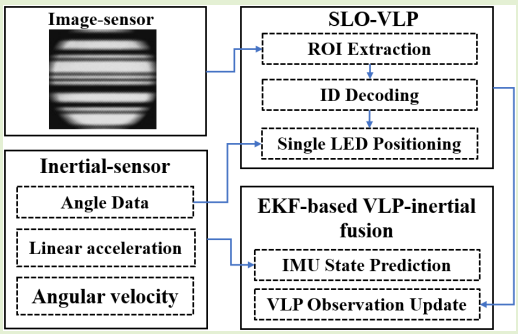
Robust Robotic Localization Using Visible Light Positioning and Inertial Fusion

Weipeng Guan, *Student Member, IEEE*, Linyi Huang, Babar Hussain, and C. Patrick Yue, *Fellow, IEEE*

***Abstract*—Accurate indoor positioning is critical in the field of location-based services and robotics. Visible light positioning (VLP) technology is a promising technique as it can provide high accuracy positioning based on the existing lighting infrastructure. However, it is difficult to meet the requirement of multiple LED anchors in range for successful and accurate positioning. In this article, we proposed a loosely-coupled VLP-inertial fusion method for VLP, with an inertial measurement unit (IMU) and rolling shutter camera, to improve positioning robustness under LED shortage/outage. The efficacy of the proposed VLP scheme as well as the robustness under LED outage, handover situation and background light interference, are verified by real-world experiments. The results show that our proposed scheme can provide an average accuracy of 2.1 cm (stationary localization) and the average computational time in low-cost embedded platforms is around 33 ms.**



***Index Terms*—Visible light positioning (VLP), robust indoor positioning, high accuracy, sensors fusion, inertial measurement unit (IMU), robotic localization.**

I. Introduction

With the increasing demand for indoor location-based services, indoor localization technology has been intensively studied in recent years. However, limited by penetration capability and multipath effect, the widely-used outdoor positioning technique, global positioning systems (GPSs), are restricted in indoor scenarios. In addition, conventional radio frequency (RF)-based indoor positioning technologies in terms of Wi-Fi, Zigbee, Bluetooth, Radio-frequency Identification (RFID) and ultra-wideband (UWB), have inherent flaws, such as, low accuracy, limited coverage, high hardware costs, and electromagnetic interference [1]. On the contrary, visible light positioning (VLP) technology uses illumination LED to transmit position information by directly modulating the light intensity at a high frequency, which is invisible to human eyes and can realize illumination and positioning simultaneously. VLP can also provide relative high accuracy since it is more immune to multipath effects and fading. Furthermore, the VLP does not cause any electromagnetic interference, thus making it suitable for use in sensitive areas such as hospitals and airplanes [2].

Manuscript received December 28, 2020; revised January 19, 2021; accepted January 19, 2021. Date of publication January 21, 2021; date of current version March 14, 2022. This work was supported by the Hong Kong Research Grants Council under General Research Fund (GRF) project no. 16215620 and the HKUST-Qualcomm Joint Innovation and Research Laboratory. The associate editor coordinating the review of this article and approving it for publication was Dr. Valérie Renaudin. *(Corresponding author: C. Patrick Yue.)*

Weipeng Guan, Babar Hussain, and C. Patrick Yue are with the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR, China (e-mail: eepatrick@ust.hk).

Linyi Huang is with the School of Automation Science and Engineering, South China University of Technology, Guangzhou 510641, China.

Digital Object Identifier 10.1109/JSEN.2021.3053342

A. Motivation

VLP methods can be divided into two categories according to the type of receiver: PD-based [3]–[6] and image sensor-based [7]–[10]. PD is not an ideal VLP device, because it is sensitive to the light intensity variation and the diffuse reflection of the light signal [11]. On the contrary, image sensor-based VLP is favored in both commerce and industry, since the good compatibility with user devices, such as mobile robots and smartphones. The coordinate information of each LED is transmitted in the format of temporally-varying intensity signals and mapped to spatially-varying striped patterns on the camera image based on the rolling shutter effect (RSE). Through the region-of-interest (ROI) of the LED extraction [12], ID decoding, and positioning calculation through imaging geometry, the coordinate of the positioning terminal can be obtained based on the global map.

In our previous works [8]–[10], we proposed high accuracy VLP algorithm based on triple, double, and single LEDs to realize centimeter-level accuracy. In [13], we firstly adopt the VLP using double LEDs into robotics localization, which achieves a positioning accuracy within 1 cm with positioning latency lower than 80 ms. In [7], a VLP system based on

mobile phone is proposed and achieves an accuracy of 7.5 cm with moving speeds of up to 18 km/h. In [1], a lightweight VLP scheme implemented on a Raspberry Pi can achieve a positioning accuracy of 3.93 cm and support the moving speed up to 38.5 km/h. Despite the promising performance of existing VLP systems, there remain some practical challenges, especially the LED shortage problem, which is mainly caused by the deployment density of LEDs, obstruction of the line-of-sight (LOS) views, limited field-of-view (FOV) of the image sensor, and so on. Although there exist some centimeter or decimeter-level accuracy VLP schemes based on single LED [8], [14], [15], it is still difficult to ensure the requirement of one known LOS LED in every corner and any area of the practical scenarios. Especially, the incomplete VLP observation due to the LOS blockage caused by a moving human body or any other items, which is the most serious problems for VLP practical application. Additionally, sensors are imperfect, and their measurements are prone to errors. The measurement noise of the VLP, which is caused by fabrication error of the image sensors, is also the bottleneck of the positioning accuracy in the field of image sensor-based VLP [11]. By fusing the data from multiple sensors, we can obtain an overall and optimal position estimation whose error is less than only using VLP.

B. Contribution

In this work, we propose a loosely-coupled VLP-inertial fusion method for robotics using RSE-based camera with single LED for VLP, and inertial measurement unit (IMU). More specifically, we employ an extended Kalman filter (EKF) for real-time 3D pose estimation (position and orientation) by fusing the relative pose measurements from the IMU with the absolute pose from the VLP measurement. As for the VLP measurement, based on our previous work [8], we firstly proposed a VLP algorithm based on single LED adopted with odometer for orientation angle calculation, termed it as single LED VLP with odometer (SLO-VLP), to achieve 3D pose measurement of the robot. Then an EKF-based pose estimation scheme is proposed to stably obtain the global poses of the robot through estimating the state transition and predicted covariance from the IMU, and updating them by the SLO-VLP observations. When one LED is observed for SLO-VLP calculation, the weight of the SLO-VLP in the optimal state estimate increases, caused by the function of the near-optimal Kalman gain. When there is no LED (without SLO-VLP observation), over a short period of time, our method can still efficiently estimate the 3D pose of the robot, which greatly improves the robustness and usability of the VLP system under LED shortage.

We highlight the contributions as:

- SLO-VLP algorithm is proposed to achieve high accuracy 3D localization just using single LED as positioning anchor.
- An EKF-based loosely-coupled SLO-VLP-inertial fusion method is proposed to handle the situation of VLP unavailable, and we can relax the assumption on the minimum number of concurrently observable LEDs required for positioning to zero. Meanwhile, the measure noise from the SLO-VLP is also updated based on the EKF to eliminate the sensor noise, while the cumulative error from the IMU can be corrected by the VLP observation.
- The proposed SLO-VLP-inertial fusion scheme is evaluated in a real-world environment based on mobile robot under the harsh environment (LED shortage/outage, handover, and background light interference). The efficacy of the proposed localization scheme and accurate 3D pose tracking, are verified with extensive experiments.

C. Organization

The remainder of this article is organized as follows. Section II introduces the related works. Section III explains the proposed methodology, including the principle of SLO-VLP, and EKF-based SLO-VLP and inertial fusion. Section IV presents the experimental evaluation and Section V is the conclusion.

II. Related Work

A. RSE-Based Optical Camera Communication

As we know, LEDs can transmit data over the air by directly modulating the light intensity at a high frequency which is invisible to human eyes but perceivable by camera or PD. Generally, PD is used as the receiver to capture the signal for the visible light communication (VLC) system. Recently, camera based on RSE can be also adopted as the receiver in optical camera communication (OCC) system. The pixels on an RSE image sensor are exposed and read out line by line instead of perceiving light at the same time at a single moment. The "ON" or "OFF" light signal can be transferred into bright or dark patterns in the captured image. For a comprehensive understanding of OCC and decoding in image-based VLP, we refer readers to the previous works [16]–[20]. In this work, we employ our universal VLC modulator [21] for retrofitting LED lamp to transmit the ID information to the RSE-based camera receiver. Each LED lamp is assigned with a unique identity (ID), which is associated with its location and stored in an ID-location database. The designed ID packet begins with a 6-bit preamble (011110), proceeded 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 can be extended by a larger payload. For the details of the VLC modulator, please refer readers to our previous works [21], [22].

B. RSE Camera-Based Visible Light Positioning

The LOS and weak perspective projection property provide information about the camera's relative position and rotation with respect to each detected LED. Decimeter/centimeter-level accuracy has been reported in some RSE-camera based VLP systems [1], [7], [9]–[11], [23]–[26]. Yet, all these camera based VLP systems treat LED lamps as point sources without geometric information. Thus, multiple known LED anchors are required at a time for trilateration or triangulation. Such approaches rely on dense deployment of LED beacons. To reduce the number of required LEDs in VLP, [14] no longer treat the captured LED images as a point as in the existing works, but as an image whose geometric features are exploited to determine the receiver's orientation and location relative

to the reference LED lamp. However, this method requires 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. Reference [27] efficiently relax the assumption on the minimum number of simultaneously observable LEDs from three to one, through the tightly-coupled visual-inertial fusion method for VLP. However, they still build up the experimental platform in the area of 5m × 4m × 2.3m with 23 LEDs, since their system required more than 4 LEDs for initialization. Reference [28] proposed a hybrid positioning algorithm consisting of VLP and pedestrian dead reckoning (PDR) to achieve decimeter-level accuracy. Reference [29] proposed single LED VLP using image and attitude information from mobile phone to achieve average accuracy of 4.45 cm. However, all of those works cannot provide stable and high accuracy under the situation of without LED. Besides, the ID-decoding rate cannot maintain 100%, which would also cause failure or major errors in VLP, and multi-sensor fusion has great potential for improving the robustness of VLP application. Therefore, in this article, we propose VLP and inertial fusion scheme to improve positioning robustness and maintain the positioning accuracy under the harsh situation of without LED, LOS blockage of VLP, or even failure of ID decoding.

III. Methodology

A. Our SLO-VLP

In our proposed scheme, the time-varying VLP signals from LEDs are perceived by the RSE-camera, which is installed vertically on the robot, as spatially-varying strip patterns. Through the VLP observation, the state of the robot can be estimated in time. To do so, firstly, we need to find the centroid imaging location or the ROI of the LED in the captured image, and then recognize its ID, and retrieve the 3D position of the captured LED from the registered LED database. After that we can obtain the absolute 3D position and azimuth of the robot from SLO-VLP calculation. Please note that we mainly deal with the VLP and inertial fusion scheme in this article, for the readers that are interested in the low complexity LED-ROI extraction and the efficient LED-ID decoding, please refer to our future works.

1) LED-ROI Extraction Scheme: The strip patterns induced by the modulated LEDs are parallel to the rows in the image and interleaving in the vertical direction, as shown in Fig. 1. For the VLP calculation, the first step is to extract the LED-ROI from the image and obtain the centroid pixel coordinates as camera measurements. To achieve this, a robust ROI extraction method with low complexity method is adopted to accurately obtain the LED-ROI even under the interference situation.

2) LED-ID Decoding Scheme: After extracting the ROI of the LED from the capture image, an efficient decoding scheme is adopted to obtain the associated world coordinate of the LED which is carried by the vertically-varying strip widths within the ROI. The demodulation algorithm allows different cameras with heterogeneous sampling rates to accurately decode the LED-ID with different modulation frequencies.

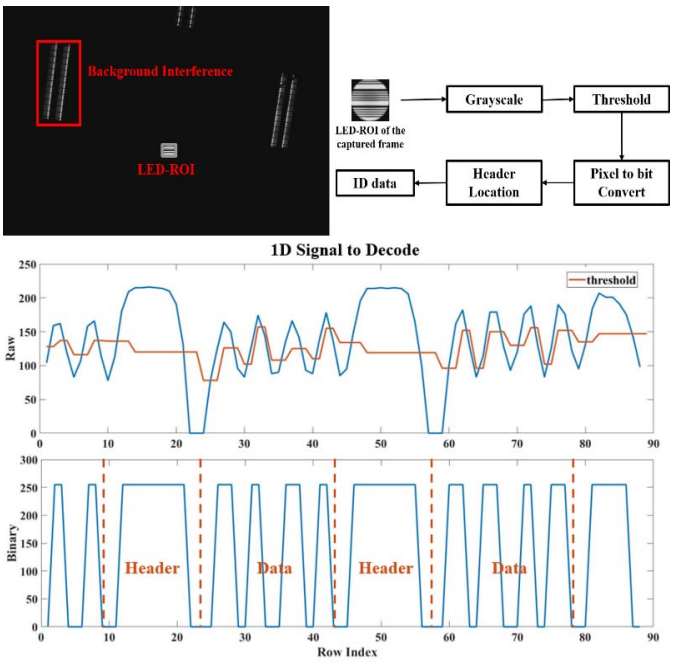


Fig. 1. Example results for ROI extraction under background light interference and VLC ID decoding.

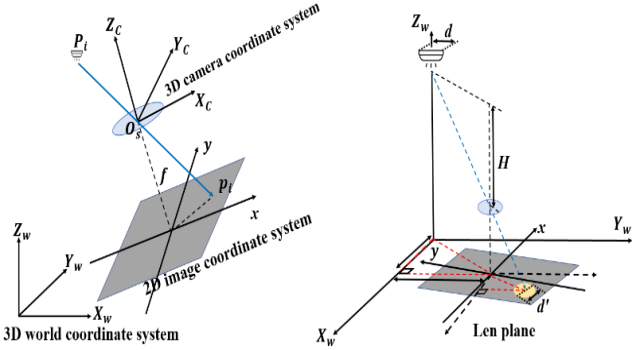


Fig. 2. The transforms among the world, camera, and image coordinate systems.

3) Single LED VLP With Odometer (SLO-VLP): The 3-D world coordinate system, 3-D camera coordinate system, and 2-D image plane coordinate system, are shown in Fig. 2. All of the LEDs $P_i = [X_i, Y_i, Z_i]^T \left(i = 1^{st}, 2^{nd}, \ldots N\right)$in the world coordinate, is mapped onto an image point $p_i = [x_i, y_i]^T \left(i = 1^{st}, 2^{nd}, \ldots N\right)$ in the image coordinate through the lens of the camera (i is the unique ID). While the p_i can be measured through the image process pipeline. Although the observation value of the imaging point is often influenced by noises [11], through the robust ROI detection method, it can be measured precisely and robustly.

We assume a fully calibrated pinhole camera with a perspective projection model, thus, for any LED $P_i$in the world coordinate, we can model the VLP system model as:

$$P_{ci} = \boldsymbol{R}(P_i - O_s) \tag{1}$$

where $P_{ci} = [X_{ci}, Y_{ci}, Z_{ci}]^T$ $(i = 1, 2, \ldots N)$ is the coordinate of the LED P_i in the 3-D camera coordinate system. $O_s = [X_s, Y_s, Z_s]^T$ is the center of the camera (positioning

state of the camera), which also refer to the coordinate of the positioning terminals. $\boldsymbol{R}$ is the rotation matrix from the 3-D world coordinate system to the 3-D camera coordinate system, as follows:

$$\boldsymbol{R} = R_x(\alpha) \bullet R_y(\beta) \bullet R_z(\gamma) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos\alpha & sin\alpha \\ 0 & -sin\alpha & cos\alpha \end{bmatrix} \bullet \begin{bmatrix} cos\beta & 0 & -sin\beta \\ 0 & 1 & 0 \\ sin\beta & 0 & cos\beta \end{bmatrix} \bullet \begin{bmatrix} cos\gamma & -sin\gamma & 0 \\ sin\gamma & cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2)$$

where α, β and γ is the angle along the X, Y, and Z axis. The three rotation angles can be estimated from the inclination sensor attached to the robot. In this article, we adopt the odometer from the mobile robot for angle calculation. The raw state definition of the odometer http://docs.ros.org/en/melodic/api/geometry_msgs/html/msg/Pose.html:

$$\mathbf{s}_0 = [{}^O_G\bar{q}^T \ {}^G P_O^T] \quad (3)$$

Here the unit quaternion[1] ${}^O_G\bar{q}^T$ represents the rotation from the global frame to the odometer frame. The vector ${}^G P_I^T$ is the position of the odometer origin in the global frame:

$${}^O_G\bar{q}^T = q_0 + q_1 i + q_2 j + q_3 k \quad (4)$$

Then, the Euler angle can be calculated through the transform form the quaternion as follows:

$$\begin{cases} \alpha = tan^{-1}\dfrac{2(q_0q_1 + q_2q_3)}{1 - 2(q_1^2 + q_2^2)} \\ \beta = -sin^{-1}(2q_0q_2 - 2q_1q_3) \\ \gamma = tan^{-1}\dfrac{2(q_0q_3 + 2q_1q_2)}{1 - 2(q_2^2 + q_3^2)} \end{cases} \quad (5)$$

Since the robot is moving horizontally on the floor, only the azimuth angle γ is needed to be considered, while the other two rotation angles can be ignored as constant matrix $\boldsymbol{C}_\alpha$ and $\boldsymbol{C}_\beta$. Then, the rotation matrix of the camera with respect to the 3-D world coordinate system can be expressed as another format:

$$\begin{aligned} \boldsymbol{R} &= \boldsymbol{C}_\alpha \cdot \boldsymbol{C}_\beta \cdot R_z(\gamma) \\ &= \begin{bmatrix} \cos(tan^{-1}\frac{2(q_0q_3+2q_1q_2)}{1-2(q_2^2+q_3^2)}) & -\sin(tan^{-1}\frac{2(q_0q_3+2q_1q_2)}{1-2(q_2^2+q_3^2)}) & 0 \\ \sin(tan^{-1}\frac{2(q_0q_3+2q_1q_2)}{1-2(q_2^2+q_3^2)}) & \cos(tan^{-1}\frac{2(q_0q_3+2q_1q_2)}{1-2(q_2^2+q_3^2)}) & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ &= \boldsymbol{C}_\alpha \cdot \boldsymbol{C}_\beta \cdot \begin{bmatrix} a & -b & 0 \\ b & a & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{aligned} \quad (6)$$

where,

$$\begin{cases} a = \cos(tan^{-1}\dfrac{2(q_0q_3 + 2q_1q_2)}{1 - 2(q_2^2 + q_3^2)}) \\ b = \sin(tan^{-1}\dfrac{2(q_0q_3 + 2q_1q_2)}{1 - 2(q_2^2 + q_3^2)}) \end{cases} \quad (7)$$

For the Z coordinate Z_s of the positioning terminal, it can be calculated by:

$$Z_s = Z_i - H \quad (8)$$

[1]https://en.wikipedia.org/wiki/Quaternion

where H is the vertical distance from center of the camera O_s to the ceiling. The LED image is no longer treated as a point as in the existing works, but as an image whose geometric features are exploited to determine Z_s. Since the rotation angles of the X and Y axis are taken to 0^o, when the robot moves horizontally, $\boldsymbol{C}_\alpha$ and $\boldsymbol{C}_\beta$ would be the identity matrix. Then vertical distance H between the LED and the lens plane can be expressed as

$$\frac{H}{f} = \frac{D}{P_d d_{pixel}} \quad (9)$$

where f is the focal length, 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. f and P_d are the intrinsic parameter of the camera. The relationship between the 3-D world coordinate of LED P_i in the 3-D camera coordinate system $P_{ci} = [X_{ci}, Y_{ci}, Z_{ci}]^T$ and the LED in the image coordinate $p_i = [x_i, y_i]^T$ can be described by:

$$\frac{X_{ci}}{x_i} = \frac{Y_{ci}}{y_i} = \frac{Z_{ci}}{-f} \quad (10)$$

Substituting (6), (8) and (10) into (1), we derive the mathematical relation between the LEDs, $[X_i, Y_i, Z_i]^T$, and the observation values of their corresponding imaging points $[x_i, y_i]^T$, then we have

$$\begin{bmatrix} x_i \\ y_i \\ \frac{D}{P_d d_{pixel}} \end{bmatrix} = \boldsymbol{C}_\alpha \cdot \boldsymbol{C}_\beta \cdot \left(-\frac{f}{H}\right) \begin{bmatrix} a & b & 0 \\ -b & a & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_i - X_s \\ Y_i - Y_s \\ Z_s - Z_i \end{bmatrix} \quad (11)$$

After obtaining the estimation coordinate of the camera's center, the 3D position and azimuth of the robot at time t can be estimated through the tf transformation[2] as follows:

$$\begin{bmatrix} x_t \\ y_t \\ z_t \\ \theta_t \end{bmatrix} = \boldsymbol{F}_{tf}\left(\begin{bmatrix} X_s \\ Y_s \\ Z_s \\ \gamma \end{bmatrix}\right) = \begin{bmatrix} r_x & 0 & 0 & 0 \\ 0 & r_y & 0 & 0 \\ 0 & 0 & r_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_s \\ Y_s \\ Z_s \\ \gamma \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \\ 0 \end{bmatrix} \quad (12)$$

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_link[3] of the robot.

B. VLP-Intertial Fusion Based on EKF

1) IMU State Prediction: Our goal is to estimate the full 3D (x, y, and z) pose of a mobile robot over time. The process can be described as a nonlinear dynamic system, with

$$\widehat{\boldsymbol{s}}_t = f(\boldsymbol{s}_{t-1}) + w_t \quad (13)$$

where $f()$ is a nonlinear state transition function of IMU state prediction, and w_t is the process noise (as the biases of inertial sensors), which is assumed to be a Gaussian distribution. $\mathbf{s}_t$ is the state of the robot's system (i.e., 3D pose and 3D orientation) at time t. Since the robots generously move horizontally, the state of the robot's system $\boldsymbol{s}_t$ is defined as a 4D pose (3D pose and 1D orientation) and the velocity along the X and Y axis:

$$\boldsymbol{s}_t = [x_t \dot{x}_t y_t \dot{y}_t \theta_t]^T \quad (14)$$

[2]https://wiki.ros.org/tf

[3]http://wiki.ros.org/navigation/Tutorials/RobotSetup/TF

The raw data of the IMU[4] in robotics is orientation quaternion, angular velocity, and linear acceleration. After the transformation from quaternion to the Euler angle (Eq. 5), the obtained value is as follows:

$$\begin{bmatrix} a_x & a_y & a_z \\ r & p & y \\ w_r & w_p & w_y \end{bmatrix} = \begin{bmatrix} a_x & a_y & 0 \\ 0 & 0 & y \\ 0 & 0 & w_y \end{bmatrix} \tag{15}$$

where the a_x, a_y,and a_z is the linear acceleration along the X, Y, and Z axes, respectively. While the w_r, w_p, and w_y is the angular velocity of the roll, pitch, and yaw, respectively. r, p, and y represents the rotation from the global frame to the IMU frame. Then, the pose of the robot can be predicted as follows:

$$\begin{cases} x_t = x_{t-1} + \dot{x}_{t-1}\Delta t + \frac{1}{2}a_{rx}\Delta t^2 \\ \dot{x}_t = \dot{x}_{t-1} + a_{rx}\Delta t \\ y_t = y_{t-1} + \dot{y}_{t-1}\Delta t + \frac{1}{2}a_{ry}\Delta t^2 \\ \dot{y}_t = \dot{y}_{t-1} + a_{ry}\Delta t \\ z_t = z_{t-1} \\ \theta_t = \theta_{t-1} + w_y \cdot \Delta t \end{cases} \tag{16}$$

While a_{rx} and a_{ry} is the component of a_x and $a_y$on the X and Y axis:

$$\begin{cases} a_{rx} = a_x cos\theta_t - a_y sin\theta_t \\ a_{ry} = a_x sin\theta_t + a_y cos\theta_t \end{cases} \tag{17}$$

It is worth mentioning that our formulation for IMU prediction can also take advantage of the control command being issued to the robot at the prediction. This means that the control is used, and it will get converted into linear acceleration and angular velocity terms, which will be used during prediction. Then, the equation (13) is changed to:

$$\hat{\boldsymbol{s}}_t = f(\boldsymbol{s}_{t-1}, \boldsymbol{u}_t) + w_t = \boldsymbol{F}'\boldsymbol{s}_{t-1} + \boldsymbol{B}\boldsymbol{u}_t + w_t \tag{18}$$

$$\boldsymbol{u}_t = [a_x \ a_y \ w_y]^{\boldsymbol{T}} \tag{19}$$

where,

$$\boldsymbol{F}' = \begin{bmatrix} 1 & \Delta t & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & \Delta t & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{20}$$

$$\boldsymbol{B} = \begin{bmatrix} 1/2cos\theta_t\Delta t^2 & -1/2sin\theta_t\Delta t^2 & 0 \\ cos\theta_t\Delta t & -sin\theta_t\Delta t & 0 \\ 1/2sin\theta_t\Delta t^2 & 1/2cos\theta_t\Delta t^2 & 0 \\ sin\theta_t\Delta t & cos\theta_t\Delta t & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \Delta t \end{bmatrix} \tag{21}$$

The state transition matrices $\boldsymbol{F}$, are defined to be the Jacobian[5] of f:

$$\boldsymbol{F} = \left.\frac{\partial f}{\partial \boldsymbol{s}}\right|(\boldsymbol{s}_{t-1}, \boldsymbol{u}_t)$$

$$= \begin{bmatrix} 1 & \Delta t & 0 & 0 & 0 & -1/2\Delta t^2(a_x sin\theta_t + a_y cos\theta_t) \\ 0 & 1 & 0 & 0 & 0 & -\Delta t(a_x cos\theta_t + a_y sin\theta_t) \\ 0 & 0 & 1 & \Delta t & 0 & 1/2\Delta t^2(a_x sin\theta_t - a_y cos\theta_t) \\ 0 & 0 & 0 & 1 & 0 & -\Delta t(a_x cos\theta_t - a_y sin\theta_t) \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{22}$$

The first stage in the EKF, shown as equations (18) and (23), is to carry out a prediction step that projects the current state estimate and error covariance forward in time:

$$\widehat{\boldsymbol{P}}_t = \boldsymbol{F}\boldsymbol{P}_{t-1}\boldsymbol{F}^{\boldsymbol{T}} + \boldsymbol{Q}_t \tag{23}$$

The estimate error covariance, $\widehat{\boldsymbol{P}}_t$, is projected via $\boldsymbol{F}$, and then perturbed by the process noise covariance $\boldsymbol{Q}_t$, which is defined as the expectation of the multiplication of process noise and its transpose:

$$\boldsymbol{Q}_t = E(w_t, w_t^T) \tag{24}$$

The process noise covariance $\boldsymbol{Q}_t$can be dynamically adjusted as follows:

$$\boldsymbol{Q}_t = \frac{1}{T}\sum_{i=0}^{T-1}(\hat{s}_{t-i} - s_{t-i})(\hat{s}_{t-i} - s_{t-i})^T \tag{25}$$

where, T is the window length ($T = 10$) of the dynamically adjustment.

2) SLO-VLP Measurement Update: The EKF employs the camera measurements of our SLO-VLP to correct its state estimate. Upon the successful SLO-VLP process, the robot receives measurements from the SLO-VLP (Eq. 12) in the form:

$$\boldsymbol{z}_t = [x_t \ y_t \ z_t \ \theta_t]^T \tag{26}$$

$$\boldsymbol{z}_t = h(\boldsymbol{s}_t) + \boldsymbol{v}_t \tag{27}$$

where $\boldsymbol{z}_t$ is the measurement at time t, h is a model that maps the state of the robot into measurement space (the VLP observation), and $\boldsymbol{v}_t$ is the normally distributed measurement noise. The standard EKF formulation specifies that observation matrix $\boldsymbol{H}$ should be a Jacobian matrix of the observation model function h. To simplify the model, we use the positioning calculation produced by SLO-VLP as the measurements, which is also in the same space as the state variables. As such, the observation matrix $\boldsymbol{H}$ is simply the identity matrix:

$$\boldsymbol{H} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{28}$$

The EKF gain can be calculated by using the SLO-VLP observation matrix $\boldsymbol{H}$ and the estimated error covariance $\widehat{\boldsymbol{P}}_t$:

$$\boldsymbol{K}_t = \widehat{\boldsymbol{P}}_t\boldsymbol{H}^{\boldsymbol{T}}[\boldsymbol{H}\widehat{P}_t\boldsymbol{H}^{\boldsymbol{T}} + \boldsymbol{R}_t]^{-1} \tag{29}$$

[4]http://docs.ros.org/en/melodic/api/sensor_msgs/html/msg/Imu.html

[5]https://en.wikipedia.org/wiki/Jacobian_matrix_and_determinant

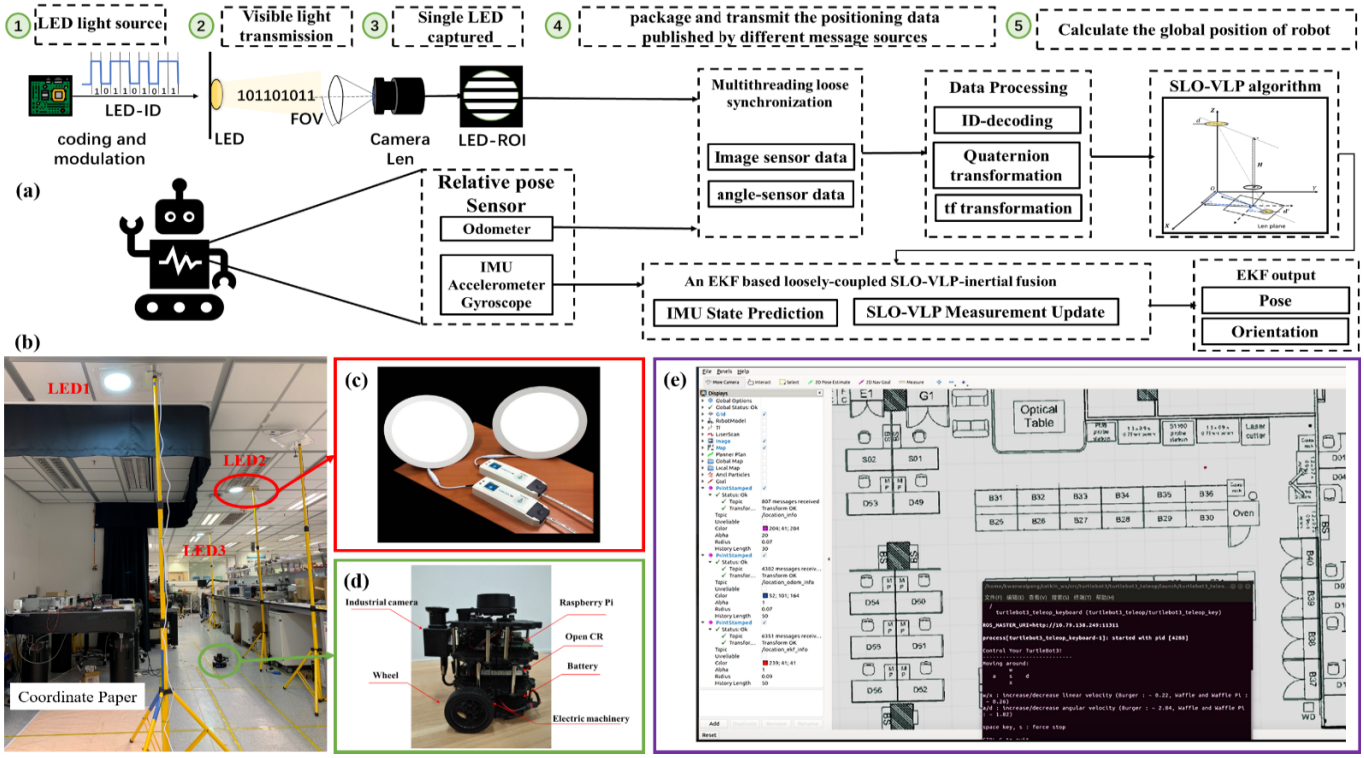


Fig. 3. Overall architecture of the proposed SLO-VLP and inertial fusion in the robotics. (a) Block diagram illustrating the full pipeline of the proposed SLO-VLP and inertial Fusion based on the EKF; (b) Experimental platform; (c) Our self-produced LED; (d) The model of the robot; (e) The indoor map of our lab.

where, $\boldsymbol{R}_t$ is measurement noise covariance of the SLO-VLP, which is defined as follows:

$$\boldsymbol{R}_t = \frac{1}{T'} \sum_{i=0}^{T'-1} v_{t-i} v_{t-i}^T \tag{30}$$

where, T' is an integer. We use the gain $\mathbf{K}_t$ to update the state vector and covariance matrix:

$$\boldsymbol{s}_t = \widehat{\boldsymbol{s}}_t + \boldsymbol{K}_t(\boldsymbol{z}_t - \boldsymbol{H}\widehat{\boldsymbol{s}}_t) \tag{31}$$

$$\boldsymbol{P}_t = (\boldsymbol{I} - \boldsymbol{K}_t\boldsymbol{H})\,\widehat{\boldsymbol{P}}_t \tag{32}$$

$\boldsymbol{I}$ is an Identity matrix. We update the calculation of the covariance of the SLO-VLP through the optimal state vector from equation (31), and T' is set to 10. Then the measurement noise/error of the SLO-VLP, including the noise from the image sensor and the odometer, can be also handled through sensor fusion during the EKF process:

$$\boldsymbol{R}_t = \frac{1}{T'} \sum_{i=0}^{T'-1} (\boldsymbol{z}_{t-i} - \boldsymbol{H}(\boldsymbol{s}_{t-i}))(\boldsymbol{z}_{t-i} - \boldsymbol{H}(\boldsymbol{s}_{t-i}))^T \tag{33}$$

IV. Evaluation

We evaluate our system through real-world experiments. We firstly evaluate the performance of the proposed SLO-VLP, and objectively show the bottleneck of just using the VLP for robot localization. After that, we verify the property of the SLO-VLP and inertial fusion to show the strong performance and the meaning of our proposed VLP-inertial fusion compared with the state-of-the-art (SOTA) methods. The video records of these two demonstrations are available at the website:

- SLO-VLP: https://www.bilibili.com/video/BV1si4y1L7HG
- VLP-inertial fusion: https://www.bilibili.com/video/BV1ZT4y1F75x

A. Experimental Settings

The framework and experimental platform of our SLO-VLP and inertial fusion scheme is shown in Fig. 3 . We set up a room-sized ($7.0 \times 3.8 \times 2.7\text{m}^3$) test field with 3 LEDs evenly mounted on the yellow pole. The experiments are performed on a Raspberry Pi 3B mobile robot (Turtlebot 3 Burger,[6] with Quad ARM Cortex-A53 Core 1.2 GHz Broadcom BCM2837 64 bit CPU and 1 GB RAM), which runs a Ubuntu Mate 16.04 OS equipped with a robot operating system[7](ROS) that is an open-source robotic framework that has been widely adopted across academia, industry, and the military around the world. 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 (our self-produced LED[8]). The LEDs' radiation surface has a circular shape of size 17.5 cm in diameter with power rating of around 18 W. The image stream is captured at around 6 Hz with a 2048×1536 resolution and recorded as ROS bags. The odometry and IMU (MPU9250[9]) is sampled at around 24 HZ

[6]https://emanual.robotis.com/docs/en/platform/turtlebot3/specifications/
[7]https://www.ros.org/
[8]http://liphy.io/hardware/
[9]https://invensense.tdk.com/products/motion-tracking/9-axis/mpu-9250/

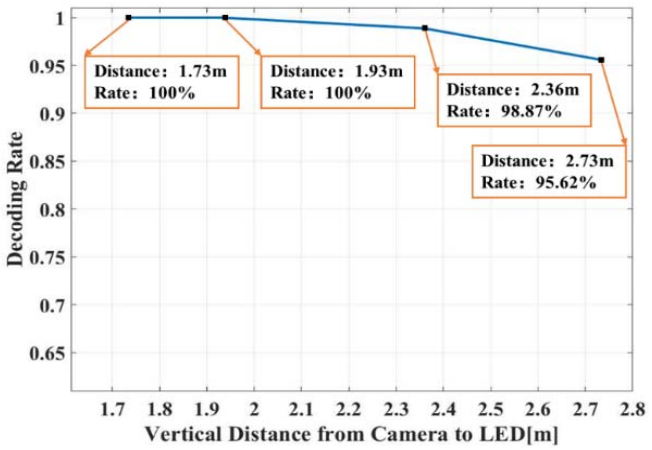


Fig. 4. The decoding performance of our SLO-VLP.

and 150 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 i7-10510U CPU @1.80 GHZ, Ubuntu 18.04) using the recorded bags from the robot. We setup the indoor grid map in our lab (Integrated Circuit Design Center, CYT-3010, HKUST). Since both the IMU and odometer provide relative motion measurements, we need pose initialization. Based on our previous work [13], we only need two LEDs to obtain millimeter-level accuracy for pose initialization. For the details about how to obtain 3D and orientation of the positioning terminal, please refer to our pervious works [9], [13]. Otherwise, the robot has to head along the global X-axes before running to obtain the initial orientation.

B. Our SLO-VLP Performance

1) VLC Decoding Performance: Since the ID decoding is the basis of the VLP. In this section, we firstly study the OCC decoding performance delivered by our hardware setup and the efficient LED-ID decoding method under various LED-to-camera distances. The decoding success rate is defined as the ratio of the number of the images with correct decoding results to the total number of images captured in a certain period. Average 1200 test data for each height, and the respective decoding rates are shown in Fig. 4.

It is clear that the decoding rate of our LED-ID decoding method, which can maintain more than 95% at a height of 2.7m, is sufficient (compare with [27] only achieves 35% decoding rate at the height of 2.5m).

2) SLO-VLP Positioning Accuracy: To evaluate the positioning accuracy of the proposed SLO-VLP system, two series of experiments were carried out. The first series was used to test the performance for motionless objects. The robot is moved in to the coordinate-paper area of Fig. 3 (b), and 325 locations were randomly chosen in the stationary localization field. At each location, there is only single VLP luminaries perceived by the camera. Then, the positioning error for each location was calculated by comparing the actual spatial position (manually observation from coordinate paper) and the estimated

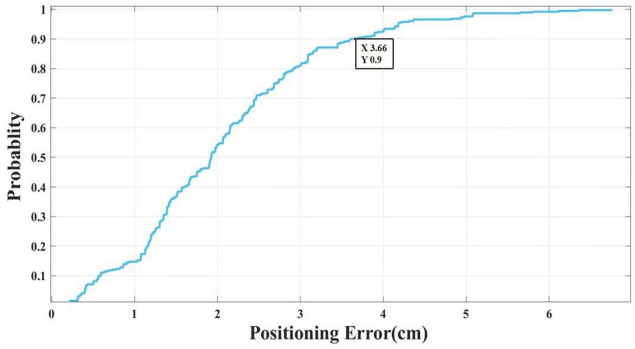


Fig. 5. The cumulative distribution function (CDF[12]) curve of positioning errors for the SLO-VLP.

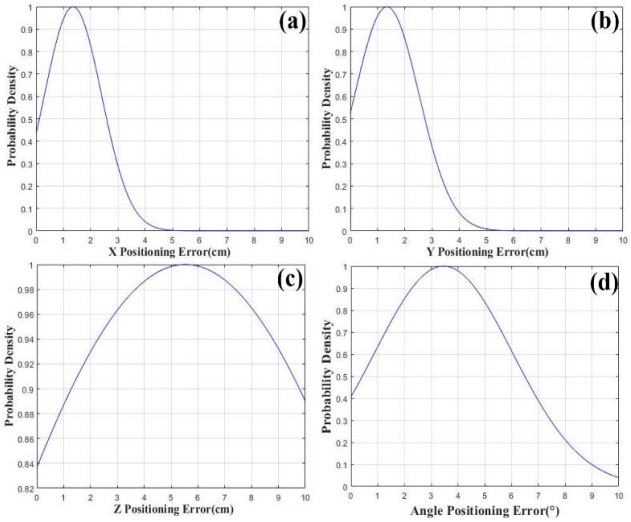


Fig. 6. The Gaussian distribution of the SLO-VLP measurement error. (a) The error distribution of x dimension;(b) The error distribution of y dimension;(c) The error distribution of z dimension;(d) The error distribution ofθ dimension.

position (from ROS calculation). The statistical result is shown is Fig. 5. More than 90% of the positioning errors are less than 3.6 cm, and the average accuracy of our SLO-VLP is around 2.1 cm, while the maximum error is 6.8 cm.

[12] We use the Gaussian distribution, which has been verified its effectiveness in the field of VLP (refer to our previous work [30]), to model the measurement noise of the SLO-VLP's output (Eq.12). Through the statistics of the mean and variance, the distribution of the measurement error (noise) for x, y, z, and θ are obtained and shown in Fig. 6. It is obvious that the measurement error of the z coordinate are more fluctuating, which is mainly caused by the change of the ROI extraction area along with the distance. Since the robot is moving horizontally, this would not influence the 2D accuracy, but we still take it into consideration for the statistical error model of the SLO-VLP.

The second series of experiments was used to test the performance for the moving mobile robot. The demonstration video of our proposed SLO-VLP is available at our website.[13]

[12] https://en.wikipedia.org/wiki/Cumulative_distribution_function

[13] Our demo of SLO-VLP is available at https://www.bilibili.com/video/BV1si4y1L7HG

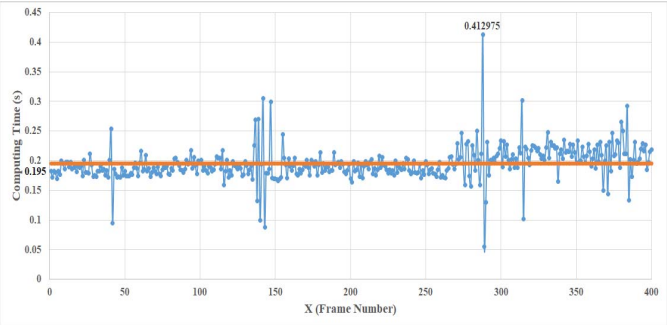


Fig. 7. The measured positioning time.

3) Real-Time Performance: Positioning speed is another key factor for localization systems. It represents the time-cost that the positioning terminal can move while 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 SLO-VLP was continuously measured 400 times to calculate the average positioning speed, as shown in Fig. 7. We run the image processing pipeline on the Raspberry Pi 3B mobile robot, and transmit to the ROS laptop for positioning calculation. The average computational time of SLO-VLP is around 195 ms.

C. The Proposed SLO-VLP and Inertial Fusion Performance

Compared with the SOTA methods (as shown in Table I), our SLO-VLP can maintain a good balance among accuracy, real-time ability, and robustness (background interference, handover). However, the performance of SLO-VLP still suffers from the shortage of decodable LEDs in reality with a chance of 5%. Furthermore, the output frequency of SLO-VLP is just around 5∼6 HZ, which is not sufficient for a fast moving robot. Although the algorithm efficiency can be improved by optimizing the image processing pipeline, it still cannot satisfy the continuity output during the movement of the robot. Finally, at least one LED is required for the SLO-VLP, which is not sufficiently robust under an LED outage situation. Therefore, in this section, we firstly build the statistical error model for the SLO-VLP, and then, the performance of the proposed VLP-inertial fusion is evaluated through the same setup are reported in the last section.

1) Statistical Error Model of the SLO-VLP: The measurement noise of the proposed SLO-VLP is modeled as white Gaussian noise (Gaussian distribution). Since the data collection process should fully cover all the possible states of each SLO-VLP observation. Then the mean value (from Fig. 6) and covariance of the measurement error of the SLO-VLP can be calculated to obtain the observation noise/error model:

$$\boldsymbol{P}\left(z_t \mid \boldsymbol{s}_t\right) = \mathcal{N}\left(\boldsymbol{H}\boldsymbol{s}_t, \boldsymbol{R}\right) = \frac{1}{\sqrt{(2\pi)^4 \left|\boldsymbol{\Sigma}\right|}} e^{-\frac{1}{2}(z_t-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(z_t-\boldsymbol{\mu})} \tag{34}$$

where, $\mathcal{N}$ is the Gaussian model, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ is the mean matrix and covariance matrix of the multivariate normal distribution, which can be calculated based on the 325 SLO-VLP observation pairs in the last section b. Based on this, we can obtain

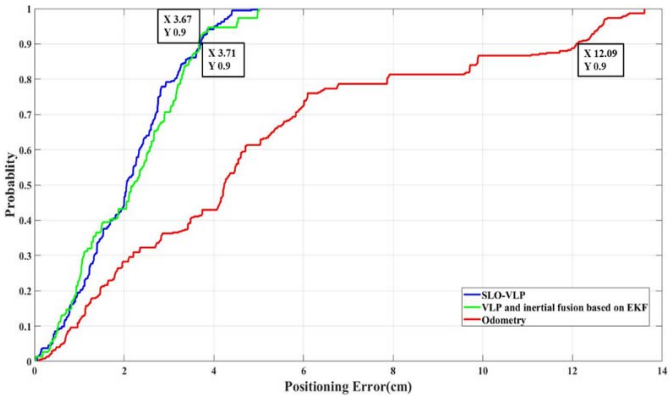


Fig. 8. CDF curve of the positioning error for the VLP-inertial fusion, SLO-VLP, and Odometry.

the initialization of the $\boldsymbol{R}_0$ for equation (33):

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_x \\ \mu_y \\ \mu_z \\ \mu_\theta \end{bmatrix} = \begin{bmatrix} 1.2383 \\ 1.3198 \\ 5.6231 \\ 2.5765 \end{bmatrix} \tag{35}$$

$$\boldsymbol{R}_0 = \Sigma = \begin{bmatrix} \sigma_{xx} & \sigma_{yx} & \sigma_{zx} & \sigma_{\theta x} \\ \sigma_{xy} & \sigma_{yy} & \sigma_{zy} & \sigma_{\theta y} \\ \sigma_{xz} & \sigma_{yz} & \sigma_{zz} & \sigma_{\theta z} \\ \sigma_{x\theta} & \sigma_{y\theta} & \sigma_{z\theta} & \sigma_{\theta\theta} \end{bmatrix} = \begin{bmatrix} 0.7223 & 0.0927 & 0.1770 & 0.1171 \\ 0.0927 & 1.1779 & -0.2214 & 0.4157 \\ 0.1770 & -0.2214 & 9.0631 & 0.5067 \\ 0.1171 & 0.4157 & 0.5067 & 3.4576 \end{bmatrix} \tag{36}$$

2) Real-Time Pose Estimation: To test the real-time localization performance, 375 locations were randomly chosen in the experimental field to evaluate the positioning accuracy of the proposed VLP-inertial fusion based on the EKF (stationary localization). Without loss of generality, we also simultaneously calculate the positioning result from the SLO-VLP and the odometry in the experiment, the result is shown in Fig. 8. The average positioning accuracy of the proposed EKF, which is 2.14 cm with the maximum error of 5.01 cm, is close to that of SLO-VLP, which is 2.04 cm with maximum error of 4.97 cm. While the average accuracy of the odometry is 4.95 cm with maximum error of 13.61 cm, while the positioning accuracy of the odometry would continue increasing due to temperature drift. The proposed EKF-based VLP-inertial fusion can achieve high accuracy when the VLP is available, while maintains robustness through the IMU when VLP is unavailable. Please note that the experiment setup is the same as the last section b, however, based on the probabilistic theory, the result of SLO-VLP might be different in different tests, but maintains a similar level. The stability of the accuracy of our proposed SKO-VLP is also evaluated through the second-time experiment.

To further assess the localization performance, we have collected data in the trial to show the positioning performance among the SLO-VLP, odometry measurement, and our proposed EKF during the movement of the mobile robot. Fig. 9 shows the estimated trajectory, which travels around 8 m at 0.22m/s (the maximum speed of our robot). Due to

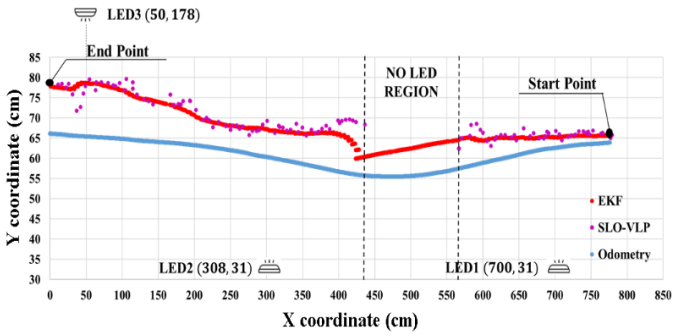


Fig. 9. The trajectory estimation of our EKF-based method compared with the trajectory from SLO-VLP and Odometry.

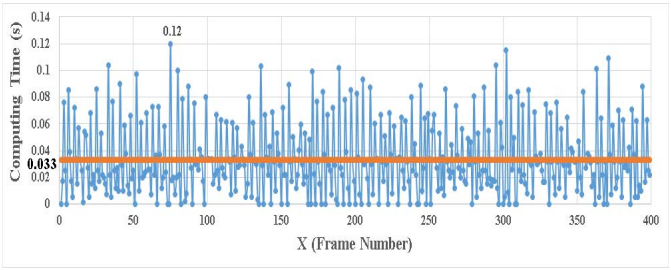


Fig. 10. The measured positioning time of the proposed VLP-inertial fusion based on the EKF.

the limitation of our hardware, we cannot provide the motion capture system to obtain the ground-truth of a moving robot. Since the average positioning accuracy of the SLO-VLP is only 2 cm, we use the trajectory of the SLO-VLP as the baseline. The EKF estimates are very close to the SLO-VLP estimation (same as the stationary localization test in Fig. 8), but we do not discuss the accuracy for the dynamic localization. What's more, even under LED outage situation, the our EKF still can provide localization, which improve the robustness of the VLP system. While the trajectory of the odometry measurement is deviated from the robot's actually position (same as the experimental observations in Fig. 8), which is caused by the cumulative error from the inertial sensor. For the proposed EKF method, the cumulative error can be corrected through the absolute information provided by VLP observation. Note that our SLO-VLP normally requires at least one LED observation in a single image. Therefore, there are no SLO-VLP results in the region of No-LED.

The runtime for our SLO-VLP and inertial fusion based on EKF is shown in Fig. 10. 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 is available, otherwise, the mainly weight of the EKF output is calculated through the inertial sensors. Please note that the inertial sensors provided data at higher rates than the camera. Even when the SLO-VLP is in-calculating (unavailable caused by calculation delay), the IMU prediction step of EKF is still used. Therefore, IMU measurement is used during the time of SLO-VLP unavailable which might be caused by the low frequency VLP output (calculation delay), lack of decodable LED, LED outage, and so on. While the cumulative error from the IMU can be corrected by the VLP observation. All of those features of the proposed VLP-inertial fusion keep well balance among accuracy, real-time, and robust performance. The video records also show the strong robustness of the proposed VLP and inertial fusion under background light interference, without any LED for VLP, and handover situation. We strongly refer our video demo.[14]

D. Discussions

The bottleneck for the application of the VLP is the real-time ability and robustness, including both limited coverage and interference. In this section, we compare the performance of our SLO-VLP, and the VLP-inertial fusion, with the SOTA works in the field of VLP in Table I. The average accuracy, computing time, and the density of the LEDs in the related experimental platform are also reported objectively.

It is obvious that the time cost (33 ms) of the proposed VLP-inertial fusion scheme achieves state-of-the-art. Note that we run the whole image process and data collection on a low-cost embedded platform Raspberry Pi 3B with a big size of captured image is 2048×1536 without any code optimization for ARM processors. After that the data (including the VLP observation and the inertial data) are transmitted to the laptop which takes charge of EKF-based pose estimation and visualizes the result in RVIZ.[15] The time cost here is the whole process covering both the image process and pose estimation (runtime in both laptop and the Raspberry), and also the time-cost of data transmission from robot to the laptop). Compare with the similar process platform with image size of 1640×1232 in [27], and 640×480 in [31], the proposed VLP-inertial based on EKF is more efficient, and hence lightweight to be used on resource-constrained computational platform. What's more, the proposed SLO-VLP and inertial fusion can also work under LED temporary outage, which greatly increases the coverage of the effective positioning area. For a large area of ($9.1 \times 4.0 \times 2.7\text{m}^3$) only 3 LEDs are required to ensure high accuracy positioning, while the other works cannot meet this high requirement. Besides, different harsh situations, such as, without LED, background light interference, occlusion, handover, are included during our experiment (details can be seen in the video demo). Therefore, the proposed SLO-VLP and inertial fusion scheme can achieve well balance among accuracy, real-time performance, and robustness.

Admittedly, in this article, we would not take the cumulative error of the odometry and the biases of the IMU into consideration. We just treat the w_t (Eq. 13) as Gaussian distribution. Base on this point, the growing yaw error of the odometry and the biases of the IMU might affect the positioning accuracy over time. In our future study, we will deeply analyze this and explore an efficient method for error calibration based on VLP observation. What's more, since the hardware limitation of our Turtlebot3 (limited maximum speed), the gyroscope bias and accelerometer bias would have little effect on the input $\mathbf{u}_t$ in Eq.18, therefore, in this article, the bias of the IMU is treated as the Gaussian distribution (w_t Eq. 13 and 18), without

[14] https://www.bilibili.com/video/BV1ZT4y1F75x

[15] http://wiki.ros.org/rviz

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.[33]	14.0	*N/A*	3	STM32F103+computer	2D	PD+IMU	7 ***-(1.5×1.5×2.5)***
R. [1]	3.9	44	2	computer	3D+yaw	Camera	9 ***-(1.8×1.8×2.0)***
R.[13]	0.8	400	2	Raspberry+computer	3D+yaw	Camera	4 ***-(1.0×1.0×1.5)***
R.[14]	17.5	*N/A*	1*	Smartphone	2D+3D	Camera	1 ***-(3.0×3.0×2.0)***
R. [8]	3.2	43	1*	computer	3D+yaw	Camera	1 ***-(0.8×0.8×2.0)***
R.[34]	2.3	60	1*	computer	2D+yaw	Camera	1 ***-(1.8×1.8×3.0)***
R.[15]	5.5	*N/A*	1	Smartphone	6D	Camera+IMU	1 ***-(2.7×1.8×1.5)***
R.[28]	13.4	*N/A*	1	Smartphone	3D+yaw	Camera+PDR	1 ***-(1.0×1.0×2.6)***
R.[27]	5.0	50	1	Raspberry+computer	3D+yaw	Camera+IMU	23 ***-(5.0×4.0×2.3)***
R.[31]	3.0	100	1	Raspberry+computer	2D+yaw	Camera+IMU	4 ***-(2.0×2.0×1.5)***
SLOVLP	2.0	195	1	Raspberry+computer	3D+yaw	Camera+odom	3 ***-(7.0×3.8×2.7)***
Our ekf	2.1	33	1 or 0	Raspberry+computer	3D+yaw	SLOVLP+IMU	3 ***-(9.1×4.0×2.7)***

"1*" means that the LED lamp with beacon (refer to Ref. [8], [14], [34])

estimating the biases. However, when it comes to general situation, the biases of the IMU are not negligible, which are needed to be estimated during the EKF. The IMU state can be added as the state of the robot (refer to [32]). These are modeled as random walk processes driven by zero-mean white Gaussian noise vectors. We credit our contribution mainly to the robust SLO-VLP scheme based on inertial sensor fusion, which derives smoothed temporally pose estimates and can be sustained in different harsh situations. Moreover, the delivered accuracy is good enough for many indoor robot applications.

V. CONCLUSION

In this article, we propose a loosely-coupled VLP-inertial fusion method for robotics localization using RSE-based camera with single or even without LED. The SLO-VLP observations lead to an estimate of the robot position, which can be fused with the IMU estimation to get the optimal update of the robot's actual position. While the IMU estimation can be used to relax the requirement on the minimum number of concurrently observable LEDs for positioning to zero. The proposed scheme is evaluated on the robotic platform under the LED shortage/outage and background light interference. The result shows that the our method has strong robustness, which can keep well balance among accuracy, real-time ability and coverage. For our future study, we will deeply analyze the cumulative error of the inertial sensor and explore more flexible and efficient method for error calibration based on VLP observation.

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] M. Maheepala, A. Z. Kouzani, and M. A. Joordens, "Light-based indoor positioning systems: A review," *IEEE Sensors J.*, vol. 20, no. 8, pp. 3971–3995, Apr. 2020.

[3] N. Huang, C. Gong, J. Luo, and Z. Xu, "Design and demonstration of robust visible light positioning based on received signal strength," *J. Lightw. Technol.*, vol. 38, no. 20, pp. 5695–5707, Oct. 15, 2020.

[4] W. Guan *et al.*, "A novel three-dimensional indoor positioning algorithm design based on visible light communication," *Opt. Commun.*, vol. 392, pp. 282–293, Jun. 2017.

[5] W. Guan *et al.*, "High-precision approach to localization scheme of visible light communication based on artificial neural networks and modified genetic algorithms," *Opt. Eng.*, vol. 56, no. 10, 2017, Art. no. 106103.

[6] 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.

[7] 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.

[8] 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. Autom., Electron. Electr. Eng. (AUTEEE)*, Nov. 2018, pp. 202–208.

[9] 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.

[10] 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.

[11] S. Chen and W. Guan, "High accuracy VLP based on image sensor using error calibration method," 2020, *arXiv:2010.00529*. [Online]. Available: http://arxiv.org/abs/2010.00529

[12] W. Guan, Z. Liu, S. Wen, H. Xie, and X. Zhang, "Visible light dynamic positioning method using improved camshift-Kalman algorithm," *IEEE Photon. J.*, vol. 11, no. 6, pp. 1–22, Dec. 2019.

[13] 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.

[14] 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.

[15] 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, Sep. 1, 2020.

[16] Y. Xiao, W. Guan, S. Wen, J. Li, Z. Li, and M. Liu, "The optical bar code detection method based on optical camera communication using discrete Fourier transform," *IEEE Access*, vol. 8, pp. 123238–123252, 2020.

[17] C. Xie, W. Guan, Y. Wu, L. Fang, and Y. Cai, "The LED-ID detection and recognition method based on visible light positioning using proximity method," *IEEE Photon. J.*, vol. 10, no. 2, pp. 1–16, Apr. 2018.

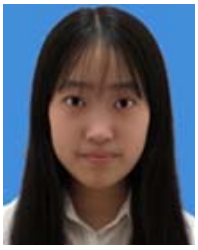
[18] K. Yu, J. He, and Z. Huang, "Decoding scheme based on CNN for mobile optical camera communication," *Appl. Opt.*, vol. 59, no. 23, pp. 7109–7113, 2020.

[19] R. Qv, L. Feng, A. Yang, P. Guo, B. Lin, and H. Huang, "A high efficient code for visible light positioning system based on image sensor," *IEEE 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.

[20] W. Guan, Y. Wu, C. Xie, L. Fang, X. Liu, and Y. Chen, "Performance analysis and enhancement for visible light communication using CMOS sensors," *Opt. Commun.*, vol. 410, pp. 531–551, Mar. 2018.

[21] B. Hussain, C. Qiu, and C. P. Yue, "A universal VLC modulator for retrofitting LED lighting and signage," in *Proc. IEEE 8th Global Conf. Consum. Electron. (GCCE)*, Oct. 2019, pp. 1008–1009.
[22] B. Hussain, C. Qiu, and C. P. Yue, "Smart lighting control and services using visible light communication and Bluetooth," in *Proc. IEEE 8th Global Conf. Consum. Electron. (GCCE)*, Oct. 2019, pp. 1–2.
[23] 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.
[24] H.-Y. Lee *et al.*, "Rollinglight: Enabling line-of-sight light-to-camera communications," in *Proc. 13th Annu. Int. Conf. Mobile Syst., Appl., Services*, 2015, pp. 167–180.
[25] 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.
[26] R. Zhang, W.-D. Zhong, K. Qian, and D. Wu, "Image sensor based visible light positioning system with improved positioning algorithm," *IEEE Access*, vol. 5, pp. 6087–6094, 2017. [Online]. Available: https://ieeexplore.ieee.org/document/7900356, doi: 10.1109/ACCESS.2017.2693299.
[27] 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.
[28] 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.
[29] Y. Ji *et al.*, "A single LED lamp positioning system based on CMOS camera and visible light communication," *Opt. Commun.*, vol. 443, pp. 48–54, Jul. 2019.
[30] M. Liu *et al.*, "Towards indoor localization using visible light communication for consumer electronic devices," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Sep. 2014, pp. 143–148.
[31] 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.
[32] M. Li and A. I. Mourikis, "Online temporal calibration for camera–IMU systems: Theory and algorithms," *Int. J. Robot. Res.*, vol. 33, no. 7, pp. 947–964, 2014.
[33] 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.
[34] 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. [Online]. Available: https://ieeexplore.ieee.org/document/9233945, doi: 10.1109/JPHOT.2020.3032448.

Weipeng Guan (Student Member, IEEE) received the B.E. and M.E. degrees from the South China University of Technology in 2016 and 2019, respectively. He is currently pursuing the Ph.D. degree with The Hong Kong University of Science and Technology. He has worked 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 published more than 40 research papers in prestigious international journals and conferences. His research interests include mobile robot, optical camera communications, visible light positioning, and so on.



Linyi Huang is currently pursuing the bachelor's degree with the South China University of Technology, majoring in automation science and engineering. Her research interests include VLC-based image communication and ROS-based robot localization.

Babar Hussain received the bachelor's degree in electronic engineering from the University of Engineering and Technology at Taxila, Pakistan. He is currently pursuing the Ph.D. degree in ECE with HKUST. His research interest includes visible light communication system design.

C. Patrick Yue (Fellow, IEEE) received the B.S. (Hons.) degree from the University of Texas at Austin, TX, USA, in 1992, and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, CA, USA, in 1994 and 1998, respectively.

Based on his Ph.D. work, he co-founded Atheros Communications in 1998. He applied his expertise in CMOS RF transistor and passive component modeling to enable the deployment of the world's first 802.11a 5-GHz CMOS RF transceiver. After four years at Atheros, he joined another Silicon Valley startup Aeluros to develop 10-Gbps CMOS serial link IC products for the optical-to-electrical module. From 2001 to 2003, he was a Consulting Assistant Professor with the EE Department at Stanford. In 2003, he joined Carnegie Mellon University, Pittsburgh, PA, as an Assistant Professor with the Department of Electrical and Computer Engineering. In 2006, he moved back to CA to join the University of California at Santa Barbara and was promoted to a Professor in 2010. Since 2011, he has been a Professor with the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology. From 2014 to 2015, he has served as the Associate Provost for Knowledge Transfer. He is currently the Director of the HKUST Integrated Circuits Design Center and the HKUST-Qualcomm Joint Innovation and Research Laboratory. He has contributed to more than 180 peer-reviewed articles and two book chapters and holds 17 U.S. patents. His current research interests include wireline communication circuits, millimeter-wave system-on-chip design, and visible and laser light communication systems and applications.

Dr. Yue is a Fellow of OSA. He was awarded the 11th Guanghua Engineering Science and Technology Youth Award by the Chinese Academy of Engineering (CAE) in 2016. He was a co-recipient of the Best Student Paper Award at the IEEE International Solid-State Circuits Conference (ISSCC) in 2003 and the IEEE International Wireless Symposium (IWS) in 2016, and the IEEE Circuits and Systems Society Outstanding Young Author Award in 2017. He has served on the technical program committees of the ISSCC, the IEEE Symposium on VLSI Circuits (VLSI-Circuits), the IEEE RFIC Symposium (RFIC), the IEEE European Solid-State Circuits Conference (ESSCIRC), the IEEE IWS, and the IEEE Asian Solid-State Circuits Conference (A-SSCC). He was an IEEE Solid-State Circuit Society (SSCS) Distinguished Lecturer in 2017 and an Elected IEEE SSCS AdCom Member from 2015 to 2017. He has been serving as the Vice-President of Membership of the IEEE SSCS. He has served as an Editor for IEEE ELECTRON DEVICE LETTERS and *IEEE Solid-State Circuits Magazine* and a Guest Editor for the IEEE TRANSACTIONS ON MICROWAVE THEORY AND TECHNIQUES. He has been an Editor of the PROCEEDINGS OF THE IEEE since 2018.