. 8, when $d_{i_k}^{\text{fea}} < d_{\text{thres}}^{\text{fea}}$, the position of VLP-Cam will be reasonably considered to be an alternative estimated position of VLP-landmark, i.e., $\mathbb{P}_{i_k}^{\text{alt}} = \tilde{X}_{i_k}$. For the i_k th observation, there is a set $\{\mathbb{P}_{i_s}^{\text{alt}} | s = 1, 2, \dots, m\}$ composed of m alternative estimated positions ($m \leq k$). The Pauta criterion³ is applied to $\{\mathbb{P}_{i_s}^{\text{alt}} | s = 1, 2, \dots, m\}$ to acquire the roughly estimated position of VLP-landmark belonging to the i_k th observation $\tilde{\mathbb{P}}_{i_k} = (\tilde{x}_{i_k}, \tilde{y}_{i_k})$. To be brief, we use VLP-Cam's 2-D position to approximate VLP-landmark's 2-D position. The error, which we aim to eliminate by the least-squares method in the following discussion, is mainly caused by the threshold $d_{\text{thres}}^{\text{fea}}$.

3) *Validation Function*: A geometry-based method is proposed to acquire a validation function, which is useful in Section III-D4. Fig. 9 describes the parameters used in the validation function. The distance from the VLP-landmark center to the VLP-Cam lens center on the 2-D world coordinate system is named as the validation distance $D_{i_k}^{\text{val}}$ and the vertical angle of $\varphi_{i_k}^{\text{fea}}$ is called the validation angle $\varphi_{i_k}^{\text{val}}$.

³A well-known data processing principle, also referred to as 3σ principle.

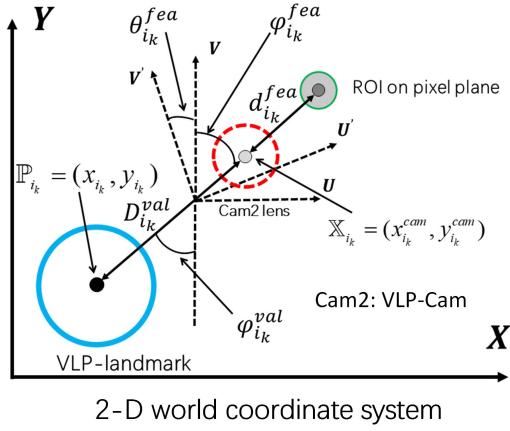


Fig. 9. Geometrical relationship between validation function's parameters and VLP-feature's parameters.

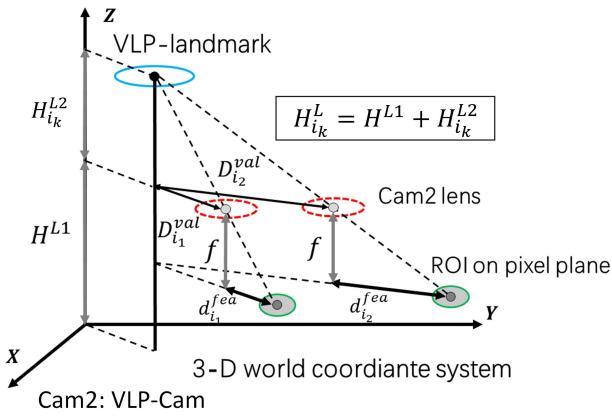


Fig. 10. Elevation relationship between VLP-landmark and VLP-Cam.

We introduce a parameter \mathbf{k}_i from the elevation relationship between VLP-landmark and VLP-Cam in Fig. 10

$$\mathbf{k}_i \triangleq \frac{f}{H_i^{L2}} \quad (8)$$

where f is the focal length of the VLP-Cam. Note that (f/H_i^{L2}) is a constant value for the i_{th} trajectory theoretically, and \mathbf{k}_i should be a constant value. By the determination of similar triangles, we have

$$\mathbf{k}_i = \frac{d_{i_1}^{fea}}{D_{i_1}^{val}} = \frac{d_{i_2}^{fea}}{D_{i_2}^{val}}. \quad (9)$$

In our system, however, the observed values of $(d_{i_k}^{fea}/D_{i_k}^{val})$ are different over time. Thus, we shall redefine $\mathbf{k}_i \triangleq \tilde{\mathbf{k}}_{i_k}$, i.e.,

$$\begin{aligned} \tilde{\mathbf{k}}_{i_k} &\triangleq \frac{d_{i_k}^{fea}}{D_{i_k}^{val}} \\ &= \frac{\sqrt{(u_{i_k}^{fea} - c_x)^2 + (v_{i_k}^{fea} - c_y)^2}}{\sqrt{(\tilde{x}_{i_k}^{cam} - \tilde{x}_{i_k})^2 + (\tilde{y}_{i_k}^{cam} - \tilde{y}_{i_k})^2}}. \end{aligned} \quad (10)$$

Similarly, by applying the Pauta criterion to the set $\{\mathbf{k}_{i_s} | s = 1, 2, \dots, k\}$, we can acquire a refined parameter $\tilde{\mathbf{k}}_{i_k}$. Note that f is an intrinsic parameter and H^{L1} is known in

advance, we can calculate the height of VLP-landmark in the i_k th observation $H_{i_k}^L$ with the following equation:

$$H_{i_k}^L = H^{L1} + H_{i_k}^{L2} = H^{L1} + \frac{f}{\tilde{\mathbf{k}}_{i_k}}. \quad (11)$$

Then, a validation function $val_{i_k}(\mathbb{X}_{i_k}, \mathbb{P}_{i_k})$ is designed to rectify the corresponding VLP-feature $fea_{i_k}^{VLP}$

$$\begin{aligned} val_{i_k}(\mathbb{X}_{i_k}, \mathbb{P}_{i_k}) &= \begin{pmatrix} \tilde{\mathbf{k}}_{i_k} * \sqrt{(x_{i_k}^{cam} - x_{i_k})^2 + (y_{i_k}^{cam} - y_{i_k})^2} \\ \arctan(x_{i_k}^{cam} - x_{i_k}, y_{i_k}^{cam} - y_{i_k}) \\ id_{i_k}^{val} \end{pmatrix} \\ &= \begin{pmatrix} \tilde{\mathbf{k}}_{i_k} * D_{i_k}^{val} \\ \varphi_{i_k}^{val} \\ id_{i_k}^{val} \end{pmatrix} \end{aligned} \quad (12)$$

where $\mathbb{X}_{i_k} = (x_{i_k}^{cam}, y_{i_k}^{cam})$ and $\mathbb{P}_{i_k} = (x_{i_k}, y_{i_k})$ are unknown variables. It is rational to artificially set $id_{i_k}^{val} = id_{i_k}^{fea}$. Noting that the parameter $\tilde{\mathbf{k}}_{i_k}$ becomes significant after several observations, the validation function is designed to be posterior.

4) Position Optimization: Two enhanced cost functions consisting of $fea_{i_k}^{VLP}$ and $val_{i_k}(\mathbb{X}_{i_k}, \mathbb{P}_{i_k})$, which will contribute to the optimization of VLP-Cam position and VLP-landmark position, respectively, are proposed. An error function is defined as follows:

$$\begin{aligned} \text{error}_{i_k}(\mathbb{X}_{i_k}, \mathbb{P}_{i_k}) &\triangleq \|fea_{i_k}^{VLP} - val_{i_k}(\mathbb{X}_{i_k}, \mathbb{P}_{i_k})\|_2 \\ &= \sqrt{\text{errdist}_{i_k}(\mathbb{X}_{i_k}, \mathbb{P}_{i_k})^2 + \text{errang}_{i_k}(\mathbb{X}_{i_k}, \mathbb{P}_{i_k})^2} \\ &\quad + (id_{i_k}^{fea} - id_{i_k}^{val})^2 \end{aligned} \quad (13)$$

where

$$\begin{aligned} \text{errdist}_{i_k}(\mathbb{X}_{i_k}, \mathbb{P}_{i_k}) &\triangleq d_{i_k}^{fea} - \tilde{\mathbf{k}}_{i_k} * D_{i_k}^{val} \\ &= \sqrt{(u_{i_k}^{fea} - c_x)^2 + (v_{i_k}^{fea} - c_y)^2} \\ &\quad - \tilde{\mathbf{k}}_{i_k} * \sqrt{(x_{i_k}^{cam} - x_{i_k})^2 + (y_{i_k}^{cam} - y_{i_k})^2} \end{aligned} \quad (14)$$

and

$$\begin{aligned} \text{errang}_{i_k}(\mathbb{X}_{i_k}, \mathbb{P}_{i_k}) &\triangleq \varphi_{i_k}^{fea} - \varphi_{i_k}^{val} \\ &= (\arctan(u_{i_k}^{fea} - c_x, v_{i_k}^{fea} - c_y) - \theta_{i_k}^{fea}) \\ &\quad - \arctan(x_{i_k}^{cam} - x_{i_k}, y_{i_k}^{cam} - y_{i_k}). \end{aligned} \quad (15)$$

Note that $id_{i_k}^{val} = id_{i_k}^{fea}$, so (13) can be rewritten as follows:

$$\begin{aligned} \text{error}_{i_k}(\mathbb{X}_{i_k}, \mathbb{P}_{i_k}) &= \sqrt{\text{errdist}_{i_k}(\mathbb{X}_{i_k}, \mathbb{P}_{i_k})^2 + \text{errang}_{i_k}(\mathbb{X}_{i_k}, \mathbb{P}_{i_k})^2}. \end{aligned} \quad (16)$$

We substitute $\tilde{\mathbb{X}}_{i_k}$ for \mathbb{X}_{i_k} in (16) and propose an enhanced cost function

$$J_1(\mathbb{P}_{i_k}) = \sum_{l=1}^k \frac{1}{2} \|\text{error}_{i_l}(\tilde{\mathbb{X}}_{i_l}, \mathbb{P}_{i_k})\|^2. \quad (17)$$

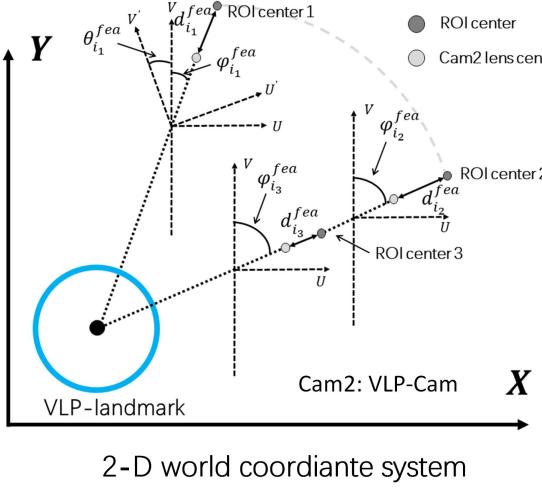


Fig. 11. Position of the VLP-Cam lens center is uniquely determined by VLP-feature.

Then the Levenberg–Marquardt method⁴ is applied to solving the following least-squares problem:

$$\hat{\mathbb{P}}_{i_k} \in \arg \min J_1(\mathbb{P}_{i_k}) \quad (18)$$

to acquire the optimized 2-D position of VLP-landmark $\hat{\mathbb{P}}_{i_k} = (\hat{x}_{i_k}, \hat{y}_{i_k})$ (corresponding 3-D position $(\hat{x}_{i_k}, \hat{y}_{i_k}, H_{i_k}^L)$). Subsequently, we can use $\hat{\mathbb{P}}_{i_k}$ in return to define another enhanced cost function

$$J_2(\mathbb{X}_{i_k}) = \sum_{l=1}^k \frac{1}{2} \|\text{error}_{i_k}(\mathbb{X}_{i_k}, \hat{\mathbb{P}}_{i_l})\|^2. \quad (19)$$

The Levenberg–Marquardt method, again, is leveraged to solve the following least-squares problem:

$$\hat{\mathbb{X}}_{i_k} \in \arg \min J_2(\mathbb{X}_{i_k}) \quad (20)$$

where $\hat{\mathbb{X}}_{i_k} = (\hat{x}_{i_k}^{\text{cam}}, \hat{y}_{i_k}^{\text{cam}})$ (correspondingly, $\hat{p}_{i_k} = (\hat{x}_{i_k}^{\text{cam}}, \hat{y}_{i_k}^{\text{cam}}, H^{L1})$) is the optimized VLP-Cam position. In our work, we borrowed Ceres library⁵ to solve the above least-squares problems. So far, we have acquired the 3-D VLP-landmark position $(\hat{x}_{i_k}, \hat{y}_{i_k}, H_{i_k}^L)$ and 2-D robot position $\hat{\mathbb{X}}_{i_k} = (\hat{x}_{i_k}^{\text{cam}}, \hat{y}_{i_k}^{\text{cam}})$.

E. Map Construction

1) *Loop Closing Based on VLP-Feature*: A loop-closing detection method based on the similarity of trajectories is proposed. We claim that the i_k th position of the robot is uniquely determined by $\text{fea}_{i_k}^{\text{VLP}}$. As shown in Fig. 11, although $d_{i_1}^{\text{fea}} = d_{i_2}^{\text{fea}}$, ROI center 1 and ROI center 3 still can be distinguished for $\varphi_{i_1}^{\text{fea}} \neq \varphi_{i_2}^{\text{fea}}$. Similarly, as $d_{i_2}^{\text{fea}} \neq d_{i_3}^{\text{fea}}$, ROI center 2 and ROI center 3 can be discriminated regardless of $\varphi_{i_2}^{\text{fea}} = \varphi_{i_3}^{\text{fea}}$. It should be pointed out that if VLP-Cam only does rotation, the feature angle will not change, e.g., the change in $\theta_{i_1}^{\text{fea}}$ does not affect $\varphi_{i_1}^{\text{fea}}$ in Fig. 11. Moreover,

as the robot is exposed to the same illumination range, we have $i d_{i_1}^{\text{fea}} = i d_{i_2}^{\text{fea}} = i d_{i_3}^{\text{fea}}$. Thus, the uniqueness of each position of the VLP-Cam lens center in i th trajectory can be guaranteed.

Based on the above properties of VLP-feature, for two different trajectories, e.g., $\text{Traj}_i = \{\text{key}_{i_k}^{\text{VLP}} | k = 1, 2, \dots\}$ and $\text{Traj}_j = \{\text{key}_{j_l}^{\text{VLP}} | l = 1, 2, \dots\}$ ($i < j$), we can evaluate their affinity with the aid of VLP-feature. For each newly updated $\text{fea}_{j_l}^{\text{VLP}}$, VWR-SLAM system loop over Traj_i and calculate the following two errors:

$$\text{reldist}_{j_l}^{i_k} \triangleq \frac{|d_{i_k}^{\text{fea}} - d_{j_l}^{\text{fea}}|}{d_{j_l}^{\text{fea}}} \quad (21)$$

and

$$\text{relang}_{j_l}^{i_k} \triangleq \frac{|\varphi_{i_k}^{\text{fea}} - \varphi_{j_l}^{\text{fea}}|}{\varphi_{j_l}^{\text{fea}}} \quad (22)$$

until a $\text{fea}_{j_l}^{\text{VLP}}$ satisfying the following conditions at once has been found.

- 1) $i d_{i_k}^{\text{fea}} = i d_{j_l}^{\text{fea}}$.
- 2) The relative error $\text{reldist}_{j_l}^{i_k}$ is upper bounded by a threshold $d_{\text{thres}}^{\text{rel}}$.
- 3) The relative error $\text{relang}_{j_l}^{i_k}$ is upper bounded by a threshold $\varphi_{\text{thres}}^{\text{rel}}$.

The $\text{fea}_{j_l}^{\text{VLP}}$ meeting the above conditions is reckoned to be a VLP-feature roughly similar to $\text{fea}_{i_k}^{\text{VLP}}$, i.e., $\text{key}_{j_l}^{\text{VLP}}$ is roughly similar to $\text{key}_{i_k}^{\text{VLP}}$. To analyze the precise similarity between Traj_i and Traj_j , the iterative closest point (ICP) [19] method is used. We assume that there are n VLP-keyframes of Traj_i achieving rough similarity with n VLP-keyframes of Traj_j one by one. By proper arrangement, we can realign the indexes and make $\text{key}_{j_s}^{\text{VLP}}$ roughly similar to $\text{key}_{i_s}^{\text{VLP}}$ for $s = 1, 2, \dots, n$. The translation vector t^* and the orientation matrix R^* computed from ICP are applied to the accumulative Euclidean distance d_{acc} defined as follows:

$$d_{\text{acc}} = \frac{1}{n} \sum_{s=1}^n \|\hat{p}_{j_s}^T - (R^* \hat{p}_{i_s}^T + t^*)\|. \quad (23)$$

If d_{acc} is smaller than an artificial threshold $d_{\text{thres}}^{\text{acc}}$, then $\{\hat{p}_{j_s} | s = 1, 2, \dots, n\}$ will be reasonably replaced by $\{\hat{p}_{i_s} | s = 1, 2, \dots, n\}$. Videlicet, the precise similarity between Traj_i and Traj_j is reckoned to be established and the loop closure is achieved.

2) *Full Optimization and Map Update*: The full optimization section uses the same scheme, as mentioned in Section III-D, to improve the accuracy of all the positions of VLP-Cam and VLP-landmark at once. Subsequently, the map will be built up using the VLP-landmark position, VLP-keyframe, and map point from ORB-SLAM2. The map can update and visualize the detected VLP-landmark, robot trajectory, and 3-D structure of the environment synchronously.

IV. EXPERIMENT AND ANALYSIS

In this section, we conduct three different experiments to verify the performance of our VWR-SLAM system. First,

⁴A celebrated nonlinear optimization method [19].

⁵A widely used library for solving least-squares problems [19].

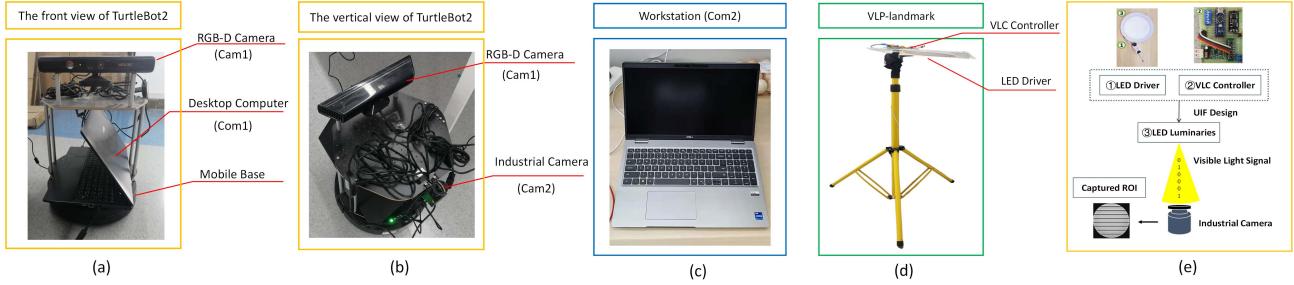


Fig. 12. Hardware setup. (a) and (b) Front view and vertical view of TurtleBot2 equipped with RGBD-Cam, Com1, and a mobile base. (c) Remote control workstation. (d) LED is used as VLP-landmark. (e) Communication process of the camera-based VLP.

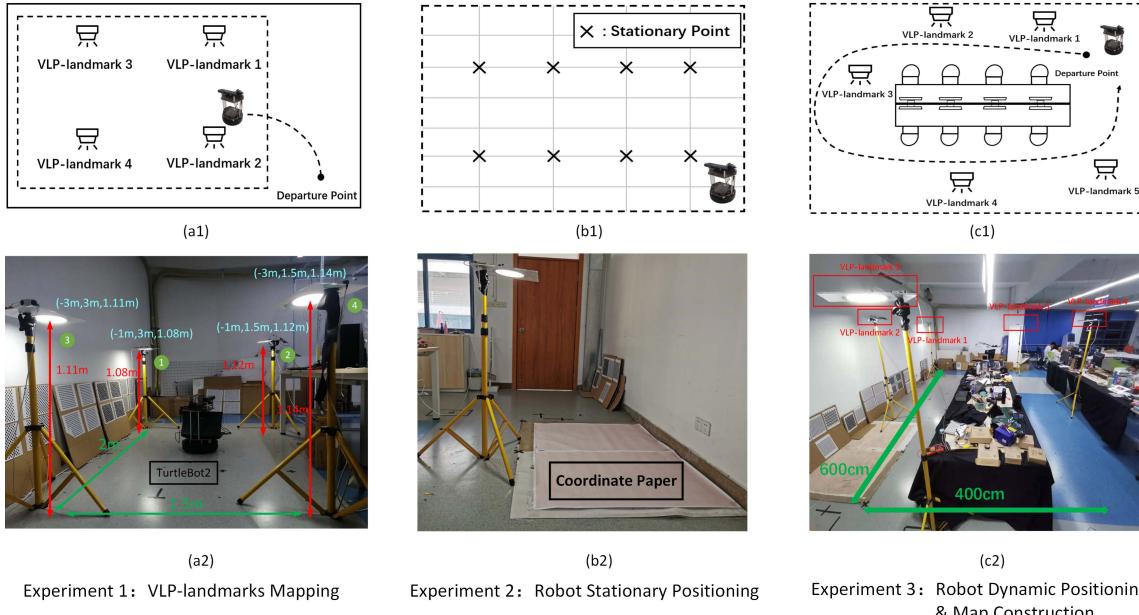


Fig. 13. Indoor environmental setup of three experiments. (a1) Four VLP-landmarks were settled at the corners of a 1.5×2 m rectangle area. (a2) Real scene of experiment 1. (b1) Eight selected points on coordinate paper for stationary positioning. (b2) Real scene of experiment 2. (c1) Five VLP-landmarks are distributed in a large-scale environment with 400 cm in width and 600 cm in length to form a sparse VLP-landmark environment. (c2) Real scene of experiment 3.

we evaluate the mapping accuracy of our VWR-SLAM by comparing the calculated 3-D localization result with the ground truth of the VLP-landmark. Then, we evaluate the robot positioning accuracy, and finally, we evaluate the real-time robot positioning and VLP-landmark map-building performance by manipulating the robot to wander around in a large-scale environment using our VWR-SLAM.

A. Hardware Setup

As shown in Fig. 12, the experiments are performed on a TurtleBot2, which is equipped with a Kinect V1 as RGBD-Cam, an industrial camera (MindVision UB-300) as VLP-Cam, a netbook (Module HASEE CW85S07 with Intel(R) Core (TM) i5-9400 and 8.00 GB RAM) as Com1, and a two-wheeled differential mobile base (iClebo Kobuki) as wheel odometer. Both RGBD-Cam and VLP-Cam are pre-calibrated with known camera intrinsic parameters. Notably, the VLP-Cam should be kept strictly vertical; otherwise, there might be irreparable estimation errors during VLP-landmark mapping. The VWR-SLAM system is implanted on Com1

which runs an Ubuntu Mate 16.04 OS equipped with a robot operating system (ROS). All the inputs from RGBD-Cam, VLP-Cam, and wheel odometer are sent to Com1 through data wire and collected by the VWR-SLAM system for further estimation. The robot is remotely controlled by a workstation laptop (Module DELL Precision 3561 with 11th Gen Intel(R) Core (TM) i7-11850H and 32 GB). The workstation sends its command to Com1 by Secure SHell (SSH) connection so that the surveyor can conduct work. Some LEDs of the same specifications (frequency 5 kHz, diameter 170 mm), whose radiation surface has a circular shape of size 18.0 cm in diameter, are mounted on yellow poles as the VLP-landmarks. The top of the pole comprises a VLC controller and an LED driver. Through the VLC controller, each LED has a unique identification frequency (UIF) correlating to the detailed position information and can transfer its information in the form of luminaries (visible light signal). Then, VLP-Cam can capture ROI and decode its ID. However, as we assume that the decoding result is inaccurate, the VWR-SLAM system cannot acquire LED position simply by LED-ID.

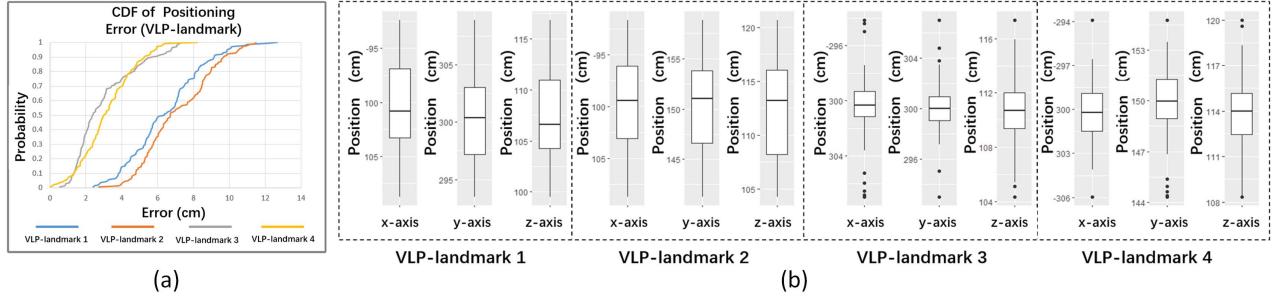


Fig. 14. Positioning results of the four VLP-landmarks in Experiment 1. (a) CDF diagram for the mapping result of the measured landmarks. (b) Boxplot of the x -axis, y -axis, and z -axis, respectively, from different VLP-Landmarks.

TABLE I
3-D MAPPING RESULT OF VLP-LANDMARK IN EXPERIMENT 1

#	VLP-landmark 1	VLP-landmark 2	VLP-landmark 3	VLP-landmark 4
Precise position (cm)	(−100, −300, 108)	(−200, 150, 112)	(−300, 300, 111)	(−300, 150, 114)
RMSE _x (cm)	3.76 (3.76 %)	4.17 (2.08 %)	2.04 (0.68 %)	2.36 (0.78 %)
RMSE _y (cm)	3.97 (1.32 %)	4.29 (2.86 %)	1.68 (0.56 %)	2.16 (1.44 %)
RMSE _z (cm)	4.01 (3.71 %)	4.37 (3.90 %)	2.41 (1.60 %)	2.22 (1.94 %)
Average Positioning Error (cm)	6.45	7.12	3.01	3.21

* Data in bold represents the best one in each row.

* RMSE_x, RMSE_y, and RMSE_z stand for the RMSE of the x -axis, y -axis, and z -axis respectively.

B. Experiment 1: VLP-Landmark Mapping

To evaluate the mapping accuracy of our VWR-SLAM system, we carried out experiment 1 to calculate the 3-D locations of 4 VLP-landmarks. As shown in Fig. 13(a1) and (a2), four LEDs were settled at the corners of a 1.5×2 m rectangle area, which were used as VLP-landmarks. During the experiment, the departure point of Turtlebot2 was considered to be the origin of the world coordinate system. The workstation Com2 sent the command and controlled the robot to go around the experimental field. To enable a stable acquisition of image frames of VLP-landmarks, the robot needs to stay under the VLP-landmark for a while. The VLP-landmark mapping result can be seen in Fig. 14 and Table I.

From the cumulative distribution function (CDF) diagram [Fig. 14(a)], we can see that the maximum mapping error of VLP-landmark localization is within 12.66 cm and the best average mapping error of one VLP-landmark is 3.01 cm. This result is acceptable and might be caused by the errors of bundle adjustment from ORB-SLAM2. The bundle adjustment can evenly allocate the accumulated error of the current observation to previous locations' errors when there are adequate RGB-D keyframes. Since VLP-landmark 1 and VLP-landmark 2 are close to the departure point of the robot, the number of the detected RGB-D keyframes is tiny, resulting in a huge accumulated error in these areas. In addition, the acquisition of RGB-D keyframes depends on the difference between two image frames from RGBD-Cam. If the movement of the robot is slight, the information for state estimation is little, which might also cause this trouble. It is noticeable that there are a few outliers in the boxplot due to the loss of information

[Fig. 14(b)]. Besides, due to the large root mean square error (RMSE) (Table I), the positioning data of VLP-landmark 1 and VLP-landmark 2 are relatively scattered, and therefore the outliers of VLP-landmark 1 and VLP-landmark 2 are covered. It is worth mentioning that the threshold of keyframe selection $d_{\text{thres}}^{\text{fea}}$ in Fig. 8 should be properly considered. If this value is too small, the calculated k -value defined in (8) will deviate from the true value, which will result in errors in VLP-landmarks' height estimation. Oppositely, if this value is too large, the 2-D positioning accuracy of VLP-landmarks will decrease. Compared with LedMapper [18], our VWR-SLAM system does not need any prior LED location. In addition, our system can work in a sparse luminaries environment, outperforming the performance of LedMapper which could be affected by the number of LEDs. To illustrate the superiority, we compare the average mapping accuracy of our method with the other state-of-the-art mapping works in Table II.

C. Experiment 2: Robot Stationary Positioning

In this section, we test the robot stationary positioning accuracy of our VWR-SLAM. We randomly select eight points on the coordinate paper and control the robot to approach each of them to measure the 2-D coordinate of the motionless points [Fig. 13(b1) and (b2)]. The ground-truth value of the robot is measured manually and compared with the localization output of our VWR-SLAM to acquire the positioning error. Similar to the experiment designed in Yan et al. [3], each static point on the coordinate was measured 30 times, and the average stationary positioning error is 1.81 cm. Fig. 15(b) and (c) illustrates that more than 90% of the positioning errors are less

TABLE II
PERFORMANCE COMPARISON WITH THE SOTA METHODS FOR ROBOT POSITIONING AND LED MAPPING

Method	Robot Average Positioning Error (cm)	LEDs Average Mapping Error (cm)	Hardware	Receiver Type	Known LEDs (at least)
R. [13]	-	10 (2D)	Raspberry+Computer	Camera+LiDAR	0
R. [18]	-	0.07 (3D)	Raspberry+Computer	Camera+IMU	2
R. [14]	-	4 (2D)	Raspberry+Computer	Camera+LiDAR	0
R. [17]	-	4.46 (3D)	Computer	AOA estimators	0
R. [3]	4.31 (2D)	-	Raspberry+computer	Camera	1
R. [4]	2.5 (3D)	-	Raspberry+computer	Camera+odometer+LiDAR	0
R. [22]	3.9 (3D)	-	Raspberry+computer	Camera	2
R. [7]	5.0 (3D)	-	Raspberry+computer	Camera+IMU	1
R. [23]	3.0 (2D)	-	Raspberry+computer	Camera+encoder+gyroscope	1
R. [24]	0.8 (3D)	-	Raspberry+computer	Camera	2
R. [25]	14.0 (2D)	-	STM32F103+computer	PD+IMU	3
R. [26]	17.5 (3D)	-	Smartphone	Camera	1
R. [27]	13.4 (3D)	-	Smartphone	Camera+PDR	1
R. [28]	2.3 (2D)	-	Computer	Camera	1
R. [29]	3.85 (3D)	-	Computer	Camera	2
Our VWR-SLAM	1.81 (2D)	3.01 (3D)	Computer	Camera+odometer	0

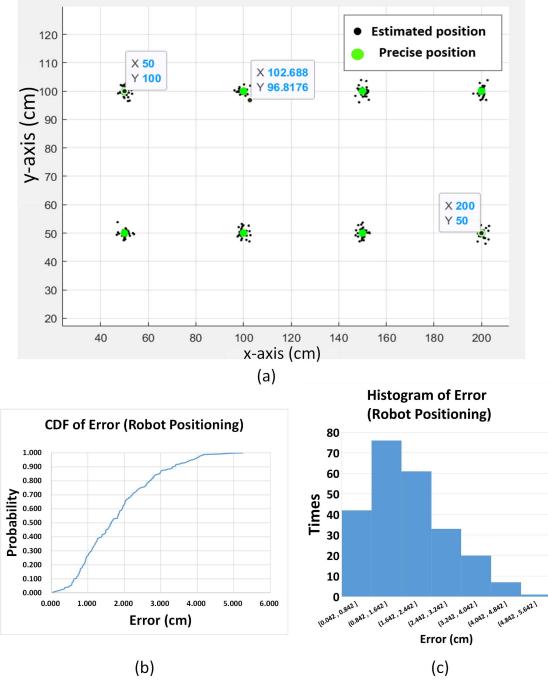


Fig. 15. Stationary positioning errors in Experiment 2. (a) Distribution of positioning results and the true value on indoor level ground. (b) CDF diagram for the robot positioning error. (c) Histogram of the distribution for the robot positioning error.

than 2.66 cm and the 95% confidence level interval of the positioning error is [1.68, 1.94] cm. Besides, in the aspect of time efficiency, the positioning calculation requires less time to get satisfactory results, different from the VLP-landmark mapping part which needs extra time for the robot to approach the LED center and collect enough VLP-keyframes. As can be seen in Table II, our VWR-SLAM system achieves satisfactory performance in robot positioning. It can also minimize the dependence on the number of known LEDs, and the average positioning accuracy is only 1.81 cm.

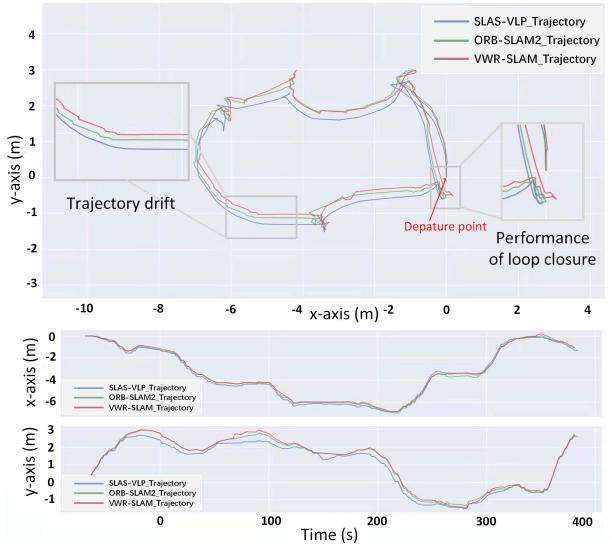


Fig. 16. Dynamic trajectory estimation of the VWR-SLAM system, ORB-SLAM2 system, and SLAS-VLP system on a 2-D world coordinate system.

D. Experiment 3: Robot Dynamic Positioning and LED Mapping in Large-Scale Environment

In this section, we test the performance of the proposed VWR-SLAM system in terms of cumulative robot positioning error and the effect of loop detection. Five LEDs are distributed as VLP-landmarks in the large-scale environment which is 400 cm in width and 600 cm in length to create a sparse LED environment [Fig. 13(c1) and (c2)]. The robot carrying two cameras for data acquisition goes around the experimenting area by remote control, completing the LED mapping and robot dynamic positioning work in real-time. To verify the dynamic robot localization performance of our tightly coupled algorithm, we compare the VWR-SLAM system with the ORB-SLAM2 and the SLAS-VLP algorithms [20]. Considering that the SLAS-VLP algorithm requires an LED with a known position, we place it artificially close to the departure

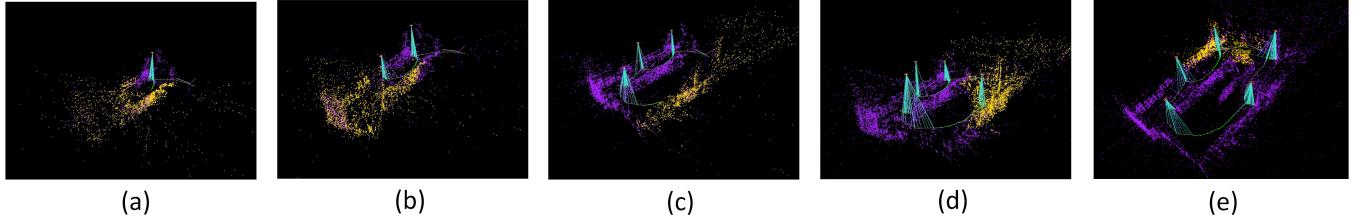


Fig. 17. Visualization of the 3-D reconstruction mapping process in experiment 3, including the 3-D sparse point clouds, the robot trajectory, and the distribution of VLP-landmarks. (a)–(e) Map construction process from one VLP-landmark to five VLP-landmarks.

point, i.e., VLP-landmark 1, in Fig. 13(c1). However, for our VWR-SLAM system, the LED-ID and location information of VLP-landmark 1 is still unknown. We record data packets through the ROS platform and run the above three projects separately on Com1. The robot's trajectory is recorded in a file in Technische Universität München (TUM) format and analyzed by evaluation of odometry (EVO) software [21].

As shown in Fig. 16, three trajectories generated from the above three systems almost coincide, but eminently the VWR-SLAM system can correct the drift caused by the wheel odometer in the SLAS-VLP algorithm. Moreover, the constraints from our proposed least-squares function [(20)] contribute to the reduction in drift. The closed-loop error of the VWR-SLAM system is the least among the three systems, which demonstrates that the loop closure module of the VWR-SLAM system works well. Fig. 17 illustrates the map construction performance. The map can update and visualize the environment's 3-D structure, the robot's trajectory, and the 3-D position of detected VLP-landmarks in real-time, which can be used as a prebuilt LED map for VLP systems. In Fig. 17, with the exploration of the robot, the effect of mapping is gradually improved, and the overall mapping is completed when the robot returns to the departure point and completes the loop closure. To reduce the memory required for program operation, dense map points are not used for mapping. Though some information in the indoor environment (such as RGB color and the object outline) cannot be displayed concretely, the distribution of the LED in the indoor environment remains clear. More details about the map construction process can be seen in our video record.

E. Discussion of Limitations

First, since we do not use IMU, the keyframes obtained by the RGB-D camera are discontinuous, which leads to a decrease in the number of industrial camera positions obtained, resulting in a decline in the accuracy of VLP-landmark location estimation. Theoretically, we can increase the number of RGB-D keyframes by elevating VLP-landmarks to improve the mapping accuracy. However, we found it inefficient by experimental verification. Second, to reduce the hardware cost, we directly use a simple angular transducer, a wheel odometer provided by the mobile base. But it yields an inferior performance of the least-squares estimation for optimizing the position to some extent. Third, the VWR-SLAM system is only suitable for a sparse interior lighting scene. That is, simultaneous positioning of more than one LED is impractical

for our system. Besides, to ensure the accuracy of LED mapping, many parameters need to be manually adjusted to be appropriate in advance. This is also a thorny problem in terms of efficiency. Furthermore, all the LED-IDs of VLP-landmarks should be different, which is a restriction to application in real-world scenarios. To overcome this problem, it is possible to use RGB-D images to correlate the uniqueness of the VLP-landmark with the environment where the VLP-landmark is located in future work.

V. CONCLUSION

This article introduces a novel system, termed VWR-SLAM, which can provide efficient and accurate mapping of modulated LEDs for VLP and achieve robust and accurate indoor localization. Compared with the traditional VLP method that needs manual surveys for the VLP-landmark map, our method requires much less human effort in building a usable LED map, thereby reducing the deployment costs of VLP systems in reality. Through tightly coupled multisensor fusion, our framework solves the full SLAM problems and offers good performance and high reliability in some extreme circumstances, such as sparse LED deployment. The experiments show that our framework can provide high precision and robust indoor localization without any prior LED position information. Meanwhile, the system can achieve a 3-D reconstruction of the VLP-landmark map with robot trajectory and indoor environment.

REFERENCES

- [1] J. Chen, D. Zeng, C. Yang, and W. Guan, "High accuracy, 6-DoF simultaneous localization and calibration using visible light positioning," *J. Lightw. Technol.*, vol. 40, no. 21, pp. 7039–7047, Nov. 1, 2022.
- [2] J. Fang et al., "High-speed indoor navigation system based on visible light and mobile phone," *IEEE Photon. J.*, vol. 9, no. 2, pp. 1–11, Apr. 2017.
- [3] Z. Yan, W. Guan, S. Wen, L. Huang, and H. Song, "Multirobot cooperative localization based on visible light positioning and odometer," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–8, 2021.
- [4] W. Guan et al., "Robot localization and navigation using visible light positioning and SLAM fusion," *J. Lightw. Technol.*, vol. 39, no. 22, pp. 7040–7051, Nov. 15, 2021.
- [5] Q. Liang and M. Liu, "A tightly coupled VLC-inertial localization system by EKF," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 3129–3136, Apr. 2020.
- [6] C. Yang, S. Wen, D. Yuan, J. Chen, J. Huang, and W. Guan, "CGA-VLP: High accuracy visible light positioning algorithm using single square LED with geomagnetic angle correction," *Photonics*, vol. 9, no. 9, p. 653, Sep. 2022.
- [7] Q. Liang, J. Lin, and M. Liu, "Towards robust visible light positioning under LED shortage by visual-inertial fusion," in *Proc. Int. Conf. Indoor Positioning Indoor Navigat. (IPIN)*, Sep. 2019, pp. 1–8.

- [8] W. Guan, L. Huang, B. Hussain, and C. P. Yue, "Robust robotic localization using visible light positioning and inertial fusion," *IEEE Sensors J.*, vol. 22, no. 6, pp. 4882–4892, Mar. 2022.
- [9] R. Mur-Artal and J. D. Tardós, "ORB-SLAM2: An open-source slam system for monocular, stereo, and RGB-D cameras," *IEEE Trans. Robot.*, vol. 33, no. 5, pp. 1255–1262, Oct. 2017.
- [10] Q. Liang, Y. Sun, L. Wang, and M. Liu, "A novel inertial-aided visible light positioning system using modulated LEDs and unmodulated lights as landmarks," *IEEE Trans. Autom. Sci. Eng.*, vol. 19, no. 4, pp. 3049–3067, Oct. 2022.
- [11] H. Song, S. Wen, C. Yang, D. Yuan, and W. Guan, "Universal and effective decoding scheme for visible light positioning based on optical camera communication," *Electronics*, vol. 10, no. 16, p. 1925, Aug. 2021.
- [12] Y. Liu et al., "Comparison of thresholding schemes for visible light communication using mobile-phone image sensor," *Opt. Exp.*, vol. 24, 3, pp. 8–1973, 2016.
- [13] R. Amsters, E. Demeester, N. Stevens, and P. Slaets, "Calibration of visible light positioning systems with a mobile robot," *Sensors*, vol. 21, no. 7, p. 2394, Mar. 2021.
- [14] R. Amsters, E. Demeester, P. Slaets, D. Holm, J. Joly, and N. Stevens, "Towards automated calibration of visible light positioning systems," in *Proc. Int. Conf. Indoor Positioning Indoor Navigat. (IPIN)*, Sep. 2019, pp. 1–8.
- [15] W. Hess, D. Kohler, H. Rapp, and D. Andor, "Real-time loop closure in 2D LIDAR SLAM," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2016, pp. 1271–1278.
- [16] C. Zhang and X. Zhang, "Pulsar: Towards ubiquitous visible light localization," in *Proc. 23rd Annu. Int. Conf. Mobile Comput. Netw.*, Oct. 2017, pp. 33–35.
- [17] K. Zhang, Z. Zhang, and B. Zhu, "Beacon LED coordinates estimator for easy deployment of visible light positioning systems," *IEEE Trans. Wireless Commun.*, vol. 21, no. 12, pp. 10208–10223, Dec. 2022.
- [18] Q. Liang, Y. Sun, C. Liu, M. Liu, and L. Wang, "LedMapper: Toward efficient and accurate LED mapping for visible light positioning at scale," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–12, 2022.
- [19] G. Xiang, Z. Tao, L. Yi, and Y. Qinrui, *14 Lectures on Visual Slam: From Theory to Practice*. Beijing, China: Publishing House Electronics Industry, 2019, pp. 117–119.
- [20] L. Huang, S. Wen, Z. Yan, H. Song, S. Su, and W. Guan, "Single LED positioning scheme based on angle sensors in robotics," *Appl. Opt.*, vol. 60, no. 21, p. 6275, 2021.
- [21] M. Grupp. (2017). *EVO: Python Package for the Evaluation of Odometry and Slam*. [Online]. Available: <https://github.com/MichaelGrupp/evo>
- [22] P. Lin et al., "Real-time visible light positioning supporting fast moving speed," *Opt. Exp.*, vol. 28, no. 10, p. 14503, 2020.
- [23] R. Amsters, D. Holm, J. Joly, E. Demeester, N. Stevens, and P. Slaets, "Visible light positioning using Bayesian filters," *J. Lightw. Technol.*, vol. 38, no. 21, pp. 5925–5936, Nov. 1, 2020.
- [24] W. Guan, S. Chen, S. Wen, Z. Tan, H. Song, and W. Hou, "High-accuracy robot indoor localization scheme based on robot operating system using visible light positioning," *IEEE Photon. J.*, vol. 12, no. 2, pp. 1–16, Apr. 2020.
- [25] Z. Li, A. Yang, H. Lv, L. Feng, and W. Song, "Fusion of visible light indoor positioning and inertial navigation based on particle filter," *IEEE Photon. J.*, vol. 9, no. 5, pp. 1–13, Oct. 2017.
- [26] R. Zhang, W.-D. Zhong, Q. Kemao, and S. Zhang, "A single LED positioning system based on circle projection," *IEEE Photon. J.*, vol. 9, no. 4, pp. 1–9, Aug. 2017.
- [27] H. Huang et al., "Hybrid indoor localization scheme with image sensor-based visible light positioning and pedestrian dead reckoning," *Appl. Opt.*, vol. 58, no. 12, pp. 3214–3221, 2019.
- [28] H. Li et al., "A fast and high-accuracy real-time visible light positioning system based on single LED lamp with a beacon," *IEEE Photon. J.*, vol. 12, no. 6, pp. 1–12, Dec. 2020.
- [29] W. Guan, X. Zhang, Y. Wu, Z. Xie, J. Li, and J. Zheng, "High precision indoor visible light positioning algorithm based on double LEDs using CMOS image sensor," *Appl. Sci.*, vol. 9, no. 6, p. 1238, Mar. 2019.



Junlin Huang is currently pursuing the bachelor's degree in mathematics and applied mathematics with the South China University of Technology, Guangzhou, China.

His current research interests include visible light positioning, visual SLAM, and distributed optimization.



Shangsheng Wen received the B.E. degree from the Wuhan University of Science and Technology, Wuhan, China, in 1984, the M.E. degree from the Anhui Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Beijing, China, in 1993, and the Ph.D. degree from South China Normal University, Guangzhou, China, in 2001.

He is a Professor with the State Key Laboratory of Luminescent Materials and Devices, South China University of Technology, Guangzhou. He has undertaken more than 20 scientific research projects and published 100 articles. His research focuses on visible light communication.



Wanlin Liang is currently pursuing the bachelor's degree in automation science and engineering with the South China University of Technology, Guangzhou, China.

Her research interests include VLC-based image communication and ROS-based robot mapping and localization.



Weipeng Guan received the B.E. and M.E. degrees from the South China University of Technology, Guangzhou, China, in 2016 and 2019, respectively. He is currently pursuing the Ph.D. degree with the University of Hong Kong, Hong Kong.

He was working with Samsung Electronics, Shenzhen, China, Huawei Technologies, Shenzhen, TP-LINK, Shenzhen, The Chinese Academy of Sciences, Beijing, China, and The Chinese University of Hong Kong, Hong Kong. He was a Technical Consultant for several companies, such as TCL, Shenzhen. He has authored or coauthored more than 60 research articles in prestigious international journals and conferences. He is also the inventor of more than 20 granted patents. His research interests include robotics, event-based VO/VIO/SLAM, and visible light positioning.