

Robot Localization and Navigation Using Visible Light Positioning and SLAM Fusion

Weipeng Guan , Linyi Huang , Shangsheng Wen , Zihong Yan , Wanlin Liang, Chen Yang, and Ziyu Liu

Abstract—Visible light positioning (VLP) is a promising technology since it can provide high accuracy indoor localization based on the existing lighting infrastructure. However, existing approaches often require dense LED distributions and persistent line-of-sight (LOS) between transmitter and receiver. What's more, sensors are imperfect, and their measurements are prone to errors. Through multi sensors fusion, we can compensate the deficiencies of stand-alone sensors and provide more reliable pose estimations. In this work, we propose a loosely-coupled multi-sensor fusion method based on VLP and Simultaneous Localization and Mapping (SLAM), using light detection and ranging (LiDAR), odometry, and rolling shutter camera. Our multi-sensor localizer can provide accurate and robust robot localization and navigation in LED shortage/outage situations. The experimental results show that our proposed scheme can provide an average accuracy of 2.5 cm with around 42 ms average positioning latency.

Index Terms—High accuracy, indoor positioning, LiDAR-SLAM, multi sensors fusion, robotics, visible light positioning (VLP).

I. INTRODUCTION

PRECISE localization is a prerequisite for many autonomous systems, such as robotics, unmanned aerial vehicles, etc. Also, indoor positioning is an especially challenging problem, where localization cannot be achieved by GPS due to the satellite signal being greatly attenuated, while the traditional radio-based indoor positioning technologies, such as Bluetooth, Wi-Fi, Radio-frequency Identification (RFID) and ultra-wideband (UWB), still have some disadvantages in terms of low accuracy, high latency, electromagnetic interference or high hardware cost [1].

Manuscript received June 23, 2021; revised August 7, 2021; accepted September 14, 2021. Date of publication September 20, 2021; date of current version November 16, 2021. 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 Grants 202010561158, 202010561155, 202110561162, 202110561165, and 202110561163, and in part by the Guangdong Provincial Training Program of Innovation and Entrepreneurship for Undergraduates under Grant S202010561272 (Corresponding author: Shangsheng Wen.)

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

Linyi Huang, Zihong Yan, Wanlin Liang, Chen Yang, and Ziyu Liu are with the School of Automation Science and Engineering, South China University of Technology, Guangzhou 510006, China (e-mail: auhly@mail.scut.edu.cn; 201830630416@mail.scut.edu.cn; aulwl@mail.scut.edu.cn; 201930364105@mail.scut.edu.cn; 201930362200@mail.scut.edu.cn).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/JLT.2021.3113358>.

Digital Object Identifier 10.1109/JLT.2021.3113358

In recent years, the localization technology based on visible light communication (VLC), which is referred as visible light positioning (VLP), has attracted intensive attention as it can provide high accuracy positioning based on the existing lighting infrastructure. Look up in almost any building and you will be able to see multiple light fittings, demonstrating that at most indoor locations [2]. The modulated LED broadcasts its unique identity (ID) via switching at a high frequency imperceptible to the human eye, but which can be recognized by photodiodes (PD) [3]–[5] and rolling shutter effect (RSE) cameras [6]–[8]. The LED-ID can be mapped/modulated once for all since they are normally fixed and not easily vulnerable to environmental changes. Hence, the “Last Mile Problem” of localization is solved via VLP and the pre-built LED landmark map. Ref. [9]–[11] design the PD-based indoor positioning system using machine learning to provide around 10 cm positioning accuracy. Besides, some state-of-the-art (STOA) camera-based VLP systems can offer centimeter-level accuracy on commodity smartphones [12], [13] or mobile robots [6], [14]. Despite the promising performance of existing VLP systems, there remain some practical challenges.

A. Motivation

One of the most urgent issues arises from the fact that VLP normally requires multiple LED observations at a time for successful positioning through trilateration or triangulation. This is because the normal LED lights offer less usable point features due to the lack of distinguishable appearance, e.g., one feature for a circular LED [15]. However, the number of decodable LEDs in a camera view is limited by a couple of practical factors, such as deployment density and geometry layout of LEDs, obstruction of the LOS views, limited field-of-view (FOV) of the camera, etc. As such, the shortage or outage of LEDs can severely deteriorate the performance of the camera-only method in reality. To address this problem, different VLP-aided inertial localization methods have been proposed [6], [15]–[17] to provide pose estimation by combining VLP and the inertial measurement unit (IMU). In [17], An Extended Kalman Filter (EKF)-based loosely-coupled VLP-inertial fusion method is proposed. The presented experiments demonstrate the functionality that the IMU helps to overcome the moments with lack of VLP coverage. However, IMU suffers drift over time (cumulative error) or measurement noises (biases). Therefore, these VLP-aided inertial localization methods suffer from errors after long-term running. To address this problem, we are motivated to adopt LiDAR sensor to compensate the accumulated error of the inertial sensors, so that the VLP-aided inertial

& LiDAR method can achieve long-term accurate and robust localization.

In addition, sensors are imperfect, and their measurements are prone to errors. All of VLP, odometry, and LiDAR also have outliers. Through multi sensors fusion, we can compensate the deficiencies of stand-alone sensors and provide more reliable and robust estimations. Furthermore, one of the most basic yet important functions of intelligent mobile robots are the ability to navigate autonomously [18]. Navigation¹ denotes the robot's ability to determine its own position in its frame of reference and then to plan a path towards some goal location. However, to the best of our knowledge, most of the works in the field of VLP only can provide discrete position (localization) and lack the ability to navigate and tracking for moving targets, which is caused by the limited VLP coverage range, positioning latency, and the lack of the obstacle detection/avoidance function. The priori pre-built LED landmark map doesn't include any obstacle information around the environment. Based on this point, we are motivated to adopt LiDAR-SLAM for occupancy grid map building and obstacle detection, adding the tracking and navigation capability into the VLP system. Through multi-sensor fusion, we can estimate quantities that are unobservable using single sensors (e.g., VLP cannot detect the obstacle information; LiDAR cannot observe the global position; odometry has cumulative error) so that we can improve robustness, and handle the sensor failures.

B. Contribution

In this work, we propose a loosely-coupled multi-sensor fusion method for VLP and LiDAR-SLAM. Integrating multi-sensor, including RSE-camera, LiDAR, and odometry, enables a robot to maximize its perceptual awareness of the environment and obtain sufficient measurements to make it promising for robust indoor localization and navigation. More specifically, the LiDAR is employed to detect the surroundings for obstacle avoidance during the navigation, and compensate the cumulative error from the odometry. While the VLP is adopted to provide high accurate pose initialization and correction for LiDAR-SLAM and odometry. Our multi-sensor localizer includes map-fusion and localization-fusion, which firstly builds a map of the previously unknown environment, and then, provide accurate and reliable pose estimation through multi-sensor fusion. In order to realize map-fusion, we develop a VLP-constrained mapping scheme based on LiDAR-SLAM (Gmapping²) to construct the occupancy grid map based on the pre-built LED landmark map. While for the localization-fusion, the loosely-coupled EKF using VLP, odometry and LiDAR can provide high accuracy indoor localization based on the previously built VLP-constrained occupancy grid map, which is the premise of reliable navigation.

We highlight the contributions as follows:

- **Map-fusion:** VLP-constrained Gmapping algorithm is proposed to align the occupancy grid map (SLAM-map) with the pre-built LED landmark map, which is the premise of the fusion between VLP and LiDAR-SLAM can be integrated. Meanwhile, the localization result can be simultaneously visualized and aligned in both SLAM-map and floorplan.

- **Localization-fusion:** Loosely-coupled multi-sensor fusion based on the EKF is proposed to achieve high accuracy localization using RSE-camera, LiDAR and odometry. Through the state prediction from odometry motion and the measurement update from LiDAR scanner and VLP, we can relax the requirement on the minimum number of concurrently observable LEDs for positioning to zero. The LiDAR can compensate the cumulative error from the odometry, the odometry contributes to smooth pose estimation, while the VLP measurement can provide high accuracy pose initialization or pose calibration when it is available.
- We embedded the navigation function into our multi-sensor fusion for VLP-SLAM. Owing to the high accuracy performance of our multi-sensor fusion scheme, the robot can provide autonomous navigation. The efficacy of the scheme is verified in real-world experiments.

C. Organization

The remainder of this paper is organized as follows. Section II introduces the related works. Section III explains the methodology of our multi-sensor localizer based on VLP-SLAM, including the VLP calculation based on single LED and odometry, VLP-constrained Gmapping, and the loosely-coupled multi-sensor fusion based on EKF. Section IV presents the experimental evaluation and Section V is the conclusion.

II. RELATED WORK

A. RSE Camera-Based VLP

LEDs can transmit data over the air by directly modulating the light intensity at a high frequency that is invisible to human eyes but perceivable by camera or PD. Generally, PD is used as the receiver to capture the signal for the VLC system. There are many works use the PD and solar cells [19]–[21] as the receiver of VLP. PD is not an ideal VLP device, since it is sensitive to the light intensity variation and the diffuse reflection of the light signal, which degrades the positioning accuracy [22]. In contrast, camera-based VLP is favored in both commerce and industry due to the high positioning accuracy achievable by imaging geometry, and the good compatibility with user devices, such as mobile robots and smartphones. The pixels on RSE-camera are exposed and read out line by line instead of perceiving light for the whole image at a single moment. Then, the temporally-varying intensity signals from the LED transmitter are mapped to spatially-varying bright or dark strip patterns on the captured image. For comprehensive understanding of optical camera communication (OCC) and LED-ID decoding for RSE-camera-based VLP, we refer readers to [23]–[28].

Camera-based VLP systems employ modulated LED luminaries mounted at known locations (e.g., on the ceiling) as artificial landmarks, associate each LED measurement with the ID coordinate through RSE-based OCC, and provide information about the camera's relative position and rotation with respect to each detected LED by the LOS and weak perspective projection property. Decimeter/centimeter-level accuracy has been reported in some RSE-camera based VLP systems [7], [8], [13], [29]–[32]. Yet, all these methods treat LED lamps as point sources without

¹[Online]. Available: https://en.wikipedia.org/wiki/Robot_navigation

²[Online]. Available: <http://wiki.ros.org/gmapping>

geometric information. Thus, multiple known LED anchors are required at a time for trilateration or triangulation calculation. Such approaches rely on dense deployment of LED beacons. To reduce the number of required LEDs in VLP, Ref. [33] no longer treats the captured LED images as a point, but as an image whose geometric features are exploited to determine the receiver's orientation and location. However, this method requires an additional marker to be placed on the LED. Another straightforward method is to employ angular sensors to measure the receiver's orientation information, thus "compensating" the missing information due to the reduced number of LED lamps. Ref. [34] proposed single LED VLP using image and attitude information from mobile phone to achieve an average accuracy of 5.5 cm. Ref. [22] proposed a hybrid positioning algorithm consisting of VLP and pedestrian dead reckoning (PDR) to achieve decimeter-level accuracy. However, the IMU is a local frame with arbitrary yaw at the start of each run, it is not identical to global information (absolute measurements) which can be directly processed with the VLP, it needs pose initialization. Ref. [15], [16] efficiently relax the assumption on the minimum number of simultaneously observable LEDs from three to one through a tightly-coupled VLP and IMU localization system. However, they still built up the experimental platform in the area of $5\text{ m} \times 4\text{ m} \times 2.3\text{ m}$ with 23 LEDs, since their system required more than 4 LEDs for initialization. In our previous work [17], we proposed a loosely-coupled VLP-inertial fusion scheme, which can relax the requirement of the observable LEDs to zero. However, the robot has to head along the x-axis of the global frame before running so that it can obtain the initial orientation. Additionally, all of the proprioceptive sensors (IMU, odometry) are subject to cumulative error. Therefore, another exteroceptive sensors acquiring information from the robot's environment, are needed to handle the LED shortage/outage problem (when VLP is unavailable) and deal with the cumulative errors from angle sensors.

B. Simultaneous Localization and Mapping

SLAM is useful for building and updating maps within unknown environments, while the robot maintains the information about its location. It consists of the concurrent construction of a model of the environment (the map) and the estimation of the pose of the robot moving within it [35]. SLAM can be divided into two categories according to the sensors: vision-based and LiDAR-based. Compared to other distance measurement devices, the LiDAR sensor makes measurements around itself with a wide scanning angle and high angular resolution. Besides, it is invariant to illumination. The high reliability and precision make LiDAR sensor a popular option for pose estimation [36]. Cartographer SLAM [37], tinySLAM [38], FastSLAM [39], and Hector SLAM [40] have been widely applied for LiDAR range-finder-based mapping and localization. Gmapping SLAM [41] is the most widely used SLAM method in robotics, it uses the Rao-Blackwellized (RB) particle filter to process the environment perceived by the robot, and works with the wheel odometry to calculate the robot's position and construct the grid map. It uses a relatively small number of particles to represent the SLAM posterior, and reduces the computational effort required to perform resampling to successfully build very accurate maps. This approach has achieved remarkable success and indeed is

able to reduce the memory consumption while producing highly accurate grid maps. Therefore, Gmapping is more suitable for low-cost mobile platforms where computational resource is limited. For comprehensive understanding of Gmapping and the RB particle filter, readers are strongly advised to the works [41], [42].

Despite its advantage, the LiDAR sensor is not perfect and has several shortcomings. First of all, the initial pose for LiDAR-SLAM is an unsolved problem, since SLAM commonly suffers from data sparsity or limited vertical FOV in real-world applications. How to bootstrap the localization without a relatively accurate manually initial guess has not been well considered [43]. Secondly, the features obtained from the LiDAR are limited, which is caused by the low vertical resolution and the sparse point cloud of it. In addition, although most LiDAR offer a horizontal FOV of 360 degrees, it receives only few usable points from the side walls in a narrow corridor environment. In other words, it is difficult to estimate the pose in environments such as large spaces or long corridors without too much variety of observations, since the depth information obtained from the LiDAR rangefinder does not change with time and is considered featureless. Therefore, the robot will easily get lost when the geometry of the environment is quite simple, such as a public space surrounded by a few long walls or circular walls [44]. Finally, the rangefinder of the LiDAR is limited, it only receives the local information around the robot, and still cannot perceive the global information of the whole indoor or map environment.

To tackle these problems mentioned above, we are motivated to adopt VLP to provide the pose initialization or calibration for LiDAR-SLAM. The VLP can be used to make the robot capable of sensing a wide range of environments through the global information provided by LED beacons. While the observation of external landmarks (LED) is also useful to provide high accuracy pose initialization for LiDAR-SLAM. We regard the localization problem as a registration problem between local perceptual information from the LiDAR and global coordinate information from the VLP. It can handle the kidnapped problems³ via the VLP observation to successfully deal with the uncertainty of the initial guess for LiDAR-SLAM. Thus, combining with VLP and SLAM, we can achieve more refined, accurate, robust, and global pose estimation, which meets the intelligent robot application demand.

III. METHODOLOGY

The proposed multi-sensor fusion for VLP-SLAM aims at enabling sufficient environment perception (mapping), robust and accurate localization, and autonomous navigation. In this section, we firstly introduce the single LED VLP with odometry (SLO-VLP). Then, the loosely-coupled multi-sensor fusion for VLP-SLAM is introduced to show how to integrate two different exteroceptive sensors (camera and LiDAR) information.

A. Visible Light Positioning Calculation

The time-varying VLP signals from LEDs are perceived by the RSE-camera as spatially-varying strip patterns. Through the VLP observation, the pose of the robot can be estimated in

³[Online]. Available: https://en.wikipedia.org/wiki/Kidnapped_robot_problem

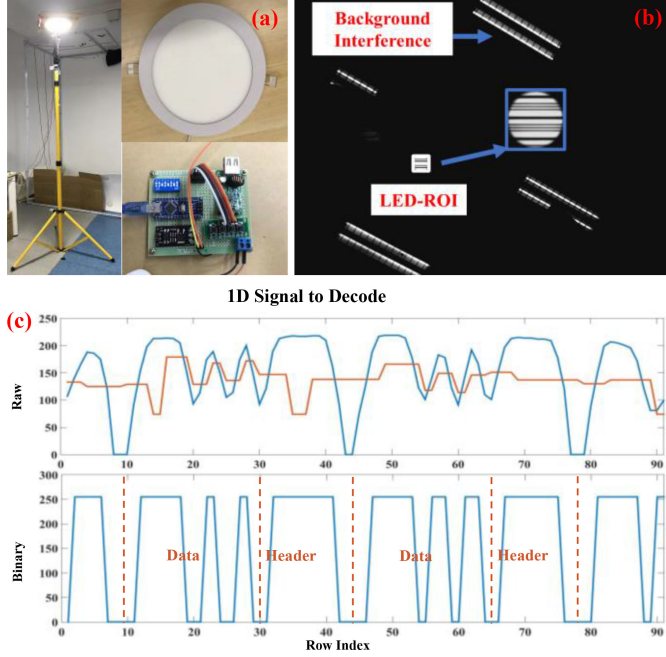


Fig. 1. (a) Self-developed VLC modulator for LED with VLP-ID broadcasting function; (b) Example results for LED-ROI extraction under background light interference; (c) The LED-ID decoding.

time. To do so, we need to find its centroid imaging location (also called as region-of-interest, ROI), recognize its identity (LED-ID), and retrieve LED's 3D position from the registered lights databases (pre-built LED landmark map). After that we can obtain the global 3D position and azimuth of the robot through SLO-VLP calculation.

1) *LED-ROI Detection*: The RSE-camera can capture barcode-like strip patterns from the high-frequency flashing LED during underexposure. Natural features are not observable, while bright objects (e.g., LED beacons, background light) can be easily distinguished, as shown in Fig. 1(b). We are interested in the image regions that carrying VLC information, termed as LED-ROI [45]. We firstly binarize the grayscale image by thresholding, and then dilate the binary image to fill the strip gaps. After that, the match template method is utilized to locate the LED-ROI in the captured image for subsequent LED-ID decoding. Since only the contour feature information is needed for ROI detection, we do not take every pixel into the computation but downsize the captured images with a constant scale, e.g., 6. To avoid the errors caused by down sampling, after the boundary of the LED-ROI is found, all the pixels nearby the obtained LED-ROI are checked again to detect the precise boundary. Since the number of pixels in the down-scaled image and the double-check pixels are much less than that in the original image, the proposed way to obtain the LED-ROI is efficient. The effectiveness of the LED-ROI detection scheme has been well evaluated in our previous work [46].

2) *VLC-ID Coding and Decoding*: For LED-ID coding, we develop a VLC modulator for retrofitting LED lamp to transmit

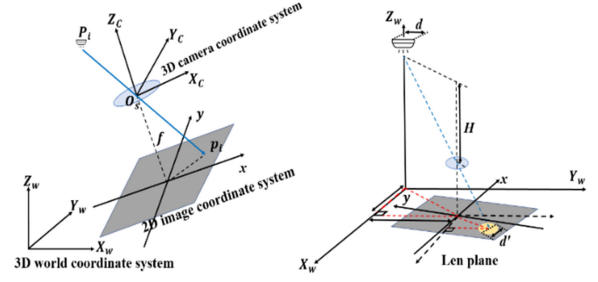


Fig. 2. The transforms among the world, camera, and image coordinate systems.

the ID information to the RSE-based camera receiver, as can be seen in Fig. 1(a). The VLC modulator can be embedded into commercial LEDs. Each LED lamp is assigned with a unique ID, which is associated with its location and stored in an ID-location database (pre-built LED landmark map). The designed ID data packet begins with a 6-bit preamble (header, 011110), preceded by 8-bit and 8 kHz-frequency OOK-coded data payload data. The payload carries one byte of ID, labeling up to 256 LEDs. The channel capacity is adequate for our experiment implementation and can be extended by a larger packet length. The VLC information is encoded by strips of varying widths. As can be seen in Fig. 1(c), the grayscale pixels in the centering column of the LED-ROI are selected for OOK demodulation. The thresholding method based on local extremum and sliding window [14] is adopted to counter the nonuniform illumination of the LEDs. The decoding rate of our LED-ID decoding method can maintain more than 95% at a height of 2.7 m, which had been evaluated in our previous work [17].

3) *Single LED VLP With Odometry (SLO-VLP)*: The pin-hole camera model is shown in Fig. 2. For VLP calculation, the position of the LED in the 3-D world coordinate $P_i = [X_i, Y_i, Z_i]^T$ ($i = 1^{st}, 2^{nd}, \dots, N$; i is associated with the VLP-ID) can be obtained through LED-ID decoding. While the LEDs mapped onto an image point $p_i = [x_i, y_i]^T$ can be measured through the LED-ROI detection. For any LEDs P_i in the world coordinate, we can model the perspective projection model between the world coordinate and the image plane coordinate p_i as:

$$\lambda \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = \begin{bmatrix} f/dx & 0 & u_0 & 0 \\ 0 & f/dy & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{R} & \mathbf{T} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix} \quad (1)$$

where λ can be represented as the depth between the LED and the camera. f (focal length), dx/dy (physical size of the pixel), and u_0/v_0 (the center pixel of the image) are formed together as Intrinsic Matrix (denoted as \mathbf{K}). \mathbf{T} represents the translation vector, which is equivalent to the receiver's position in the 3-D world coordinate system. \mathbf{R} is the rotation matrix from the world coordinate system to the camera coordinate system (2) shown at the bottom of this page, where α , β and γ is the angle along the X, Y, and Z axis, and s refer to $\sin(\cdot)$, c refer to $\cos(\cdot)$. The

$$\mathbf{R} = R_x(\alpha) \cdot R_y(\beta) \cdot R_z(\gamma) = \begin{bmatrix} c\beta * c\gamma & -c\beta * s\gamma & s\beta \\ c\alpha * s\gamma + c\gamma * s\alpha * s\beta & c\alpha * c\gamma - s\alpha * s\beta * s\gamma & -c\beta * s\alpha \\ s\alpha * s\gamma - c\alpha * c\beta * c\gamma & c\gamma * s\alpha + c\alpha * s\beta * s\gamma & c\alpha * c\beta \end{bmatrix} \quad (2)$$

three rotation angles can be estimated from the proprioceptive sensor (such as IMU or odometry) attached to the robot. Since the robot is moving horizontally on the floor, only the azimuth angle γ needs to be considered, while the other two rotation angles can be ignored as constant matrixes.

The pose of the robot moving in the 2D plane consists of coordinates and heading orientation, therefore we work with a four-dimensional state vector $\mathbf{s} = [X_s \ Y_s \ Z_s \ \gamma]^T$. In this paper, we discuss the mobile robot which moves in 2D environment, it is horizontal, and the rotation angles along the X and Y axis are taken to 0° (C_α and C_β are degenerate into identity matrix). Then the Z-coordinate Z_s of the positioning terminal can be calculated by:

$$Z_s = Z_i - \lambda \quad (3)$$

where λ is equal as the vertical distance from center of the camera O_s to the ceiling. The LED image is no longer treated as a point, but as an image whose geometric features are exploited to determine Z_s . Then vertical distance λ between the LED and the lens plane can be expressed as:

$$\frac{\lambda}{f} = \frac{D}{P_d d_{pixel}} \quad (4)$$

where D is the physical diameter of the LED, d_{pixel} is the pixel distance of the LED-ROI, and P_d is the conversion of the pixel distance and physical distance. Substituting (4) and (5) into (1), we derive the mathematical relation between the LEDs, $[X_i, Y_i, Z_i]^T$, and the observation values of their corresponding imag points $[x_i, y_i]^T$, then we can obtain the position of the camera's center $[X_s \ Y_s \ Z_s]$:

$$\frac{fD}{P_d d_{pixel}} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = \mathbf{K} \cdot \begin{bmatrix} \cos\gamma & -\sin\gamma & 0 & X_s \\ \sin\gamma & \cos\gamma & 0 & Y_s \\ 0 & 0 & 1 & Z_s \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix} \quad (5)$$

After estimating the coordinate of the camera's center, the 3D position and azimuth of the robot $\mathbf{s}_t = [x_t \ y_t \ z_t \ \theta_t]^T$ at time t can be estimated through the TF transformation⁴ as follows:

$$\begin{aligned} \begin{bmatrix} x_t \\ y_t \\ z_t \\ \theta_t \end{bmatrix} &= \mathbf{F}_{tf} \left(\begin{bmatrix} X_s \\ Y_s \\ Z_s \\ \gamma \end{bmatrix} \right) \\ &= \begin{bmatrix} r_x & 0 & 0 & 0 \\ 0 & r_y & 0 & 0 \\ 0 & 0 & r_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_s \\ Y_s \\ Z_s \\ \gamma \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \\ 0 \end{bmatrix} \end{aligned} \quad (6)$$

where, r_x, r_y, r_z and t_x, t_y, t_z are the rotation and translation coefficient from the center of the camera to the base_link⁵ of the robot. However, the odometry suffers drift over time, in which case the assumption of perfectly-known yaw information will gradually be problematic. While observing a circular LED without any asymmetry appearance cannot give orientation information. Therefore, this may have problems when long-term

drifts occur. In this paper, we utilize a LiDAR scan matcher component that corrects the robot's orientation estimated with a raw odometry by matching LiDAR scan data with previous scans, then the cumulative error of the odometry can be compensated:

$$\theta_t = \gamma = yaw_{odom} + yaw_{LiDAR \rightarrow map} \quad (7)$$

where yaw_{odom} is the raw orientation angle from the odometry, while $yaw_{LiDAR \rightarrow map}$ is the included angle between the output TF of the LiDAR-SLAM and the map_frame⁶.

B. Multi-Sensor Fusion for VLP and SLAM

The proposed multi-sensor fusion for VLP-SLAM includes map-fusion and localization-fusion. For the map-fusion, the VLP-constrained Gmapping is proposed to construct the occupancy grid map⁷ based on the pre-built LED landmark map. While for the localization-fusion, an EKF-based pose integration scheme is designed to stably obtain the global poses of the robot through odometry, LiDAR-SLAM and SLO-VLP localization.

1) *VLP-Constrained Gmapping for Map-Fusion*: Mapping is the fundamental prerequisites for localization and navigation. Both LiDAR-SLAM and VLP are map-based localization method, the localization outputs of them are corresponding to the origin of their respective map_frame (VLP-map and SLAM-map). As for the VLP-map, it is a pre-build LED landmark feature map, which manually indicate the localization of the LED and store in the registered lights databases through LED-ID coding. While the SLAM-map is an occupancy grid map (binary map) of navigable regions and boundary walls, which are represented through white and black (boundaries and obstacles), respectively, for visualization purposes. The occupancy grid map is created by obtaining the information from odometry and LiDAR scanning.

To some extended, the SLAM-map belongs to the local map, which is based on the starting point of the SLAM Gmapping process (as the origin). Since the position of the obstacles in the grid map, which is obtained from the LiDAR scanner, are related to the odometry during the Gmapping process, while the start position and direction of the robot is viewed as the origin and x-axis in the odom_frame⁸. On the other hand, the VLP-map is global, which is consist with the floorplan/blueprint of the building. The origin of the VLP-map is artificially defined, which is related to the installation coordinate of the LED luminaries. If the origin of the VLP-map and SLAM-map are not aligned, the integration between VLP and SLAM would be mismatched.

To overcome this problem, a VLP-constrained Gmapping method is proposed to ensure the alignment between the SLAM-map and VLP-map. Since the obstacle or occupancy information detected in the SLAM-map is collected from LiDAR based on the current pose of the mobile robot. The SLO-VLP is adopted into the Gmapping SLAM as the initialization point. Then, the accuracy of mileage information from the odometry and LiDAR can be improved and constrained by fusing the SLO-VLP measurement data. The output SLAM-map would be associated with

⁴[Online]. Available: <https://wiki.ros.org/tf>

⁵[Online]. Available: <http://wiki.ros.org/navigation/Tutorials/RobotSetup/TF>

⁶[Online]. Available: <https://www.ros.org/reps/rep-0105.html#map>

⁷[Online]. Available: https://en.wikipedia.org/wiki/Occupancy_grid_mapping

⁸[Online]. Available: <https://www.ros.org/reps/rep-0105.html#odom>

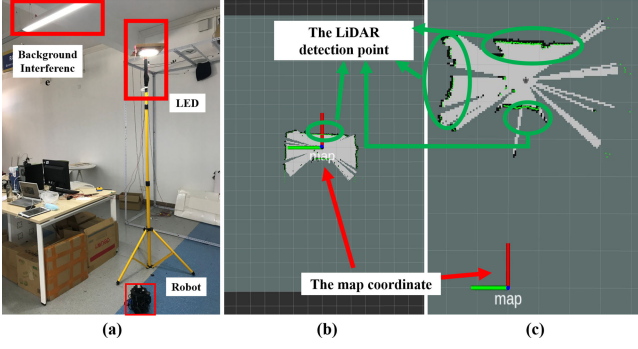


Fig. 3. The grip map made at the beginning of Point B; “map” coordinate is the origin of the map. (a) The test field; (b) The occupancy grid map built by traditional Gmapping; (c) The occupancy grid map built by the proposed VLP-constrained Gmapping.

the LED beacons in the environment, so that the localization of the LiDAR-SLAM can be integrated with the VLP measurement.

Without the VLP-constrained, the origin of the SLAM-map made from traditional Gmapping is the position of the robot at the beginning of the start-up, which may have a drift. For example, the robot starts at Point A, then the origin of the SLAM-map made from Gmapping is Point A, which is same as the origin of the LED feature map (also Point A). However, if the robot starts at Point B, then the origin of the SLAM-map made from Gmapping is Point B, which has a drift from Point A. As can be seen in Fig. 3, without the VLP-constrained, the SLAM-map made from Gmapping is termed as the start point (Point B) as the origin of the SLAM-map. While with the VLP-constrained Gmapping, the origin of the SLAM-map can be drifted to the same point as the origin (Point A) of the pre-built LED landmark map. Thus, the proposed method can align the LiDAR-SLAM results with the VLP global map, so that the localization output of the LiDAR-SLAM can be integrated with the SLO-VLP. After the map-fusion, the information about a prior-unknown environment is stored as a certainty grid map, which is a two-dimensional array of identical units that keep their probability of being occupied by obstacles. This makes great contributions to the navigation function for VLP system.

2) *Loosely-Couple EKF for Localization-Fusion*: After map building, we adopt Adaptive Monte Carlo Localization (AMCL)⁹ for LiDAR-SLAM to track the pose of the robot against the map from VLP-constrained Gmapping. AMCL is a probabilistic localization system for a robot moving in 2D. It uses a particle filter to track the pose of a robot against a known map with the LiDAR scanner match and odometry. Since AMCL is not the contribution in this paper, we would not describe the details of it, we refer readers to the book Probabilistic Robotics¹⁰ for the concept of AMCL.

The full pipeline of the proposed multi-sensor localization fusion based on EKF is illustrated in Fig. 4. Our goal is to estimate the full 3D (x, y, θ) pose of a mobile robot in 2D space, and the state of the robot at time t can be represented as a vector $\mathbf{S}_t = [x_t \ y_t \ \theta_t]^T$. The EKF is adopted for state estimation of dynamic systems, which works as formulas (8) ~ (12). First, the

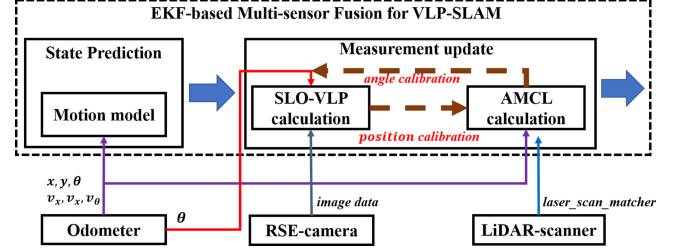


Fig. 4. The pipeline of the proposed EKF-based VLP-SLAM.

current estimated state $\hat{\mathbf{s}}_t$ of the robot is predicted based on the previous state \mathbf{s}_{t-1} with the state transition matrices \mathbf{F} , and the control command \mathbf{u}_t . For our application, the prediction step is based on the motion model of the odometry measurements and the twist.¹¹ During the state prediction, the estimated error covariance, $\hat{\mathbf{P}}_t$, is projected via \mathbf{F} , and then perturbed by the process noise covariance \mathbf{Q}_t .

$$\hat{\mathbf{s}}_t = f(\mathbf{s}_{t-1}, \mathbf{u}_t) = \mathbf{F}\mathbf{s}_{t-1} + \mathbf{B}\mathbf{u}_t \quad (8)$$

$$\hat{\mathbf{P}}_t = \mathbf{F}\mathbf{P}_{t-1}\mathbf{F}^T + \mathbf{Q}_t \quad (9)$$

The second step is the measurement update, the EKF incorporates the measurements \mathbf{z}_t from the sensors, including the camera measurements from our SLO-VLP and the LiDAR scanning from AMCL, as input to correct its state estimation. The EKF gain \mathbf{K}_t can be calculated by using the observation matrix \mathbf{H} and the estimated error covariance $\hat{\mathbf{P}}_t$:

$$\mathbf{K}_t = \hat{\mathbf{P}}_t \mathbf{H}^T [\mathbf{H}\hat{\mathbf{P}}_t\mathbf{H}^T + \mathbf{R}_t]^{-1} \quad (10)$$

where, \mathbf{R}_t is measurement noise covariance. Finally, the gain \mathbf{K}_t is also used to update the state vector and covariance matrix:

$$\mathbf{s}_t = \hat{\mathbf{s}}_t + \mathbf{K}_t(\mathbf{z}_t - \mathbf{H}\hat{\mathbf{s}}_t) \quad (11)$$

$$\mathbf{P}_t = (\mathbf{I} - \mathbf{K}_t\mathbf{H})\hat{\mathbf{P}}_t \quad (12)$$

While the cumulative error or drift from the odometry can be estimated and compensated by the AMCL, since it can publish a *map_frame* to *odom_frame* transform, which essentially fixes the pose of the robot in the map frame.

IV. EVALUATION

In this section, we evaluate our system through real-world experiments, shown in Fig. 5(a). We firstly evaluate the performance of the proposed VLP-constrained Gmapping, which shows that our VLP-constrained Gmapping can drift or align the origin of the SLAM building map with the floorplan map. After that, we verify the property of the proposed multi-sensor fusion for VLP-SLAM, and show the strong performance and the robustness of our method compared with the state-of-the-art (SOTA) methods. Also, we embed the navigation function into our system to present the navigation and obstacle avoidance function, as an indication of the overall positioning and navigation performance.

⁹[Online]. Available: <http://wiki.ros.org/amcl>

¹⁰[Online]. Available: <http://www.probabilistic-robotics.org/>

¹¹[Online]. Available: <http://wiki.ros.org/navigation/Tutorials/RobotSetup/Odom>

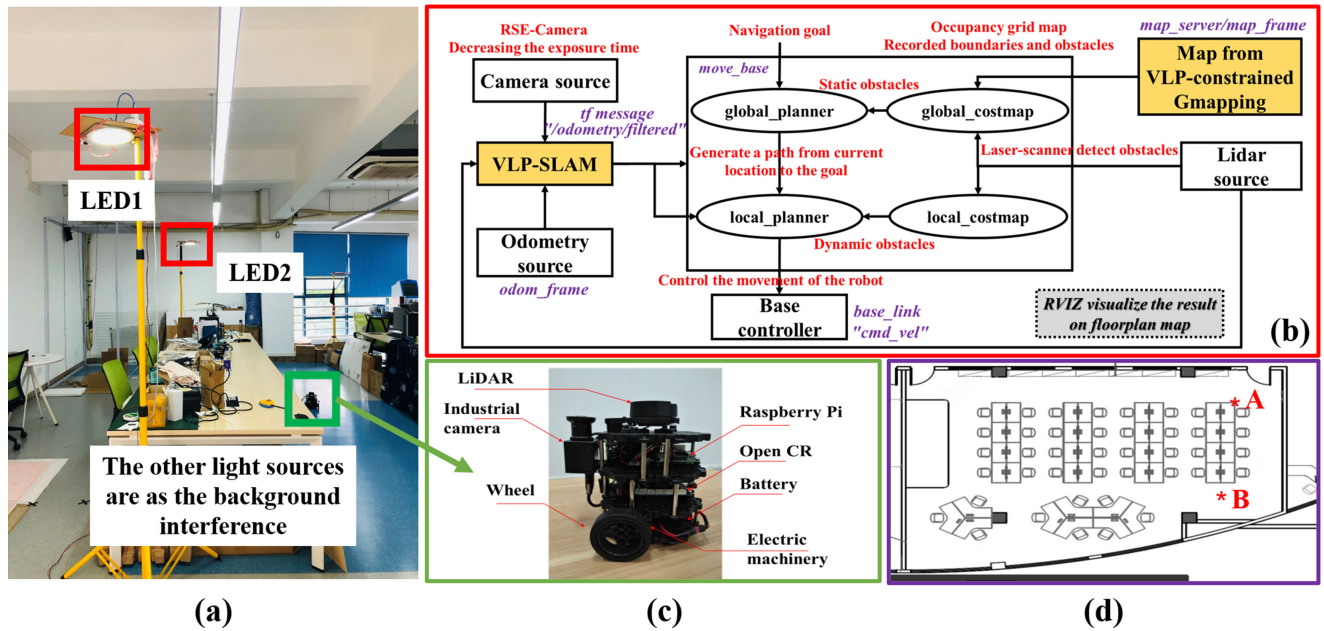


Fig. 5. (a) Experimental platform; (b) The computational graph for robot navigation based on our VLP-SLAM; (c) The mobile robot; (d) The indoor map (A point is the origin of the pre-built LED landmark map, B point is selected as a random point different from the origin).

A. Experimental Settings

We setup a room-sized (around $8.4 \times 4.9 \text{ m}^2$) test field with 4 LEDs mounted on the yellow pole with 2.2 m high. The experiments are performed on a Raspberry Pi 3B mobile robot (Turtlebot 3 Burger¹² with Quad ARM Cortex-A53 Core 1.2 GHz Broadcom BCM2837 64 bits CPU and 1 GB RAM), which runs an Ubuntu Mate 16.04 OS equipped with a robot operating system (ROS¹³). The LEDs' radiation surface has a circular shape of size 18.0 cm in diameter. We control the RSE camera sensor (MindVision UB-300) settings by minimizing the exposure time, so as to see clear strip patterns from the modulated LEDs. The image stream is captured at around 6 Hz with a 2048×1536 resolution and recorded as ROS bags. The LiDAR sensor (HLS-LFCD-LDS¹⁴) we used is a 2D LiDAR scanner capable of sensing 360 degrees with one-degree angular resolution and 12 cm \sim 3.5 m detection distance. The LiDAR is connected and powered by the Raspberry Pi to collect a set of data around the robot to use for SLAM. The odometry and LiDAR is sampled at around 24 HZ and 5 HZ, respectively. Yet for these sensors, hardware synchronization is not available. Therefore, a multithreading loose synchronization framework is used in our algorithm. We run our algorithm on a laptop computer (Intel(R) Core(TM) i5-10210U CPU @ 1.60GHz 2.11 GHz, Ubuntu 18.04) using the recorded bags from the robot. The major computation is done by the computer, and the physical components like the LiDAR sensors, RSE-camera, odometry, etc. are interfaced through the Raspberry Pi robot, which communicates serially with the computer through publishing or subscribing the message of the related ROS packages. The indoor

map of our lab (R&C lab, Innovation and Entrepreneurship Incubation Base, SCUT) is shown in Fig. 5(d), and is changed to the occupancy grid map so that it can be visualized in the RVIZ.¹⁵

B. Map-Fusion Performance

In this section, we firstly study the performance of our map-fusion. The pre-built LED landmark map is fixed and predefined along with the floorplan map of the building (Fig. 5(d)). With the proposed VLP-constrained Gmapping process, the VLP-map and the generated SLAM map can be aligned and integrate together, so that the fusion between the VLP and LiDAR-SLAM is meaningful. It is worth to mention that, in this paper, we don't evaluate the map building accuracy, since the Gmapping is a related accurate map building method [47], [48]. The comparison of grid maps made through the traditional Gmapping (beginning of Point A and Point B in Fig. 5. d), and the proposed VLP-constrained Gmapping (beginning of Point B) are visualized in Fig. 6, where we can see that the resulting maps have similar structure.

Although the mapping performances among these three maps are similar and good (fit the floorplan of our lab), the origin of the map is different. As can be seen in Fig. 7 (visualize the robot in Fig. 6), we visualize the robot (stays at the same Point B with the same SLO-VLP calculation result) in RVIZ. However, in Fig. 7(b), the position of the robot on the map drifts out of range, which is caused by the drift or mismatch of the origin between the VLP-map and the SLAM-map. Since the SLAM and VLP are map-based localization method, when the origin of these two maps is not aligned, the localization fusion between them would be mismatched. Therefore, our VLP-constrained Gmapping can

¹²[Online]. Available: <https://emanual.robotis.com/docs/en/platform/turtlebot3/overview/>

¹³[Online]. Available: <https://www.ros.org/>

¹⁴[Online]. Available: http://wiki.ros.org/hls_lfcd_lds_driver

¹⁵[Online]. Available: <http://wiki.ros.org/rviz>

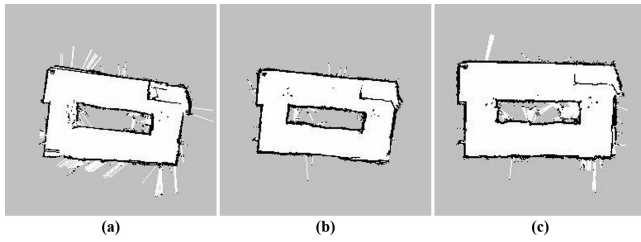


Fig. 6. The resulting map (occupancy grid map) through traditional Gmapping (a) start at Point A; and (b) start at Point B; (c) The resulting map through proposed VLP-constrained Gmapping start at Point B.

align the LiDAR scan relative to the pre-built LED landmark map. Based on Fig. 6 and Fig. 7, we can know that our proposed VLP-constrained Gmapping has the same performance with the traditional Gmapping, at the same time, it can drift the origin of the occupancy grid map to the origin of the LED feature map, so that it can realize the alignment between the localization output of LiDAR-SLAM and VLP.

C. Localization-Fusion Performance

1) *Positioning Accuracy*: To evaluate the positioning accuracy of our multi-sensor localizer, two series of experiments are carried out. The first series are used to test the performance of motionless objects. 900 locations were randomly chosen in the experimental field to evaluate the positioning accuracy of the proposed multi-sensor fusion for VLP-SLAM. Without loss of generality, we also simultaneously calculate the positioning result from the SLO-VLP (900 samples) and the AMCL (50 samples) in the experiment. The results can be seen in Fig. 9. The average positioning accuracy of the proposed multi-sensor fusion for VLP-SLAM is 2.48 cm with a maximum error of 5.73 cm, which is close to that of SLO-VLP, which is 2.23 cm with maximum error of 5.74 cm (maintains as similar level to our previous work [17] [49]). While the average accuracy of the AMCL (LiDAR-SLAM) is 6.97 cm with a maximum error of 14.18 cm.

2) *Robust and Real-Time Pose Estimation*: To further assess the real-time localization performance, we collect the data in trials to show the positioning performance among the SLO-VLP, AMCL, and our proposed VLP-SLAM during the movement of the robot. Fig. 9 shows the estimated trajectories, which travel around 17.6 m. Due to the limitation of our hardware, we cannot provide the motion capture system to obtain the ground-truth of a moving robot. Therefore, we do not discuss the accuracy for the dynamic localization. For the proposed VLP-SLAM method, on the one hand, the SLO-VLP can provide high accuracy positioning, which can not only used as the pose measurement for the robot but also can be used as the pose initialization and calibration for the AMCL. On the other hand, with the AMCL and the odometry, it still can provide localization when the LED is outage (no LED coverage region). Therefore, the derived pose estimates from our VLP-SLAM can get smoothed temporally and can be sustained over a short time when LEDs are not available. It is worth to mention that the AMCL is the calculation of robot pose through the LiDAR scanner. When the robot is under no-LED region, the LiDAR localization calculation

(AMCL) can be used to compensate the information loss of the VLP, so that the overall robot localization system can work well without LED density deployment.

3) *Real-Time Performance*: Positioning speed is another key factor for localization systems. Especially, when the positioning terminal is moving, it can continuously receive information from the positioning system and calculate the current position in time. In this section, the computational time for position calculation of our VLP-SLAM is continuously measured 200 times to calculate the average positioning time-consumption, as shown in Fig. 10. The average computational time of our VLP-SLAM is around 42ms. There are some special points of positioning time, which fluctuate from about 100 ms to 120 ms. These points are the EKF output when SLO-VLP or LiDAR-SLAM calculation is available, otherwise, the main weight of the EKF output is calculated through the odometry sensor which is high frequency sampled. Please note that the odometry provided data at higher rates than the camera or LiDAR scanner, even when the SLO-VLP is in-calculating (unavailable caused by calculation delay), the odometry prediction step of the EKF is still used. Thanks to the high-rate pose estimation from odometry, our VLP-SLAM can maintain smooth pose estimation and compensate the delay of the VLP or LiDAR-SLAM calculation.

D. Navigation Performance

Once mapping and localization are successfully done, the navigation can be easily achieved since our VLP-SLAM can provide real-time and accuracy localization. As can be seen in in Fig. 5(c), we embed our VLP-SLAM method into the Navigation Stack¹⁶ in ROS to navigate the robot. The robot navigation framework is based on our multi-sensor fusion for VLP and SLAM, which uses the proposed multi-sensor localizer to estimate the pose of the robot, and the map data from the proposed VLP-constrained Gmapping for path planning. ROS package `move_base`¹⁷ is adopted to accomplish autonomous navigation. This package uses the localization information from our multi-sensor localizer and provides move-commands to the mobile base of the robot to move safely in the environment without colliding with obstacles. The framework maintains two costmap¹⁸ each for the local and global planner. The global planner is based on A Star search algorithm, while the local planner uses the dynamic window approach [50]. Once a navigation goal is received, the global path planning algorithm takes charge of generating an appropriate path from the start pose to the target pose, while the local path planning is responsible for generating velocity commands. Thanks to the reliable and accurate pose estimation from our VLP-SLAM fusion, the robot can move steadily according to the planned path to reach the designated location.

During the automatic navigation, the VLP provides high accuracy pose estimation, initialization, and correction for SLAM, which contributes to pose initialization for global path planning. Odometry makes great contributions to maintain real-time capability and compensate the delay of the VLP or SLAM

¹⁶[Online]. Available: <http://wiki.ros.org/navigation/Tutorials/RobotSetup>

¹⁷[Online]. Available: http://wiki.ros.org/move_base

¹⁸[Online]. Available: http://wiki.ros.org/costmap_2d

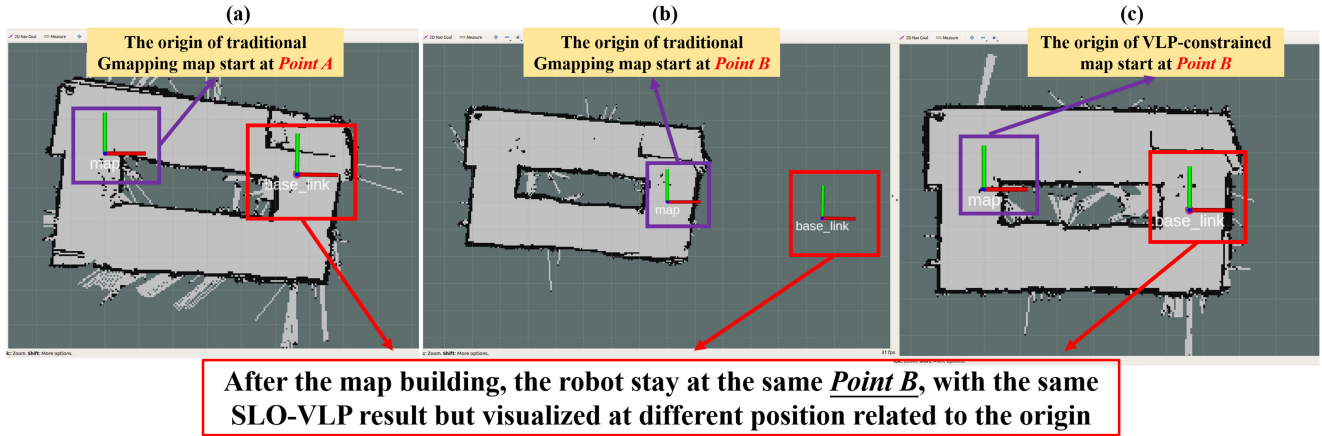


Fig. 7. Using the RVIZ to visualize the position of the robot (stay at same point B; “map” coordinate is the origin of the map; “base_link” coordinate is the position of the robot through SLO-VLP calculation): (a) The SLAM-map made by traditional Gmapping starting at Point A; (b) The SLAM-map made by traditional Gmapping start at Point B; (c) The SLAM-map made by proposed VLP-constrained Gmapping start at Point B.

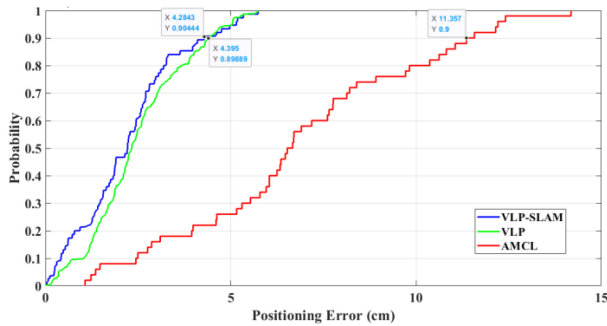


Fig. 8. CDF curve of the positioning error for the VLP-SLAM fusion, SLO-VLP, and AMCL.

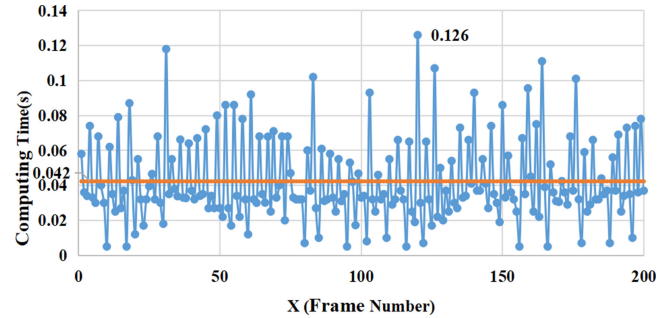


Fig. 10. The measured positioning / calculation time of our Multi-sensor fusion for VLP-SLAM.

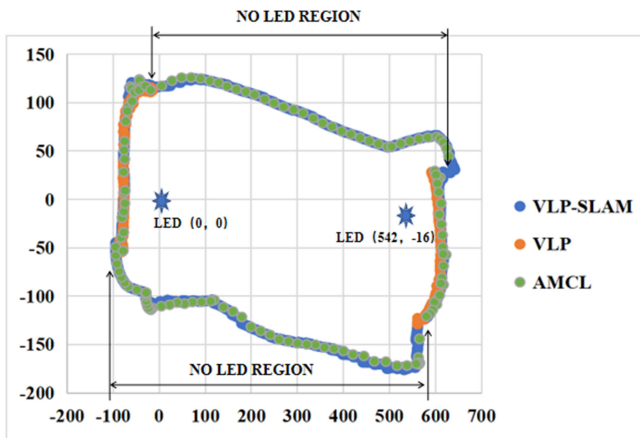


Fig. 9. The trajectory estimation of our multi-sensor fusion for the VLP-SLAM method with SLO-VLP and AMCL. (a) Test One and (b) Test Two for similar paths.

calculation. While the LiDAR-SALM can compensate the cumulative error from the odometry, and the dynamic obstacles can be detected through the LiDAR scanner via the update of the local cost-map to realize the dynamic obstacle avoidance. All of

these contribute to the reliable navigation for robotics. Thus, the crucial problem of autonomous initialization and localization in robotics can be solved. For the details of the navigation test, we strongly recommend readers to our video demo:

- <https://www.bilibili.com/video/BV1Qw411Z7HX>
- <https://www.bilibili.com/video/BV1EP4y1h7Fy/>
- <https://www.bilibili.com/video/BV1vq4y1K7Tx/>
- <https://www.bilibili.com/video/BV1JX4y137Q7>

E. Discussion

In this section, we compare the performance of our VLP-SLAM multi-sensor localizer with the SOTA works in the field of VLP in Table I. The average positioning accuracy, computing time, and the density of the LEDs in the related experimental platform are also reported objectively. Compared to other approaches, our method performs favorably, with less latency (42 ms), high accuracy (2.5 cm) and no LED density deployment required. It is worth to mention that, without losing generality, the proposed multi-sensor fusion for VLP-SLAM method has also been evaluated in the Hong Kong University of Science and Technology to show the reproducibility. Extensive experiments evaluation can be seen in our technology report [49]. Note that the experiment setup is different between this work and our technology report [49] with same algorithm, however, based on

TABLE I
PERFORMANCE COMPARISON WITH THE SOTA

Method	Average Accuracy (cm)	Time Cost (ms)	Require LEDs (at least)	Hardware	Dimension	Receiver Type	Density LEDs-Coverage ($L \times W \times H$)m ³
R. [51]	14.0	<i>N/A</i>	3	STM32F103+computer	2D	PD+IMU	7 $-(1.5 \times 1.5 \times 2.5)$
R. [1]	3.9	44	2	Raspberry+computer	3D+yaw	Camera	9 $-(1.8 \times 1.8 \times 2.0)$
R. [14]	0.8	400	2	Raspberry+computer	3D+yaw	Camera	4 $-(1.0 \times 1.0 \times 1.5)$
R. [33]	17.5	<i>N/A</i>	1*	Smartphone	2D+3D	Camera	1 $-(3.0 \times 3.0 \times 2.0)$
R. [52]	3.2	43	1*	computer	3D+ysaw	Camera	1 $-(0.8 \times 0.8 \times 2.0)$
R. [53]	2.3	60	1*	computer	2D+yaw	Camera	1 $-(1.8 \times 1.8 \times 3.0)$
R. [34]	5.5	<i>N/A</i>	1	Smartphone	6D	Camera+IMU	1 $-(2.7 \times 1.8 \times 1.5)$
R. [22]	13.4	<i>N/A</i>	1	Smartphone	3D+yaw	Camera+PDR	1 $-(1.0 \times 1.0 \times 2.6)$
R. [16]	5.0	<i>N/A</i>	1	Raspberry+computer	3D+yaw	Camera+IMU	23 $-(5.0 \times 4.0 \times 2.3)$
R. [15]	5.0	50	1	Raspberry+computer	3D+yaw	Camera+IMU	25 $-(5.0 \times 4.0 \times 2.3)$
R. [6]	3.0	78#	1	Raspberry+computer	2D+yaw	Camera+encoder+gyroscope	4 $-(2.0 \times 2.0 \times 1.5)$
R. [6]	3.0	131.8#	1	Raspberry+computer	2D+yaw	Camera+encoder+gyroscope	4 $-(2.0 \times 2.0 \times 1.5)$
R. [6]	2.2	81.5#	1	Raspberry+computer	2D+yaw	Camera+encoder+gyroscope	4 $-(2.0 \times 2.0 \times 1.5)$
R. [54]	4.3	<i>N/A</i>	1	Raspberry+computer	3D+yaw	Camera+odom	3 $-(5.0 \times 1.8 \times 2.7)$
R. [55]	2.5	184	1	Raspberry+computer	3D+yaw	Camera+odom	3 $-(6.8 \times 2.7 \times 2.7)$
R. [17]	2.0	195	1	Raspberry+computer	3D+yaw	Camera+odom	3 $-(7.0 \times 3.8 \times 2.7)$
R. [17]	2.1	33	1 or 0	Raspberry+computer	3D+yaw	Camera+odom+IMU	3 $-(9.1 \times 4.0 \times 2.7)$
Our [49]	2.0	52	1 or 0	Raspberry+computer	3D+yaw	Camera+odom+LiDAR	4 $-(12.0 \times 10.8 \times 2.7)$
Ours	2.5	42	1 or 0	Raspberry+computer	3D+yaw	Camera+odom+LiDAR	2 $-(8.4 \times 4.9 \times 2.2)$

“*” means that the LED lamp with beacon marker.

“#” means that the position estimation was performed in post-processing, not real-time.

the probabilistic theory, the localization result might have slight fluctuation in different platform, but maintains a similar level (2.0 cm in [49]). While the real-time performance is different since the processor is different (52 ms in [49]). But these can show the stability of the accuracy of our method is also evaluated. As for the LED density coverage, the VLP coverage area depends on the FOV of the camera and the height of the LED-installation, our multi-sensor fusion for VLP-SLAM method can perform well under LED outage region through LiDAR and odometry.

Compared to the VLP-aided IMU method [6], [15]–[17], which suffers from bias or drift during long-term running. Our work can cope with the cumulative error through the LiDAR scan matching measurement. Through the multi-sensor fusion, especially the LiDAR scanning compensation, the orientation angle from the odometry can perform well with considerable drift after long-term experiments. The single LED VLP, which is based on the orientation angle from the odometry, can obtain low-drift and accuracy pose estimation. Therefore, the proposed multi-sensor fusion for VLP-SLAM can not only achieve well balance among accuracy, real-time performance, and robustness, but also compensate the deficiencies of stand-alone sensors and provide more reliable, more robust positioning and navigation.

For the real-time performance, please note that we run the whole image process and data collection on a low-cost embedded platform Raspberry Pi 3B with a large-size captured image of 2048×1536 without any code optimization for ARM processors. After that, the data (including the VLP observation, LiDAR scanning, and odometry data) are transmitted to the laptop which takes charge of multi-sensor fusion pose estimation and visualizes the result in RVIZ. The time cost here is the whole process covering both the image process and pose estimation (runtime on both laptop and Raspberry), and also the time-cost of data transmission from robot to the laptop. Compared to the similar process platform with image size of 1640×1232 in Ref. [15], and 640×480 in Ref. [6], our proposed multi-sensor fusion for VLP-SLAM is more efficient, and hence lightweight to be used on resource-constrained computational platforms.

For the pose initialization, since the odometry only provides relative motion measurements, the initial pose is needed with

the direction/orientation with respect to the map. For LiDAR-SLAM localization, the AMCL also needs to initialize its filters, which is generally done with an initial pose message given by the user. In our VLP-SLAM method, the VLP can provide position initialization for the AMCL, while the AMCL can provide reliable orientation estimation for the SLO-VLP. Based on this, the robot no longer needs to head along a fixed direction (e.g., x-axis) before running. When the manually initial pose (position and orientation) of the robot is not accurate, the LiDAR-SLAM can adjust the orientation of the robot, while the SLO-VLP can provide higher accuracy position estimation with the more correct orientation from the LiDAR. Meanwhile, the high position estimation from SLO-VLP can update the overall pose estimation for the EKF, which can also act on the AMCL to improve the measurement from the AMCL. The interaction between VLP and LiDAR-SLAM provides reliable and accurate pose estimation. The details of the random pose initialization can be visualized through the video demo.¹⁹

Furthermore, for LiDAR-SLAM, it is difficult to estimate the pose in the large space or long corridors without a sufficient variety of observations. Since similar geometric regions in these environment may mislead the convergence of AMCL. When the robot is not under the VLP coverage area, the wrong initialization for LiDAR-SLAM would cause the robot getting lost since the geometric regions of these two parts are quite similar. However, when VLP is available, the position of the robot can be calibrated to the correct position. The evaluation of the VLP calibration for LiDAR-SLAM can be seen in the video demo.²⁰

V. CONCLUSION

In this paper, we pursued reliable, real-time, and accurate pose estimation for mobile robot through the multi-sensor fusion method based on EKF for VLP-SLAM. We relaxed the

¹⁹Pose initialization for VLP-SLAM: [Online]. Available: <https://www.bilibili.com/video/BV1VX4y1G7wN>

²⁰VLP calibration for SLAM: [Online]. Available: <https://www.bilibili.com/video/BV1iA411N76g/>

assumption on the minimum number of concurrently observable LEDs from three to zero for the RSE-camera based VLP system through integrating with odometry and LiDAR sensor. While the cumulative error of the odometry can be compensated and corrected through the LiDAR scanning. The experiment result shows that the our method has strong robustness with low latency (42 ms) and high accuracy (2.5 cm), which can achieve well balance among accuracy, real-time ability and coverage. In future work, we will deeply study the tightly coupled VLP and SLAM for incorporating the LED markers into the maps, which could construct the global map and optimize poses with sufficient features, together with a method to model and reduce data uncertainty.

REFERENCES

- [1] P. Lin *et al.*, "Real-time visible light positioning supporting fast moving speed," *Opt. Exp.*, vol. 28, no. 10, pp. 14503–14510, 2020.
- [2] J. Armstrong, Y. A. Sekercioglu, and A. Neild, "Visible light positioning: A roadmap for international standardization," *IEEE Commun. Mag.*, vol. 51, no. 12, pp. 68–73, Dec. 2013.
- [3] N. Huang *et al.*, "Design and demonstration of robust visible light positioning based on received signal strength," *J. Lightw. Technol.*, vol. 38, no. 20, pp. 5695–5707, 2020.
- [4] Y. Cai, W. Guan, Y. Wu, C. Xie, Y. Chen, and L. Fang, "Indoor high precision three-dimensional positioning system based on visible light communication using particle swarm optimization," *IEEE Photon. J.*, vol. 9, no. 6, pp. 1–20, Dec. 2017.
- [5] B. Chen *et al.*, "Performance comparison and analysis on different optimization models for high-precision three-dimensional visible light positioning," *Opt. Eng.*, vol. 57, no. 12, 2018, Art. no. 125101.
- [6] R. Amsters *et al.*, "Visible light positioning using Bayesian filters," *J. Lightw. Technol.*, vol. 38, no. 21, pp. 5925–5936, 2020.
- [7] Y. Yang, J. Hao, and J. Luo, "CeilingTalk: Lightweight indoor broadcast through LED-camera communication," *IEEE Trans. Mobile Comput.*, vol. 16, no. 12, pp. 3308–3319, Dec. 2017.
- [8] Y. Kuo *et al.*, "Luxapose: Indoor positioning with mobile phones and visible light," in *Proc. 20th Annu. Int. Conf. Mobile Comput. Netw.*, 2014, pp. 447–458.
- [9] S. Song *et al.*, "Employing DIALux to relieve machine-learning training data collection when designing indoor positioning systems," *Opt. Exp.*, vol. 29, no. 11, pp. 16887–16892, 2021.
- [10] Y. Chuang *et al.*, "Visible light communication and positioning using positioning cells and machine learning algorithms," *Opt. Exp.*, vol. 27, no. 11, pp. 16377–16383, 2019.
- [11] Y. Wu *et al.*, "Received-signal-strength (RSS) based 3D visible-light-positioning (VLP) system using kernel ridge regression machine learning algorithm with sigmoid function data preprocessing method," *IEEE Access*, vol. 8, pp. 214269–214281, 2020.
- [12] F. An *et al.*, "A tilt visible light positioning system based on double LEDs and angle sensors," *Electron.*, vol. 10, no. 16, 2021, Art. no. 1923.
- [13] 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.
- [14] W. Guan *et al.*, "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.
- [15] Q. Liang and M. Liu, "A tightly coupled VLC-inertial localization system by EKF," *IEEE Robot. Automat. Lett.*, vol. 5, no. 2, pp. 3129–3136, Apr. 2020.
- [16] 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.*, Oct. 2019, pp. 1–8.
- [17] W. Guan *et al.*, "Robust robotic localization using visible light positioning and inertial fusion," *IEEE Sensors J.*, early access, 2021. doi: [10.1109/JSEN.2021.3053342](https://doi.org/10.1109/JSEN.2021.3053342).
- [18] V. Nguyen, A. Harati, and R. Siegart, "A lightweight SLAM algorithm using orthogonal planes for indoor mobile robotics," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct./Nov. 2007, pp. 658–663.
- [19] C. Hsu *et al.*, "Visible light positioning and lighting based on identity positioning and RF carrier allocation technique using a solar cell receiver," *IEEE Photon. J.*, vol. 8, no. 4, pp. 1–7, Aug. 2016.
- [20] C. Hong *et al.*, "Angle-of-arrival (AOA) visible light positioning (VLP) system using solar cells with third-order regression and ridge regression algorithms," *IEEE Photon. J.*, vol. 12, no. 3, pp. 1–5, Jun. 2020.
- [21] C. Hong *et al.*, "Visible light positioning (VLP) system using low-cost organic photovoltaic cell (OPVC) for low illumination environments," *Opt. Exp.*, vol. 28, no. 18, pp. 26137–26142, 2020.
- [22] 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.
- [23] P. Hu *et al.*, "High speed LED-to-Camera communication using color shift keying with flicker mitigation," *IEEE Trans. Mobile Comput.*, vol. 19, no. 7, pp. 1603–1617, Jul. Apr. 2019.
- [24] R. Qv *et al.*, "A high efficient code for visible light positioning system based on image sensor," *Access*, vol. 7, pp. 77762–77770, 2019. [Online]. Available: <https://search.datacite.org/works/10.1109/access.2019.2921601>, doi: [10.1109/access.2019.2921601](https://doi.org/10.1109/access.2019.2921601).
- [25] W. Guan *et al.*, "Performance analysis and enhancement for visible light communication using CMOS sensors," *Opt. Commun.*, vol. 410, pp. 531–551, 2018.
- [26] J. He *et al.*, "Efficient sampling scheme based on length estimation for optical camera communication," *IEEE Photon. Technol. Lett.*, vol. 31, no. 11, pp. 841–844, Jun. 2019.
- [27] Z. Zhou *et al.*, "Performance enhancement scheme for RSE-based underwater optical camera communication using de-bubble algorithm and binary fringe correction," *Electronics*, vol. 10, no. 8, 2021, Art. no. 950.
- [28] Z. Zhou, S. Wen, and W. Guan, "RSE-based optical camera communication in underwater scenery with bubble degradation," in *Proc. Opt. Fiber Commun. Conf. Exhib.*, 2021.
- [29] S. Chen and W. Guan, "High accuracy VLP based on image sensor using error calibration method," 2020, *arXiv:2010.00529*.
- [30] W. Guan *et al.*, "High-precision indoor positioning algorithm based on visible light communication using complementary metal-oxide-semiconductor image sensor," *Opt. Eng.*, vol. 58, no. 2, 2019, Art. no. 024101.
- [31] W. Guan *et al.*, "High precision indoor visible light positioning algorithm based on double LEDs using CMOS image sensor," *Appl. Sci.*, vol. 9, no. 6, 2019, Art. no. 1238.
- [32] H. Lee *et al.*, "Rollinglight: Enabling line-of-sight light-to-camera communications," in *Proc. 13th Annu. Int. Conf. Mobile Syst., Appl., Serv.*, 2015, pp. 167–180.
- [33] R. Zhang, W. Zhong, Q. Kema, and S. Zhang, "A single LED positioning system based on circle projection," *IEEE Photon. J.*, vol. 9, no. 4, pp. 1–9, Aug. 2017.
- [34] H. Cheng, C. Xiao, Y. Ji, J. Ni, and T. Wang, "A single LED visible light positioning system based on geometric features and CMOS camera," *IEEE Photon. Technol. Lett.*, vol. 32, no. 17, pp. 1097–1100, Jul. 2020.
- [35] C. Cadena *et al.*, "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age," *IEEE Trans. Robot.*, vol. 32, no. 6, pp. 1309–1332, Dec. 2016.
- [36] H. Ye, Y. Chen, and M. Liu, "Tightly coupled 3D lidar inertial odometry and mapping," in *Proc. Int. Conf. Robot. Automat.*, May 2019, pp. 3144–3150.
- [37] W. Hess *et al.*, "Real-time loop closure in 2D LIDAR SLAM," in *Proc. IEEE Int. Conf. Robot. Automat.*, May. 2016, pp. 1271–1278.
- [38] B. Steux and O. El Hamzaoui, "tinySLAM: A slam algorithm in less than 200 lines C-language program," in *Proc. 11th Int. Conf. Control Automat. Robot. Vis.*, Dec. 2010, pp. 1975–1979.
- [39] M. Montemerlo *et al.*, "FastSLAM: A factored solution to the simultaneous localization and mapping problem," *Aaai/Laai*, 2002, Art. no. 593598.
- [40] S. Kohlbrecher *et al.*, "Hector open source modules for autonomous mapping and navigation with rescue robots," in *Robot Soccer World Cup*, Springer, Berlin, Heidelberg, 2013, pp. 624–631.
- [41] G. Grisetti, C. Stachniss, and W. Burgard, "Improved techniques for grid mapping with rao-blackwellized particle filters," *IEEE Trans. Robot.*, vol. 23, no. 1, pp. 34–46, Feb. 2007.
- [42] G. Grisetti, C. Stachniss, and W. Burgard, "Improving grid-based slam with rao-blackwellized particle filters by adaptive proposals and selective resampling," in *Proc. IEEE Int. Conf. Robot. Automat.*, Apr. 2005, pp. 2432–2437.
- [43] T. Caselitz, B. Steder, M. Ruhnke, and W. Burgard, "Monocular camera localization in 3D lidar maps," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2016, pp. 1926–1931.
- [44] S. Chan, P. Wu, and L. Fu, "Robust 2D indoor localization through laser SLAM and visual SLAM fusion," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Oct. 2018, pp. 1263–1268.

- [45] W. Guan, X. Chen, M. Huang, Z. Liu, Y. Wu, and Y. Chen, "High-speed robust dynamic positioning and tracking method based on visual visible light communication using optical flow detection and Bayesian forecast," *IEEE Photon. J.*, vol. 10, no. 3, pp. 1–22, Jun. 2018.
- [46] H. Song *et al.*, "Robust LED region-of-interest tracking for visible light positioning with low complexity," *Opt. Eng.*, vol. 60, no. 5, 2021, Art. no. 053102.
- [47] J. M. Santos, D. Portugal, and R. P. Rocha, "An evaluation of 2D SLAM techniques available in robot operating system," in *Proc. IEEE Int. Symp. Saf., Secur., Rescue Robot.*, Oct. 2013, pp. 1–6.
- [48] K. Krinkin *et al.*, "Evaluation of modern laser based indoor slam algorithms," in *Proc. 22nd Conf. Open Innovations Assoc. (FRUCT)*, May 2018, pp. 101–106.
- [49] W. Guan, B. Hussain, and P. Yue, "Technology report: Robotic localization and navigation system for visible light positioning and SLAM," 2021, *arXiv: 2104.14755*.
- [50] E. Marder-Eppstein, E. Berger, T. Foote, B. Gerkey, and K. Konolige, "The office marathon: Robust navigation in an indoor office environment," in *Proc. IEEE Int. Conf. Robot. Automat.*, May 2010, pp. 300–307.
- [51] 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.
- [52] W. Guan, S. Wen, H. Zhang, and L. Liu, "A novel three-dimensional indoor localization algorithm based on visual visible light communication using single LED," in *Proc. IEEE Int. Conf. Automat., Electron. Elect. Eng.*, Nov. 2018, pp. 202–208.
- [53] 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.
- [54] Z. Yan *et al.*, "Multi-robot cooperative localization based on visible light positioning and odometer," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–8, Jun. 2021, Art. no. 7004808.
- [55] L. Huang *et al.*, "Single LED positioning scheme based on angle sensors in robotics," *Appl. Opt.*, vol. 60, no. 21, pp. 6275–6287, 2021.

Weipeng Guan received the B.E. and M.E. degrees from the South China University of Technology, China, in 2016 and 2019, respectively. He is currently working toward the Ph.D. degree. He was working with Samsung Electronics Company Ltd., Huawei Technologies Company Ltd., TP-LINK, the Chinese Academy of Sciences, and The Chinese University of Hong Kong. He has authored or coauthored more than 50 research articles in prestigious international journals and conferences. His research interests include mobile robot, optical camera communications, visible light positioning, sensor fusion and so on.

Linyi Huang will receive the B.S. degree from the South China University of Technology in 2022, and is the Undergraduate Student with the South China University of Technology, majoring in automation science and engineering. Her current research interests include VLC-based image communication and ROS based robot localization.

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 authored or coauthored hundreds of articles. His research interest focuses on visible light communication.

Zihong Yan will receive the B.S. degree from the South China University of Technology in 2022, and is the Undergraduate Student with the South China University of Technology, Guangzhou, China, majoring in automation science and engineering. His current research interests include visible light positioning, robot localization, and multirobot cooperative localization.

Wanlin Liang will receive the B.S. degrees from the South China University of Technology in 2023, and is the Undergraduate Student with the South China University of Technology, majoring in automation science and engineering. Her current research interests include VLC-based image communication and ROS based robot mapping and localization.

Chen Yang, is a Student with the South China University of Technology and is working toward the Bachelor degree of Science in automation. She entered R&C studio in 2020 and assisted the team do research. Her research interests include picture processing and visible light positioning algorithm.

Ziyu Liu will receive the B.S. degrees from the South China University of Technology in 2023, and is the Undergraduate Student with the South China University of Technology, majoring in automation science and engineering. Her current research interest focuses on ROS based multi robot cooperative slam.