Computer Vision-based Localization with Visible Light Communications

Lin Bai, Yang Yang, *Member, IEEE*, Mingzhe Chen, *Member, IEEE*, Chunyan Feng, *Senior Member, IEEE*, Caili Guo, *Senior Member, IEEE*, Walid Saad, *Fellow, IEEE*, and Shuguang Cui, *Fellow, IEEE*

***Abstract*—Visible light positioning and computer vision-based localization have the potential to be cost-effective technologies for accurate indoor localization. However, the feasibility of existing methods in this domain is limited. In this paper, a novel visible light communication (VLC)-assisted perspective-four-line algorithm (V-P4L) is proposed for practical indoor localization. The basic idea of V-P4L is to jointly use VLC and computer vision techniques to achieve high localization accuracy regardless of LED height differences. In particular, the space-domain information is first exploited to estimate the orientation and coordinate information of a single rectangular LED luminaire in the camera coordinate system based on plane geometry theory and solid geometry theory. Then, by using time-domain information transmitted by VLC and the estimated luminaire information, the proposed V-P4L can estimate the position and pose of the camera using single-view geometry theory and the linear least square (LLS) method. To further mitigate the effect of height differences among LEDs on localization accuracy, a correction algorithm based on the LLS method and a simple optimization method is proposed. Due to the combination of time- and space-domain information, V-P4L can achieve accurate localization using a single luminaire without limitation on the correspondences between the features and their projections in conventional perspective-n-line (PnL) algorithms. Simulation results show that the position error caused by the proposed V-P4L algorithm is always less than $15\,\mathrm{cm}$ and the orientation error is always less than 4° using popular indoor luminaires. Experimental results with real hardware show that the average position error is less than 3 cm under both similar and different heights for the LEDs.**

***Index Terms*—Camera, computer vision based localization, perspective-n-line (PnL), visible light positioning.**

I. Introduction

ACCURATE indoor localization is increasingly important due to the surge of position-dependent services such as position tracking, navigation, virtual reality, and robot movement control [2]. In this area, visible light positioning (VLP) techniques and computer vision-based localization have the advantage of high accuracy and low cost. VLP solutions use visible light signals to determine the position of a receiver [3]–[5]. Due to the strong directionality and low multipath interference of visible light, using visible light for localization can achieve high accuracy, e.g., see [3], [5] and [6]. Meanwhile, since VLP uses light-emitting doides (LEDs) as transmitters [7], VLP has relatively low cost on infrastructure (e.g., see [3] and [8]). Computer vision-based localization is another localization technique that uses the images of the reference features captured by cameras to estimate the position and pose of cameras [9] and [10]. However, VLP and computer vision based localization also have their own challenges. In particular, VLP cannot estimate the pose and position of the receiver using a single beacon, while computer vision based localization cannot be used in the dark. To further improve the performance of VLP and computer vision based localization, one can consider novel approaches that combine both technologies.

A. Related Works

Typical VLP algorithms include proximity [11], fingerprinting [12], time of arrival (TOA) [13], angle of arrival (AOA) [14], received signal strength (RSS) [15]–[17], and image sensing [18]. In these existing algorithms, proximity and fingerprinting cannot estimate the receiver pose. Meanwhile, the accuracy of proximity is low [15] while fingerprinting requires at least three luminaires to overcome ambiguity challenges [19]. AOA algorithms use the estimated angles of arrival signals from multiple luminaires to determine the receiver position [14], [20] and [21]. Therefore, AOA algorithms are sensitive to the estimated angles, which can significantly affect the accuracy. TOA algorithms use the transmission time of the signals transmitted from multiple luminaires to estimate the receiver position [13], [22] and [23]. In this case, TOA algorithms require strict synchronizations between the transmitters and the receiver [24]. Image sensing algorithms use the geometric relation between the transmitters and the receiver to estimate the receiver position, which require multiple luminaires [18], [24] and [25]. Among these VLP algorithms, RSS algorithms are most widely-used due to their high accuracy and low cost [24]. However, RSS algorithms also require multiple luminaires for localization, e.g., see [17], [26], and [27]. Moreover, RSS algorithms require accurate channel

This work was supported by NSF CNS-1909372.

L. Bai, Y. Yang and C. Feng are with the Beijing Key Laboratory of Network System Architecture and Convergence, School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China (e-mail: bailin2126@bupt.edu.cn; young0607@bupt.edu.cn; cyfeng@bupt.edu.cn).

M. Chen is with the Department of Electrical Engineering, Princeton University, Princeton, NJ, 08544, USA (e-mail: mingzhec@princeton.edu).

C. Guo is with Beijing Laboratory of Advanced Information Networks, School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China (e-mail: guocaili@bupt.edu.cn).

W. Saad is with the Wireless@VT, Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA, 24060, USA, (email: walids@vt.edu).

S. Cui is currently with the School of Science and Engineering, Shenzhen Research Institute of Big Data and Future Network of Intelligence Institute (FNii), the Chinese University of Hong Kong, Shenzhen, China, 518172 (e-mail: shuguangcui@cuhk.edu.cn).

A preliminary version of this work [1] appears in the proceeding of IEEE GLOBECOM.

models, which is challenging to be achieved in practice. A popular assumption in RSS algorithms is that the radiation pattern of LEDs follows a Lambertian model which may not be applicable for many luminaires especially when a lampshade is used [28]. Meanwhile, the estimated channel gain is affected by sunlight, dust, and shadowing in practice [29]. Therefore, the feasibility of RSS algorithms is limited.

Meanwhile, a number of computer vision-based localization techniques exist. In particular, the perspective-n-line (PnL) and perspective-n-point (PnP) algorithms are performed by analyzing n correspondences between three dimensional (3D) reference features and their two dimensional (2D) projections on the image (i.e., 3D-2D correspondences), where the features are either points or lines [30]. PnP methods use point features, while PnL methods employ line features. Compared with point features, line features carry richer information [30]. Therefore, compared with PnP methods, PnL methods can achieve higher detection accuracy and are more robust to occlusions [31]. PnL methods can estimate the pose and position of the camera using the iterative methods or non-iterative methods [30]. Both iterative and non-iterative PnL methods need 3D-2D correspondences given in advance, which is difficult to obtain [32], [33]. To circumvent this challenge, the work in [33] proposed a method to match the 3D-2D correspondences for a scenario in which the number of the vertical lines are more than that of horizontal lines. However, this method cannot be applied to a scenario in which there is no significant difference between the numbers of horizontal and vertical lines, such as the scenario in which the rectangular beacons are deployed on the ceiling. Therefore, the feasibility of the method in [33] is constrained.

In summary, VLP algorithms require multiple luminaires and accurate channel models, while computer vision-based localization algorithms have limitation on 3D-2D correspondences. Therefore, both VLP and computer vision-based localization algorithms have limitations on the feasibility.

B. Contributions

The main contribution of this paper is a novel visible light communication (VLC) assisted perspective-four-line algorithm (V-P4L), which can achieve practical and accurate indoor localization. *To the authors' best knowledge, this is the first localization algorithm that can achieve accurate position and pose estimation without the limitation on 3D-2D correspondence using a single luminaire*. The key contributions of this paper include:

- We propose an indoor localization algorithm, dubbed V-P4L, which uses a camera to simultaneously capture the information in both the time and space domains[1] of luminaires so as to achieve high practicality and high accuracy. Based on the plane geometry theory and solid geometry theory, V-P4L first estimates the luminaires' information in the camera coordinate system using the space-domain information captured by the camera. Then, based on the single-view geometry theory and the time-domain information of luminaires, V-P4L can estimate the pose and position of the camera exploiting the luminaires' information in different coordinate systems. In this way, V-P4L can accurately estimate the position and pose of the receiver using a single luminaire or a luminaire array.
- To overcome the limitation on the 3D-2D correspondences in computer vision-based localization algorithms, the time-domain information transmitted by VLC is exploited in V-P4L to properly match 3D-2D correspondences. In particular, based on the time-domain information, the camera can obtain the coordinates information of LEDs in the world coordinate system. Then, based on the perspective projection theory, the 3D-2D correspondences can be properly matched using the information of LEDs in different coordinate systems. As such, V-P4L can achieve higher practicality compared to that of the existing algorithms in [30], [32], [33].
- To further improve the practicality of V-P4L, we propose an algorithm that extends V-P4L to correct height differences among LEDs based on the information in both time and space domains. When the heights of LEDs are different, based on the single-view geometry theory and the LLS method, the proposed correction algorithm can first estimate the pose and 2D position of the camera. Then, based on the single-view geometry theory and a simple optimization method, the proposed correction algorithm can estimate the 3D position of the camera. This, in turn, allows the use of V-P4L regardless of the height differences among LEDs.

Simulation results show that the position error caused by the proposed V-P4L algorithm is always less than $15\,\mathrm{cm}$ and the orientation error is always less than 4° using popular indoor luminaires. Additionally, the proposed algorithm can be used regardless of the height difference among LEDs. When LEDs have the same height, the accuracy of V-P4L is 5 cm. When LEDs have different heights, the accuracy of V-P4L is less than 17 cm. Moreover, V-P4L is less sensitive to image noise than conventional VLP and computer vision-based localization algorithms. Experimental results with real hardware show that the average position error is less than 3 cm under both similar and different heights for the LEDs.

The rest of the paper is organized as follows. Section II introduces the system model. Section III calculates the luminaire information in the camera coordinate system. The proposed V-P4L algorithm is presented in Section IV, and the proposed correction algorithm is detailed in Section V. Simulation and experimental results are presented in Section VI. Conclusions are drawn in Section VII.

The following notations are used throughout the paper: A and a with or without subscript denote scalars; $\boldsymbol{v}$ denotes a column vector and $\boldsymbol{A}$ stands for a matrix; $|A|$ denotes the absolute value of A; $\boldsymbol{A}^{\mathrm{T}}$, $\boldsymbol{A}^{-1}$, $\det(\boldsymbol{A})$ and $|\boldsymbol{A}|$ indicate the transpose, inverse, determinant and norm of $\boldsymbol{A}$, respectively; $\boldsymbol{v} \times \boldsymbol{u}$ denotes the cross product of $\boldsymbol{v}$ and $\boldsymbol{u}$; $\boldsymbol{v} \cdot \boldsymbol{u}$, $\boldsymbol{A} \cdot \boldsymbol{v}$ and $\boldsymbol{A}\boldsymbol{B}$ denotes the dot products of $\boldsymbol{v}$ and $\boldsymbol{u}$, $\boldsymbol{A}$ and $\boldsymbol{v}$, and $\boldsymbol{A}$ and $\boldsymbol{B}$, respectively; $\hat{A}$, $\hat{\boldsymbol{A}}$ and $\hat{\boldsymbol{v}}$ represent the estimate of A,

[1]The time domain information refers to the information transmitted in a given time slot. This type of information includes the coordinates of the luminaire's vertices transmitted through VLC. The space domain information means the geometric information captured by the camera such as the shape of the luminaire.

TABLE I
SUMMARY OF OUR MAIN NOTATIONS.

Symbol	Meaning	Symbol	Meaning
$(u_0, v_0)^{\mathrm{T}}$	Pixel coordinate of O^{i}	$\phi_{l_{ij}}$	Rotation angle from Y^{i}-axis to l_{ij} in the counter clockwise direction
f	Focal length	$\rho_{l_{ij}}$	Distance from O^{i} to l_{ij}
f_u, f_v	Focal ratios	$\boldsymbol{n}^{\mathrm{w}}_{\mathrm{LED}}/\boldsymbol{n}^{\mathrm{c}}_{\mathrm{LED}}$	Normal vector of the luminaire in WCS/CCS
d_x, d_y	Physical size of each pixel	Π_{ij}	Lateral face determined by the vertices P_i, P_j and O^{c}
P_i	The ith vertex of the luminaire	$\boldsymbol{n}^{\mathrm{c}}_{\Pi_{ij}}$	Normal vector of Π_{ij} in CCS
p_i	Projection of P_i on the image plane	S	Area of the luminaire
$p^{\mathrm{p}}_i/p^{\mathrm{i}}_i$	Pixel/Image coordinate of p_i	H	Distance from O^{c} to the luminaire
$P^{\mathrm{c}}_i/P^{\mathrm{w}}_i$	Camera/World coordinate of P_i	V	Volume of rectangular pyramid $O^{\mathrm{c}}-P_1P_2P_3P_4$
L_{ij}	3D reference line connecting P_i and P_j	V_i	Volume of triangular pyramid $O^{\mathrm{c}}-P_iP_jP_k$
l_{ij}	2D projection of L_{ij} on the image plane	φ, θ, ψ	Euler angles corresponding to the $X^{\mathrm{c}}-$axis, $Y^{\mathrm{c}}-$axis and $Z^{\mathrm{c}}-$axis
$\boldsymbol{n}^{\mathrm{c}}_{L_{ij}}$	Direction vector of L_{ij} in CCS	$\boldsymbol{R}^{\mathrm{w}}_{\mathrm{c}}/\boldsymbol{t}^{\mathrm{w}}_{\mathrm{c}}$	Pose/Position of the camera in WCS

$\boldsymbol{A}$ and $\boldsymbol{v}$, respectively. In addition, there are some special or important symbols used throughout in this paper, which are listed in Table I with their meaning.

II. System Model

A diagram of our considered system is illustrated in Fig. 1. Figure 1(a) shows a typical indoor scenario, where LED lamps in a parallelogram shape uniformly mounted on the ceiling. In the considered scenario, the LED luminaires are transmitter, while the receiver can be a smart phone or other smart terminal with a camera on it. V-P4L can be implemented based on a single parallelogram luminaire which is also a LED array. We will introduce V-P4L using a single rectangular LED luminaire[2] hereinafter. Figure 1(b) shows the geometric relation between the rectangular LED luminaire and the camera. The luminaire's shape and the world coordinates of the luminaires' vertices are known in advance, while the position and pose of the receiver should be estimated.

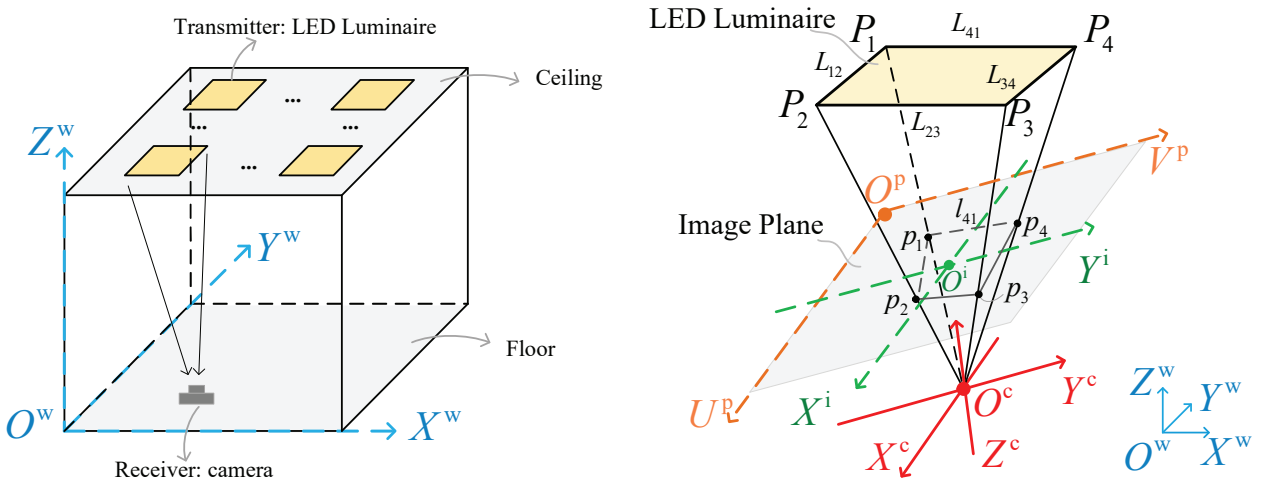
In V-P4L, four coordinate systems are used for localization: the 2D pixel coordinate system (PCS) $(O^{\mathrm{p}}, U^{\mathrm{p}}, V^{\mathrm{p}})$ on the image plane[3], the 2D image coordinate system (ICS) $\left(O^{\mathrm{i}}, X^{\mathrm{i}}, Y^{\mathrm{i}}\right)$ on the image plane, the 3D camera coordinate system (CCS) $(O^{\mathrm{c}}, X^{\mathrm{c}}, Y^{\mathrm{c}}, Z^{\mathrm{c}})$, and the 3D world coordinate system (WCS) $(O^{\mathrm{w}}, X^{\mathrm{w}}, Y^{\mathrm{w}}, Z^{\mathrm{w}})$. In PCS, ICS, and CCS, the axes U^{p}, X^{i}, and X^{c} are parallel to each other and, similarly, V^{p}, Y^{i}, and Y^{c} are also parallel to each other. Additionally, O^{p} is at the upper left corner of the image plane and O^{i} is at the center of the image plane. Moreover, O^{i} is called the principal point, whose pixel coordinate is $(u_0, v_0)^{\mathrm{T}}$. In contrast, O^{c} is called the camera optical center. Furthermore, O^{i} and O^{c} are on the optical axis. The distance between O^{c} and O^{i} is the focal length f, and thus the z-coordinate of the image plane in CCS is $z^{\mathrm{c}} = f$. The pixel coordinate of the principal point $(u_0, v_0)^{\mathrm{T}}$ and the focal length f can be known in advance by the camera calibration [15], [35].

The rectangular LED luminaire is constructed by four vertices P_i ($i \in \{1,2,3,4\}$). The four 3D reference lines L_{ij} ($i, j \in \{1,2,3,4\}, i \neq j$) are the edges of the luminaire. In other words, L_{ij} is the line segment whose endpoints are points P_i and P_j ($i, j \in \{1,2,3,4\}, i \neq j$). In addition, $\boldsymbol{P}^{\mathrm{w}}_i = (x^{\mathrm{w}}_i, y^{\mathrm{w}}_i, z^{\mathrm{w}}_i)^{\mathrm{T}}$ is the world coordinate of vertex i of the luminaire. This coordinate is assumed to be known at the transmitter and can be transmitted by VLC as time-domain information [6], [8]. In the rectangular luminaire, only a single LED at a vertex transmits the coordinates and order of the four vertices of the luminaire. Without loss of generality, we assume that the LED at P_1 transmits information. Moreover, the unit normal vector of the luminaire in WCS, $\boldsymbol{n}^{\mathrm{w}}_{\mathrm{LED}}$, can be calculated by the world coordinates of the luminaire's vertices. In other words, the orientation of the luminaire in WCS is also known in advance.

On the other hand, the receiver is a calibrated standard pinhole camera which is not coplanar with the luminaire. The pose and position of the receiver need to be estimated. As shown in Fig. 1(b), the transmitter and the receiver produce a rectangular pyramid $O^{\mathrm{c}} - P_1P_2P_3P_4$ where O^{c} is called the apex, the rectangle $P_1P_2P_3P_4$ is called the base face and P_i is called the ith vertex. In the rectangular pyramid $O^{\mathrm{c}} - P_1P_2P_3P_4$, we define Π_{ij} as the lateral face determined by the vertices P_i, P_j ($i, j \in \{1,2,3,4\}, i \neq j$) and O^{c}. In the camera, point p_i is the projection of point P_i on the image plane. Moreover, line l_{ij} is the projection on the image plane of the 3D reference line L_{ij}. The point p_i and the line l_{ij} can be easily obtained by the camera. In Fig. 1(b), the image plane is tilted to show that camera is not necessarily parallel to the rectangle $P_1P_2P_3P_4$. Note

[2]The parallelogram luminaire consists of multiple LEDs. In this luminaire, only a single LED at a vertex transmits information. Therefore, compared with existing VLP algorithms, V-P4L can reduce the cost of implementation and the interference between LEDs.

[3]As shown in Fig. 1, the image plane is a virtual plane. In this work, the camera is a standard pinhole camera. The actual image plane is behind the camera optical center (i.e., the pinhole), o^{c}. To show the geometric relations more clearly, the virtual image plane is set up in front of o^{c} as done in prior works (e.g., [34]). In particular, the virtual image plane and the actual image plane are centrally symmetric, and o^{c} is the center of symmetry.



(a) A typical indoor localizaion scenario of the proposed algorithm. (b) The geometric relation between the LED luminaire and the camera.

Fig. 1. Illustrative system diagram of the proposed algorithm.

that several existing PnL algorithms assume that the 3D-2D correspondences $(L_{ij} \iff l_{ij})$ are known in advance, which is impractical [33]. In contrast, here, to estimate the pose and position of the camera without the 3D-2D correspondences, the camera is used to simultaneously capture the time- and space-domain information. The pixel coordinate of p_i is given by $\boldsymbol{p}_i^{\mathrm{p}} = \left(u_i^{\mathrm{p}}, v_i^{\mathrm{p}}\right)^{\mathrm{T}}$ and obtained by the camera through image processing [18]. Based on the single-view geometry theory, the image coordinate of p_i, $\boldsymbol{p}_i^{\mathrm{i}} = \left(x_i^{\mathrm{i}}, y_i^{\mathrm{i}}\right)^{\mathrm{T}}$, can be obtained as

$$\boldsymbol{p}_i^{\mathrm{i}} = \begin{bmatrix} d_x & 0 \\ 0 & d_y \end{bmatrix} \cdot \boldsymbol{p}_i^{\mathrm{p}} - \begin{bmatrix} u_0 d_x \\ v_0 d_y \end{bmatrix}, \tag{1}$$

where d_x and d_y are, respectively, the physical sizes of each pixel in the x and y directions on the image plane. The camera's intrinsic parameters, including $(u_0, v_0)^{\mathrm{T}}$ and the focal ratio $f_u = \frac{f}{d_x}$ and $f_v = \frac{f}{d_y}$, can be known in advance by using camera calibration [15], [35].

Based on the single-view geometry theory, the transformation from CCS to WCS can be expressed as [30]

$$\boldsymbol{P}^{\mathrm{w}} = \boldsymbol{R}_{\mathrm{c}}^{\mathrm{w}} \cdot \boldsymbol{P}^{\mathrm{c}} + \boldsymbol{t}_{\mathrm{c}}^{\mathrm{w}}, \tag{2}$$

where $\boldsymbol{P}^{\mathrm{w}}$ and $\boldsymbol{P}^{\mathrm{c}}$ are the world and camera coordinates of the same object, respectively. In addition, $\boldsymbol{R}_{\mathrm{c}}^{\mathrm{w}} \in \mathbb{R}^{3\times3}$ and $\boldsymbol{t}_{\mathrm{c}}^{\mathrm{w}} \in \mathbb{R}^{3\times1}$ denote the pose and the position of the camera in WCS, respectively[4]. The task of localization essentially consists of finding $\boldsymbol{R}_{\mathrm{c}}^{\mathrm{w}}$ and $\boldsymbol{t}_{\mathrm{c}}^{\mathrm{w}}$. In particular, $\boldsymbol{R}_{\mathrm{c}}^{\mathrm{w}}$ denotes the pose of the camera which contains the rotation angles of WCS corresponding to the X^{c}-axis, Y^{c}-axis, and Z^{c}-axis, while $\boldsymbol{t}_{\mathrm{c}}^{\mathrm{w}}$ denotes the position of the camera which contains x-coordinate, y-coordinate, and z-coordinate of the camera in WCS. As introduced in Section I, accurate $\boldsymbol{R}_{\mathrm{c}}^{\mathrm{w}}$ and $\boldsymbol{t}_{\mathrm{c}}^{\mathrm{w}}$ are difficult to obtain in both VLP and computer vision-based localization due to the inherent drawbacks of the two techniques. However, V-P4L can address this challenge due to the combination of time- and space-domain information. In the following sections, we will introduce the details of V-P4L.

[4]The pose of the camera can be represented by both $\boldsymbol{R}_{\mathrm{c}}^{\mathrm{w}} \in \mathbb{R}^{3\times3}$ and $\boldsymbol{R}_{\mathrm{w}}^{\mathrm{c}} \in \mathbb{R}^{3\times3}$ [30], [36]. In particular, $\boldsymbol{R}_{\mathrm{c}}^{\mathrm{w}}$ denotes the rotation matrix from CCS to WCS, while $\boldsymbol{R}_{\mathrm{w}}^{\mathrm{c}}$ denotes the rotation matrix from WCS to CCS. The absolute values of the rotation angles in $\boldsymbol{R}_{\mathrm{c}}^{\mathrm{w}}$ and $\boldsymbol{R}_{\mathrm{w}}^{\mathrm{c}}$ are the same. In this paper, we denote the pose of the camera by $\boldsymbol{R}_{\mathrm{c}}^{\mathrm{w}}$.

III. Calculation of the Luminaire Information in CCS

Next, we calculate the information of the luminaire in CCS, including its normal vector $\boldsymbol{n}_{\mathrm{LED}}^{\mathrm{c}}$ and its vertices' coordinates $\boldsymbol{P}_i^{\mathrm{c}}$, $i \in \{1,2,3,4\}$ based on the space-domain information. Since the luminaire and the camera construct a rectangular pyramid which contains many geometric features such as points, lines, faces, and polyhedrons, plane geometry theory and solid geometry theory can be used to estimate $\boldsymbol{n}_{\mathrm{LED}}^{\mathrm{c}}$ and $\boldsymbol{P}_i^{\mathrm{c}}$, $i \in \{1,2,3,4\}$ in two steps. In the first step, $\boldsymbol{n}_{\mathrm{LED}}^{\mathrm{c}}$ is estimated based on the plane geometry theory and the solid geometry theory. Then, based on $\boldsymbol{n}_{\mathrm{LED}}^{\mathrm{c}}$, $\boldsymbol{P}_i^{\mathrm{c}}$, $i \in \{1,2,3,4\}$ are estimated by the solid geometry theory.

A. The Normal Vector of the Luminaire in CCS

Next, we introduce the calculation of the luminaire's orientation information $\boldsymbol{n}_{\mathrm{LED}}^{\mathrm{c}}$. In ICS, the point-normal form equation of a given l_{ij} can be expressed as [32]

$$x^{\mathrm{i}} \cos\phi_{l_{ij}} + y^{\mathrm{i}} \sin\phi_{l_{ij}} = \rho_{l_{ij}}, \tag{3}$$

where $\left(x^{\mathrm{i}}, y^{\mathrm{i}}\right)^{\mathrm{T}}$ is the image coordinate of a point on l_{ij} which can be obtained by the single-view geometry theory, $\phi_{l_{ij}}$ is the rotation angle from Y^{i}-axis to l_{ij} in the counter clockwise direction, and $\rho_{l_{ij}}$ is the distance from O^{i} to l_{ij}. Since p_i and p_j are on l_{ij}, $\phi_{l_{ij}}$ and $\rho_{l_{ij}}$ can be obtained based on the image coordinates of p_i and p_j. From (1), there are two points with image coordinates $\boldsymbol{p}_{l_{ij},1}^{\mathrm{i}} = \left(\frac{\rho_{l_{ij}}}{\cos\phi_{l_{ij}}}, 0\right)^{\mathrm{T}}$ and $\boldsymbol{p}_{l_{ij},2}^{\mathrm{i}} = \left(0, \frac{\rho_{l_{ij}}}{\sin\phi_{l_{ij}}}\right)^{\mathrm{T}}$ on l_{ij}. Since the two points are also on the image plane, their camera coordinates are $\boldsymbol{p}_{l_{ij},1}^{\mathrm{c}} = \left(\frac{\rho_{l_{ij}}}{\cos\phi_{l_{ij}}}, 0, f\right)^{\mathrm{T}}$ and $\boldsymbol{p}_{l_{ij},2}^{\mathrm{c}} = \left(0, \frac{\rho_{l_{ij}}}{\sin\phi_{l_{ij}}}, f\right)^{\mathrm{T}}$. Since the two points and $O^{\mathrm{c}} = (0,0,0)^{\mathrm{T}}$ are on Π_{ij}, Π_{ij} in CCS in the general form as

$$A_{\Pi_{ij}} x^{\mathrm{c}} + B_{\Pi_{ij}} y^{\mathrm{c}} + C_{\Pi_{ij}} z^{\mathrm{c}} = 0, \tag{4}$$

where $A_{\Pi_{ij}} = f\cos\phi_{l_{ij}}$, $B_{\Pi_{ij}} = f\sin\phi_{l_{ij}}$, and $C_{\Pi_{ij}} = -\rho_{l_{ij}}$.

In CCS, the general form equation of rectangle $P_1P_2P_3P_4$ can be given by:

$$A_{\mathrm{LED}}x^{\mathrm{c}} + B_{\mathrm{LED}}y^{\mathrm{c}} + C_{\mathrm{LED}}z^{\mathrm{c}} = 1, \tag{5}$$

where A_{LED}, B_{LED}, and C_{LED} are unknown constants. From (5), the normal vector of the rectangle $P_1P_2P_3P_4$ will be $(A_{\mathrm{LED}}, B_{\mathrm{LED}}, C_{\mathrm{LED}})^{\mathrm{T}}$. Let $\boldsymbol{n}^{\mathrm{c}}_{\Pi_{ij}} = \left(A_{\Pi_{ij}}, B_{\Pi_{ij}}, C_{\Pi_{ij}}\right)^{\mathrm{T}}$ $(i,j \in \{1,2,3,4\}, i \neq j)$ be the normal vector of Π_{ij} and $\boldsymbol{v}^{\mathrm{c}}_{L_{ij}} \in \mathbb{R}^3$ $(i,j \in \{1,2,3,4\}, i \neq j)$ be the direction vector of L_{ij}. Since L_{ij} is the intersection line of the rectangle $P_1P_2P_3P_4$ and Π_{ij}, we have $\boldsymbol{v}^{\mathrm{c}}_{L_{ij}} = (A_{\mathrm{LED}}, B_{\mathrm{LED}}, C_{\mathrm{LED}})^{\mathrm{T}} \times \boldsymbol{n}^{\mathrm{c}}_{\Pi_{ij}}$. Based on the solid geometry theory, we have

$$\begin{cases} \boldsymbol{v}^{\mathrm{c}}_{L_{34}} \cdot \boldsymbol{n}^{\mathrm{c}}_{\Pi_{12}} = 0, \\ \boldsymbol{v}^{\mathrm{c}}_{L_{41}} \cdot \boldsymbol{n}^{\mathrm{c}}_{\Pi_{23}} = 0. \end{cases} \tag{6}$$

Define $m = \frac{A_{\mathrm{LED}}}{C_{\mathrm{LED}}}$ and $n = \frac{B_{\mathrm{LED}}}{C_{\mathrm{LED}}}$. We can now obtain m and n as the functions of $A_{\Pi_{ij}}$, $B_{\Pi_{ij}}$, and $C_{\Pi_{ij}}$ by solving (6). Therefore, the normalized normal vector of the rectangle $P_1P_2P_3P_4$ (i.e., the orientation of the luminaire) in CCS can be expressed as

$$\boldsymbol{n}^{\mathrm{c}}_{\mathrm{LED}} = (\cos\alpha, \cos\beta, \cos\gamma)^{\mathrm{T}}, \tag{7}$$

where

$$\begin{cases} \cos\alpha = \frac{m}{\sqrt{m^2+n^2+1}}, \\ \cos\beta = \frac{n}{\sqrt{m^2+n^2+1}}, \\ \cos\gamma = \frac{1}{\sqrt{m^2+n^2+1}}. \end{cases} \tag{8}$$

B. Camera Coordinates of the Luminaire's Vertices

Given the normal vector of the luminaire estimated in Subsection III-A, we now can compute the luminaire's coordinate information $\boldsymbol{P}^{\mathrm{c}}_i$ $(i \in \{1,2,3,4\})$. Since P_1 is the intersection point of the rectangle $P_1P_2P_3P_4$, Π_{12}, and Π_{41}, its camera coordinate can be calculated as $\boldsymbol{P}^{\mathrm{c}}_1 = \frac{\boldsymbol{M}_{P_1}}{C_{\mathrm{LED}}}$, where

$$\boldsymbol{M}_{P_1} = \begin{bmatrix} m & n & 1 \\ A_{\Pi_{12}} & B_{\Pi_{12}} & C_{\Pi_{12}} \\ A_{\Pi_{41}} & B_{\Pi_{41}} & C_{\Pi_{41}} \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}. \tag{9}$$

The other three $\boldsymbol{M}_{P_i}$ $(i \in \{2,3,4\})$ can be calculated in a similar method as done in (9). In general, the camera coordinate $\boldsymbol{P}^{\mathrm{c}}_i$ $(i \in \{1,2,3,4\})$ will be given by:

$$\boldsymbol{P}^{\mathrm{c}}_i = \frac{\boldsymbol{M}_{P_i}}{C_{\mathrm{LED}}}. \tag{10}$$

From (10), we can observe that $\boldsymbol{P}^{\mathrm{c}}_i$ can be represented by C_{LED}. Next, we compute C_{LED} based on the solid geometry theory.

The volume of the rectangular pyramid $O^{\mathrm{c}} - P_1P_2P_3P_4$ is given by: $V = \frac{1}{3}SH$, where S is the area of the luminaire and is known in advance. Additionally, $H = \frac{1}{C_{\mathrm{LED}}\sqrt{m^2+n^2+1}}$ is the distance from O^{c} to the rectangle $P_1P_2P_3P_4$. For the triangular pyramid $O^{\mathrm{c}} - P_1P_2P_3$, its volume can be calculated as

$$V_1 = \frac{1}{6}\left|\det(\boldsymbol{M}_{V_1})\right|, \tag{11}$$

where $\boldsymbol{M}_{V_1} = [\boldsymbol{P}^{\mathrm{c}}_1, \boldsymbol{P}^{\mathrm{c}}_2, \boldsymbol{P}^{\mathrm{c}}_3]^{\mathrm{T}}$. Substituting (10) into (11), we have $V_1 = \frac{q_1}{C^3_{\mathrm{LED}}}$ where $q_1 = \frac{1}{6}\left|\det(\boldsymbol{M}_{q_1})\right|$, with $\boldsymbol{M}_{q_1} = [\boldsymbol{M}_{P_1}, \boldsymbol{M}_{P_2}, \boldsymbol{M}_{P_3}]^{\mathrm{T}}$. The volumes of the other three triangular pyramid $O^{\mathrm{c}} - P_2P_3P_4$, $O^{\mathrm{c}} - P_3P_4P_1$, and $O^{\mathrm{c}} - P_4P_1P_2$, are represented by V_2, V_3, and V_4, respectively, which can be obtained in the same way. Since $V = \frac{1}{2}\sum_{i=1}^{4} V_i$, C_{LED} can be calculated as

$$C_{\mathrm{LED}} = \sqrt{\frac{3\sum_{i=1}^{4} q_i\sqrt{m^2+n^2+1}}{2S}}. \tag{12}$$

Substituting (12) into (10), $\boldsymbol{P}^{\mathrm{c}}_i$ $(i \in \{1,2,3,4\})$ can be obtained.

Note that in the conventional P4L algorithm [32], the camera can only estimate the relative position between the camera and the beacons, i.e., the camera coordinates of the beacons. In contrast, V-P4L can obtain the absolute position of the camera based on the following implementation.

IV. Basic V-P4L Algorithm

We now propose the basic V-P4L algorithm that can be used for scenarios in which the LEDs have the same height. This algorithm has three steps. In the first step, based on the orientation information of the luminaire estimated in Section III, the camera's rotation angles corresponding to the $X^{\mathrm{c}}-$axis and $Y^{\mathrm{c}}-$axis can be obtained by single-view geometry theory which expresses the relationship among the coordinates of the LEDs, the position of the camera and the pose of the camera. Then, based on the time-domain information captured by the camera, the basic V-P4L algorithm can properly match the 3D-2D correspondences, so as to obtain the rotation angles corresponding to the $Z^{\mathrm{c}}-$axis and the 2D coordinate of the camera. Finally, based on the single-view geometry theory, V-P4L can estimate the z-coordinate of the camera.

A. Rotation Angles on the $X^{\mathrm{c}}-$Axis and $Y^{\mathrm{c}}-$Axis

Let $\boldsymbol{R}_X$, $\boldsymbol{R}_Y$, and $\boldsymbol{R}_Z$ be the rotation matrices of WCS along the $X^{\mathrm{c}}-$axis, $Y^{\mathrm{c}}-$axis, and $Z^{\mathrm{c}}-$axis, respectively. $\boldsymbol{R}_X$, $\boldsymbol{R}_Y$, and $\boldsymbol{R}_Z$ are given by [37],

$$\boldsymbol{R}_X = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi & -\sin\varphi \\ 0 & \sin\varphi & \cos\varphi \end{bmatrix}, \tag{13}$$

$$\boldsymbol{R}_Y = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix}, \tag{14}$$

$$\boldsymbol{R}_Z = \begin{bmatrix} \cos\psi & -\sin\psi & 0 \\ \sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{15}$$

where $\varphi \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right]$, $\theta \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right]$, and $\psi \in (-\pi, \pi]$ are the unknown Euler angles corresponding to the $X^{\mathrm{c}}-$axis, $Y^{\mathrm{c}}-$axis, and $Z^{\mathrm{c}}-$axis, respectively[5]. The rotation matrix $\boldsymbol{R}^{\mathrm{w}}_{\mathrm{c}}$ from CCS to WCS can be given by [37]

$$\boldsymbol{R}^{\mathrm{w}}_{\mathrm{c}} = \boldsymbol{R}_Z\boldsymbol{R}_Y\boldsymbol{R}_X. \tag{16}$$

Since LEDs have the same height, the normal vector of the luminaire can be given by $\boldsymbol{n}^{\mathrm{w}}_{\mathrm{LED}} = (0,0,1)^{\mathrm{T}}$. Based on single-view geometry theory, the relationship between $\boldsymbol{n}^{\mathrm{w}}_{\mathrm{LED}} = (0,0,1)^{\mathrm{T}}$ and $\boldsymbol{n}^{\mathrm{c}}_{\mathrm{LED}} = (\cos\alpha, \cos\beta, \cos\gamma)^{\mathrm{T}}$ will be given as [30]

$$\boldsymbol{n}^{\mathrm{w}}_{\mathrm{LED}} = \boldsymbol{R}^{\mathrm{w}}_{\mathrm{c}} \cdot \boldsymbol{n}^{\mathrm{c}}_{\mathrm{LED}}. \tag{17}$$

[5]The orientation of the receiver also can be obtained by the three rotation angles.

Therefore, we have

$$\begin{cases} \cos\alpha = -\sin\theta, \\ \cos\beta = \cos\theta \cdot \sin\varphi, \\ \cos\gamma = \cos\theta \cdot \cos\varphi. \end{cases} \tag{18}$$

The estimated rotation angles $\hat{\varphi}$ and $\hat{\theta}$ can be obtained by solving (18).

B. Rotation Angles on the $Z^{\mathrm{c}}-$Axis and the 2D Coordinate of the Camera

Given the calculated rotation angles $\hat{\varphi}$ and $\hat{\theta}$, the rotation angles corresponding to $Z^{\mathrm{c}}-$axis and the 2D coordinates of the camera can be estimated. Based on the single-view geometry theory, the relationship between $\boldsymbol{P}_i^{\mathrm{c}}$ $(i \in \{1,2,3,4\})$ and $\boldsymbol{P}_i^{\mathrm{w}} = (x_i^{\mathrm{w}}, y_i^{\mathrm{w}}, z_i^{\mathrm{w}})^{\mathrm{T}}$ can be given as [30]

$$\boldsymbol{P}_i^{\mathrm{w}} = \boldsymbol{R}_{\mathrm{c}}^{\mathrm{w}} \cdot \boldsymbol{P}_i^{\mathrm{c}} + \boldsymbol{t}_{\mathrm{c}}^{\mathrm{w}}, \tag{19}$$

where $\boldsymbol{P}_i^{\mathrm{w}}$ is known in advance and can be obtained by the camera. In addition, $\boldsymbol{P}_i^{\mathrm{c}}$ is the space-domain information that is estimated in Subsection III-B. Moreover, $\boldsymbol{t}_{\mathrm{c}}^{\mathrm{w}} = (t_x, t_y, t_z)^{\mathrm{T}}$ is the 3D world coordinate of the camera. In (19), there are four unknown parameters ψ, t_x, t_y, and t_z. If the 3D-2D correspondences are known in advance, we can easily obtain the four unknown parameters using the world and camera coordinates of the four vertices. However, we can only match the vertex P_1 which transmits information and its projection p_1. Fortunately, based on the perspective projection theory, the projection of the rectangle is still a quadrangle. Additionally, the order of the four LEDs and the their projections are corresponding in the clockwise direction, as shown in Fig. 1. Therefore, based on the time-domain information transmitted by VLC, the 3D-2D correspondences can be properly matched. We can now calculate the four unknown parameters ψ, t_x, t_y, and t_z based on the LEDs' information in both the time and space domains. For mathematical analysis, we define

$$\begin{aligned} \boldsymbol{R}_Y \boldsymbol{R}_X &= \begin{bmatrix} \cos\theta & \sin\theta\cdot\sin\varphi & \sin\theta\cdot\cos\varphi \\ 0 & \cos\varphi & -\sin\varphi \\ -\sin\theta & \cos\theta\cdot\sin\varphi & \cos\theta\cdot\cos\varphi \end{bmatrix} \\ &= \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix}, \end{aligned} \tag{20}$$

and thus we can rewrite (19) as

$$\boldsymbol{P}_i^{\mathrm{w}} = \begin{bmatrix} \cos\psi & -\sin\psi & 0 \\ \sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix} \boldsymbol{P}_i^{\mathrm{c}} + \boldsymbol{t}_{\mathrm{c}}^{\mathrm{w}}. \tag{21}$$

The three unknown parameters ψ, t_x, and t_y in (21) can be calculated by the LLS estimator, which can be expressed in a matrix form as

$$\boldsymbol{A}_{ij}\boldsymbol{x} = \boldsymbol{b}_{ij}, \tag{22}$$

where $\boldsymbol{A}_{ij} = [\boldsymbol{A}_i; \boldsymbol{A}_j]$ and $\boldsymbol{b}_{ij} = [\boldsymbol{b}_i; \boldsymbol{b}_j]$ $(i, j \in \{1,2,3,4\}, i \neq j, r \neq s)$, with

$$\begin{aligned} \boldsymbol{A}_i &= \begin{bmatrix} [a_1,a_2,a_3]^{\mathrm{T}} \cdot \boldsymbol{P}_i^{\mathrm{c}} & -[b_1,b_2,b_3]^{\mathrm{T}} \cdot \boldsymbol{P}_i^{\mathrm{c}} & 1 & 0 \\ [b_1,b_2,b_3]^{\mathrm{T}} \cdot \boldsymbol{P}_i^{\mathrm{c}} & [a_1,a_2,a_3]^{\mathrm{T}} \cdot \boldsymbol{P}_i^{\mathrm{c}} & 0 & 1 \end{bmatrix} \\ &(i \in \{1,2,3,4\}), \end{aligned} \tag{23}$$

$$\boldsymbol{x} = [\cos\psi, \sin\psi, t_x, t_y]^{\mathrm{T}}, \tag{24}$$

and

$$\boldsymbol{b}_i = [x_i^{\mathrm{w}}, y_i^{\mathrm{w}}]^{\mathrm{T}} \, (i \in \{1,2,3,4\}). \tag{25}$$

Therefore, the unknown parameters can be given by:

$$\hat{\boldsymbol{x}} = (\boldsymbol{A}_{rs}^{\mathrm{T}}\boldsymbol{A}_{rs})^{-1}\boldsymbol{A}_{rs}^{\mathrm{T}}\boldsymbol{b}_{ij}, \tag{26}$$

where $\hat{\boldsymbol{x}} = \left[\hat{\cos\psi}, \hat{\sin\psi}, \hat{t}_x, \hat{t}_y\right]^{\mathrm{T}}$ is the estimation of $\boldsymbol{x}$.

Hence, based on the information in both time and space domains, the basic V-P4L algorithm can properly match the 3D-2D correspondences. Meanwhile, the rotation angles corresponding to the $Z^{\mathrm{c}}-$axis ψ and the 2D coordinate of the camera $(t_x, t_y)^{\mathrm{T}}$ can be obtained.

C. Computation of the z-Coordinate of the Camera

In Subsection IV-B, we have obtained ψ and $(t_x, t_y)^{\mathrm{T}}$. In (21), there is still one unknown parameter t_z. Recall that, here, the LEDs have the same height, i.e., $z_1^{\mathrm{w}} = z_2^{\mathrm{w}} = z_3^{\mathrm{w}} = z_4^{\mathrm{w}}$. Based on the single-view geometry theory, we can obtain the relationship between z_i^{w} and t_z from (21) as

$$z_i^{\mathrm{w}} = [c_1, c_2, c_3]^{\mathrm{T}} \cdot \boldsymbol{P}_i^{\mathrm{c}} + t_z. \tag{27}$$

The estimated z-coordinate of the camera in WCS $\hat{t}_z$ can be calculated as

$$\hat{t}_z = \frac{1}{4}\left(\sum_{i=1}^{4} z_i^{\mathrm{w}} - \sum_{i=1}^{4} [\hat{c}_1, \hat{c}_2, \hat{c}_3]^{\mathrm{T}} \cdot \boldsymbol{P}_i^{\mathrm{c}}\right), \tag{28}$$

where $\hat{c}_k$ $(k \in \{1,2,3\})$ is the estimation of c_k.

Therefore, the estimated pose $\hat{\boldsymbol{R}}_{\mathrm{c}}^{\mathrm{w}}$ and position of the camera $\hat{\boldsymbol{t}}_{\mathrm{c}}^{\mathrm{w}}$ can be obtained without the ideal 2D-3D correspondence assumption. In this way, when LEDs have the same height, the basic V-P4L algorithm can be implemented for high accuracy at low complexity. However, when LEDs have different heights, the normal vector of the luminaire is not $(0,0,1)^{\mathrm{T}}$, i.e., the three rotation angles do not satisfy (18), which means the basic V-P4L algorithm cannot be implemented in the scenarios. Therefore, we will propose a correction algorithm for V-P4L in Section V for localization in the scenarios.

V. Correction algorithm for V-P4L

Most existing studies including our approach in Subsection IV assume that LEDs have the same height [16], [24]. However, this may not always be true in practice. For instance, ceilings may be tilted due to imperfect decoration or as a deliberate design. In these scenarios, the localization accuracy can be significantly degraded. Therefore, next, we propose a correction algorithm for V-P4L, called V-P4L-DH, that extends the basic V-P4L algorithm to the scenarios in which the LEDs have different heights. Based on the single-view geometry theory and the LLS method, V-P4L-DH can obtain the 2D position of the camera. Based on the 2D localization, V-P4L-DH can achieve 3D localization by a simple optimization method.

A. 2D Localization

For the 2D-localization case in which the z-coordinate of the camera t_z is known in advance, based on single-view geometry theory, (27) can be rewritten as

$$[c_1, c_2, c_3]^{\mathrm{T}} \cdot \boldsymbol{P}_i^{\mathrm{c}} = z_i^{\mathrm{w}} - t_z, \tag{29}$$

where $i \in \{1,2,3,4\}$. Based on the LLS method, we can estimate c_k ($k \in \{1,2,3\}$) by solving (29) using four $\boldsymbol{P}_i^{\mathrm{c}}$ and z_i^{w}. Then, from (20), we have:

$$\begin{cases} \hat{c}_1 = -\sin\theta, \\ \hat{c}_2 = \cos\theta \cdot \sin\varphi, \\ \hat{c}_3 = \cos\theta \cdot \cos\varphi, \end{cases} \tag{30}$$

where $\hat{c}_k$ ($k \in \{1,2,3\}$) is the estimation of c_k. The estimated rotation angles corresponding to the $X^{\mathrm{c}}-$axis $\hat{\varphi}$ and $Y^{\mathrm{c}}-$axis $\hat{\theta}$ can be obtained by solving (30). Then, the pose of the camera $\hat{\boldsymbol{R}}_{\mathrm{c}}^{\mathrm{w}}$ and 2D position of the camera $\hat{\boldsymbol{t}}_{\mathrm{c,2D}}^{\mathrm{w}} = \left(\hat{t}_x, \hat{t}_y\right)^{\mathrm{T}}$ can be obtained according to (20)$-$(26), where $\hat{t}_x$ and $\hat{t}_y$ are the estimated x and y-coordinates of the camera, respectively.

B. 3D Localization

For the 3D-localization case, the z-coordinate of the camera t_z is unknown in advance. For an indoor scenario, the value of t_z must be within $[0, H_{\mathrm{m}})$, where H_{m} is the maximum height of the room. Based on the above 2D-localizatoin algorithm, for different $t_z \in [0, H_{\mathrm{m}})$, we can obtain different $\hat{\boldsymbol{R}}_{\mathrm{c}}^{\mathrm{w}}(t_z)$ and $\hat{\boldsymbol{t}}_{\mathrm{c,2D}}^{\mathrm{w}}(t_z)$. Based on the estimated normal vector of the luminaire in CCS $\boldsymbol{n}_{\mathrm{LED}}^{\mathrm{c}}$, we have [30]:

$$\hat{\boldsymbol{n}}_{\mathrm{LED}}^{\mathrm{w}}(t_z) = \hat{\boldsymbol{R}}_{\mathrm{c}}^{\mathrm{w}}(t_z) \cdot \boldsymbol{n}_{\mathrm{LED}}^{\mathrm{c}}, \tag{31}$$

where $\hat{\boldsymbol{n}}_{\mathrm{LED}}^{\mathrm{w}}(t_z)$ is the estimated normal vector of the luminaire in WCS when the z-coordinate of the camera is t_z. Since the world coordinates of the luminaire's vertices $\boldsymbol{P}_i^{\mathrm{w}}$ ($i \in \{1,2,3,4\}$) are known in advance, the actual normal vector of the luminaire in WCS $\boldsymbol{n}_{\mathrm{LED}}^{\mathrm{w}}$ can be calculated as

$$\boldsymbol{n}_{\mathrm{LED}}^{\mathrm{w}} = \left(\boldsymbol{P}_i^{\mathrm{w}} - \boldsymbol{P}_j^{\mathrm{w}}\right) \times \left(\boldsymbol{P}_i^{\mathrm{w}} - \boldsymbol{P}_k^{\mathrm{w}}\right), \tag{32}$$

where $i,j,k \in \{1,2,3,4\}, i \neq j \neq k$. Therefore, the difference between the estimated and actual normal vectors of the luminaire in WCS can be given as

$$\Delta G(t_z) = \left|\boldsymbol{n}_{\mathrm{LED}}^{\mathrm{w}} - \hat{\boldsymbol{n}}_{\mathrm{LED}}^{\mathrm{w}}(t_z)\right|. \tag{33}$$

The estimated z-coordinate of the camera $\hat{t}_z$ will satisfy the following equation

$$\hat{t}_z = \arg\min_{t_z} \Delta G(t_z). \tag{34}$$

To reduce the complexity of V-P4L-DH, we propose a K-step segmentation optimization strategy. This strategy can significantly reduce the search range of t_z through each step, and it allows us to directly eliminate most of the false candidate values of t_z. In particular, in the first step, we set the interval between adjacent t_z is ε_1, such that

$$t_z \in \{0, \varepsilon_1, 2\varepsilon_1, 3\varepsilon_1, \ldots, H_{\mathrm{m}}\}. \tag{35}$$

Substituting all the t_z into (33), we have $\Delta G(t_z) = \left\{\Delta G(t_{z,0}), \Delta G(t_{z,1}), \ldots, \Delta G\left(t_{z,\frac{H_{\mathrm{m}}}{\varepsilon_1}}\right)\right\}$. According to (34), we can find the minimal $\Delta G(t_z)$. We denote the minimal $\Delta G(t_z)$ by $\Delta G(t_{z,s})$ where $t_{z,s} = \arg\min_{t_z} \Delta G(t_z), s \in \left\{0, 1, \ldots, \frac{H_{\mathrm{m}}}{\varepsilon_1}\right\}$. In the second step, we reduce the range of t_z to $(\varepsilon_1(s-1), \varepsilon_1(s+1))$. We set

$$t_z \in \{\varepsilon_1(s-1) + \varepsilon_2, \varepsilon_1(s-1) + 2\varepsilon_2, \ldots, \varepsilon_1(s+1) - \varepsilon_2\}, \tag{36}$$

where ε_2 is the interval between adjacent t_z, and $\varepsilon_2 < \varepsilon_1$. Then, we can find the minimal $\Delta G(t_z)$ and the corresponding t_z. We repeat the process for K steps until we can find the precise $\hat{t}_z$ according to (34), so as to obtain the optimal $\hat{\boldsymbol{R}}_{\mathrm{c}}^{\mathrm{w}}$ and $\hat{\boldsymbol{t}}_{\mathrm{c,2D}}^{\mathrm{w}}$. Therefore, based on the simple optimization method, we can obtain the pose and 3D position of the receiver when LEDs have different heights. Although V-P4L-DH requires iterations to estimate the position and pose of the receiver, its computation complexity can be low to achieve high localization accuracy. Based on the proposed K-step segmentation optimization strategy, the computation complexity of V-P4L-DH can be expressed as $\mathcal{O}\left(\frac{H_{\mathrm{m}}}{\varepsilon_1} + \frac{2H_{\mathrm{m}}}{\varepsilon_1\varepsilon_2} + \frac{2^2 H_{\mathrm{m}}}{\varepsilon_1\varepsilon_2\varepsilon_3} + \cdots\right)$, where $\frac{2^{k-1}H_{\mathrm{m}}}{\varepsilon_1\varepsilon_2\cdots\varepsilon_k}$ is the number of iterations in the kth step. For a typical H_{m}, tens of iterations are basically sufficient to achieve centimeter-level accuracy.

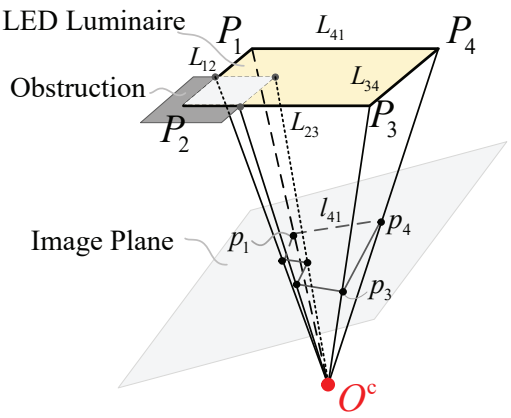


Fig. 2. A occlusion scenario where the projection of P_2, p_2, is blocked by barriers and not on the image plane.

V-P4L only requires the camera to capture the information transmitted by one vertex of the rectangle luminaire. Therefore, if the rest vertices are blocked by barriers, V-P4L still can be used for localization. For instance, in Fig. 2, the LED at P_1 transmits information. If the projection p_2 of P_2 is blocked by barriers and not on the image plane as shown in Fig. 2, the pixel coordinate of p_2 can be determined by the intersection of l_{12} and l_{23}, and thus V-P4L can be still successively implemented. Hence, V-P4L can be used for localization under partial occlusion regardless of the height differences among LEDs without the limitation on 3D-2D correspondences using a single luminaire.

V-P4L algorithm is summarized in Algorithm 1. For brevity, we only give the first step of the segmentation optimization strategy in Algorithm 1.

VI. Simulation and Experimental Results and Analysis

A. Simulation Setup

For our simulations, we consider a positioning system with a rectangular luminaire and a camera. The LED luminaire is deployed in the center of the ceiling. The length of the luminaire which is along the X^{w}-axis is set to 120 cm [12], [38], and the width of luminaire which is along the Y^{w}-axis will be varied according to configurations. We set the

Algorithm 1 V-P4L Algorithm.

Input:

$i, j \in \{1, 2, 3, 4\}, i \neq j$;
$\boldsymbol{P}_i^{\mathrm{w}} = (x_i^{\mathrm{w}}, y_i^{\mathrm{w}}, z_i^{\mathrm{w}})^{\mathrm{T}}$;
$\boldsymbol{p}_j^{\mathrm{p}} = \left(u_j^{\mathrm{p}}, v_j^{\mathrm{p}}\right)^{\mathrm{T}}$;
u_0, v_0, f, f_u, f_v, d_x, d_y and ε_1;
The shape of the luminaire, and the order of the four vertices of the luminaire.

Obtain $\boldsymbol{p}_i^{\mathrm{i}}$ according to (1).
Calculate $\phi_{l_{ij}}$ and $\rho_{l_{ij}}$ by $\boldsymbol{p}_i^{\mathrm{i}}$ and $\boldsymbol{p}_j^{\mathrm{i}}$.
Compute $A_{\Pi_{ij}} = f\cos\phi_{l_{ij}}$, $B_{\Pi_{ij}} = f\sin\phi_{l_{ij}}$, and $C_{\Pi_{ij}} = -\rho_{l_{ij}}$, and then calculate $\boldsymbol{n}_{\Pi_{ij}}^{\mathrm{c}} = \left(A_{\Pi_{ij}}, B_{\Pi_{ij}}, C_{\Pi_{ij}}\right)^{\mathrm{T}}$.
Compute $\boldsymbol{v}_{L_{ij}}^{\mathrm{c}} = \left(A_{\mathrm{LED}}, B_{\mathrm{LED}}, C_{\mathrm{LED}}\right)^{\mathrm{T}} \times \boldsymbol{n}_{\Pi_{ij}}^{\mathrm{c}}$.
Calculate m and n according to $\boldsymbol{v}_{L_{34}}^{\mathrm{c}} \cdot \boldsymbol{n}_{\Pi_{12}}^{\mathrm{c}} = 0$ and $\boldsymbol{v}_{L_{41}}^{\mathrm{c}} \cdot \boldsymbol{n}_{\Pi_{23}}^{\mathrm{c}} = 0$.
Estimate $\cos\alpha$, $\cos\beta$, and $\cos\gamma$ according to (8), and then calculate $\boldsymbol{n}_{\mathrm{LED}}^{\mathrm{c}}$ according to (7).
Calculate $\boldsymbol{M}_{P_i}$ according to (9).
Calculate $q_i = \frac{1}{6}\left|\det(\boldsymbol{M}_{q_i})\right|$.
Find C_{LED} according to (12).
Estimate $\boldsymbol{P}_i^{\mathrm{c}}$ according to (10).
if $z_1^{\mathrm{w}} = z_2^{\mathrm{w}} = z_3^{\mathrm{w}} = z_4^{\mathrm{w}}$ **then**
  $\boldsymbol{n}_{\mathrm{LED}}^{\mathrm{w}} = (0, 0, 1)^{\mathrm{T}}$.
  Estimate $\hat{\varphi}$ and $\hat{\theta}$ according to (18).
  Estimate $\hat{\psi}$, $\hat{t}_x$, and $\hat{t}_y$ according to (22) - (26).
  Estimate $\hat{t}_z$ according to (28) if t_z is unknown in advance.
else
  if t_z is known in advance **then**
    Obtain $\hat{c}_i$ $(i \in \{1, 2, 3\})$ using the same method of solving (26).
    Estimate $\hat{\varphi}$ and $\hat{\theta}$ according to (30).
    Estimate $\hat{\psi}$, $\hat{t}_x$, and $\hat{t}_y$ according to (20) - (26) to obtain $\hat{\boldsymbol{R}}_{\mathrm{c}}^{\mathrm{w}}$ and $\hat{\boldsymbol{t}}_{\mathrm{c,2D}}^{\mathrm{w}}$.
  else
    Calculate $\boldsymbol{n}_{\mathrm{LED}}^{\mathrm{w}}$ according to (32).
    for $s = 0 \rightarrow \frac{H_{\mathrm{m}}}{\varepsilon_1}$ **do**
      $\hat{t}_{z,s} = s\varepsilon_1$.
      Estimate $\hat{\boldsymbol{R}}_{\mathrm{c}}^{\mathrm{w}}(t_{z,s})$ and $\hat{\boldsymbol{t}}_{\mathrm{c,2D}}^{\mathrm{w}}(t_{z,s})$.
      Calculate $\hat{\boldsymbol{n}}_{\mathrm{LED}}^{\mathrm{w}}(t_{z,s}) = \hat{\boldsymbol{R}}_{\mathrm{c}}^{\mathrm{w}}(t_{z,s}) \cdot \boldsymbol{n}_{\mathrm{LED}}^{\mathrm{c}}$.
      Calculate $\Delta G(t_{z,s}) = \left|\boldsymbol{n}_{\mathrm{LED}}^{\mathrm{w}} - \hat{\boldsymbol{n}}_{\mathrm{LED}}^{\mathrm{w}}(t_{z,s})\right|$.
    end for
    Estimate $\hat{t}_z = \min_{t_z} \Delta G(t_z)$. Meanwhile, obtain $\hat{\boldsymbol{R}}_{\mathrm{c}}^{\mathrm{w}} = \hat{\boldsymbol{R}}_{\mathrm{c}}^{\mathrm{w}}(\hat{t}_z)$ and $\hat{\boldsymbol{t}}_{\mathrm{c,2D}}^{\mathrm{w}} = \hat{\boldsymbol{t}}_{\mathrm{c,2D}}^{\mathrm{w}}(\hat{t}_z)$.
  end if
end if

Output: $\hat{\boldsymbol{R}}_{\mathrm{c,est}}^{\mathrm{w}}$ and $\hat{\boldsymbol{t}}_{\mathrm{c,est}}^{\mathrm{w}} = \left(\hat{\boldsymbol{t}}_{\mathrm{c,2D}}^{\mathrm{w}}, \hat{t}_z\right)^{\mathrm{T}}$.

rectangular luminaire tilt with various angles along the Y^{w}-axis to represent LEDs with different heights. The simulation parameters are listed in Table II. Since V-P4L combines VLC and computer vision based localization, for baseline comparisons, we use a VLP algorithm named enhanced camera assisted received signal strength ratio algorithm (eCA-RSSR) [16], and a typical computer vision algorithm termed P4L algorithm [32]. All statistical results are averaged over 1000 independent runs. For each simulation run, the receiver positions are selected in the room randomly. The pinhole camera is calibrated. The image noise is modeled as a white Gaussian noise having an expectation of zero and a standard deviation of 2 pixels [36]. Since the image noise affects the pixel coordinate of the luminaire's projection on the image plane, the pixel coordinate is obtained by processing 20 images for the same position. Moreover, we use a two-step segmentation optimization strategy, and we set $\varepsilon_1 = 10\,\mathrm{cm}$ and $\varepsilon_2 = 1\,\mathrm{cm}$, which means the number of iterations is 50.

Note that since eCA-RSSR requires three LEDs for localization, we assume that the four LEDs at the vertices of the luminaire are used for eCA-RSSR and we choose the three LEDs with the highest RSSs to achieve best performance for eCA-RSSR. Additionally, all the LEDs transmit different information in eCA-RSSR, which is not required by V-P4L and the P4L algorithms. Therefore, compared with V-P4L, the VLC link of eCA-RSSR is more complex. Furthermore, eCA-RSSR relies on the perfect Lambertian pattern model. We set a random deviation $\delta_1 \leq 10\%$ for the Lambertian

TABLE II
SIMULATION PARAMETERS.

Parameter	Value	
Room size (length × width × height)	$5\,\mathrm{m} \times 5\,\mathrm{m} \times 3\,\mathrm{m}$	
Length and width of LED luminaire	Length	Width
	$120\,\mathrm{cm}$	$20\,\mathrm{cm} - 100\,\mathrm{cm}$
LED semi-angle, $\Phi_{1/2}$	60°	
Principal point of camera	$(u_0, v_0) = (320, 240)$	
Focal ratio of camera	$f_u = f_v = 800$	
The distance between the PD and the camera in eCA-RSSR, d_{pc}	$1\,\mathrm{cm}$	

pattern model for eCA-RSSR, conservatively. On the other hand, the P4L algorithm exploits a rectangle to estimate the position and pose of the camera. The P4L algorithm assumes that the camera knows the 3D-2D correspondences, which is impractical since the beacon in the P4L algorithm cannot convey time-domain information to the camera. We set a random error rate $\delta_2 \leq 10\%$ for the 3D-2D correspondences for P4L algorithm, conservatively. Moreover, the P4L algorithm can only obtain the relative position. For comparison with V-P4L, we transform the relative position into the absolute position for the P4L algorithm.

We evaluate the performance of V-P4L in terms of the accuracy of its position and pose estimation. We define position error as $E_P = |\boldsymbol{r}^{\mathrm{w}}_{\mathrm{true}} - \boldsymbol{r}^{\mathrm{w}}_{\mathrm{est}}|$, where $\boldsymbol{r}^{\mathrm{w}}_{\mathrm{true}} = \left(x^{\mathrm{w}}_{r,\mathrm{true}}, y^{\mathrm{w}}_{r,\mathrm{true}}, z^{\mathrm{w}}_{r,\mathrm{true}}\right)$ and $\boldsymbol{r}^{\mathrm{w}}_{\mathrm{est}} = \left(x^{\mathrm{w}}_{r,\mathrm{est}}, y^{\mathrm{w}}_{r,\mathrm{est}}, z^{\mathrm{w}}_{r,\mathrm{est}}\right)$ are the actual and estimated world coordinates of the receiver, respectively. Additionally, the accuracy of pose estimation can be measured by the orientation error which is defined as $E_O = |\Theta_{\mathrm{true}} - \Theta_{\mathrm{est}}|$, where Θ_{true} and Θ_{est} are the actual and estimated rotation angles, respectively.

In this section, we evaluate the performance of V-P4L under various tilted angles and width values of the luminaire, and different image noise levels. Hereinafter, the scenario in which the LEDs have the same height is denoted by SH, and the scenario in which the LEDs have different heights is denoted by DH. Since the basic algorithm of V-P4L is used when LEDs have the same height, we denote the basic V-P4L algorithm by V-P4L-SH in figures.

B. Simulation Results

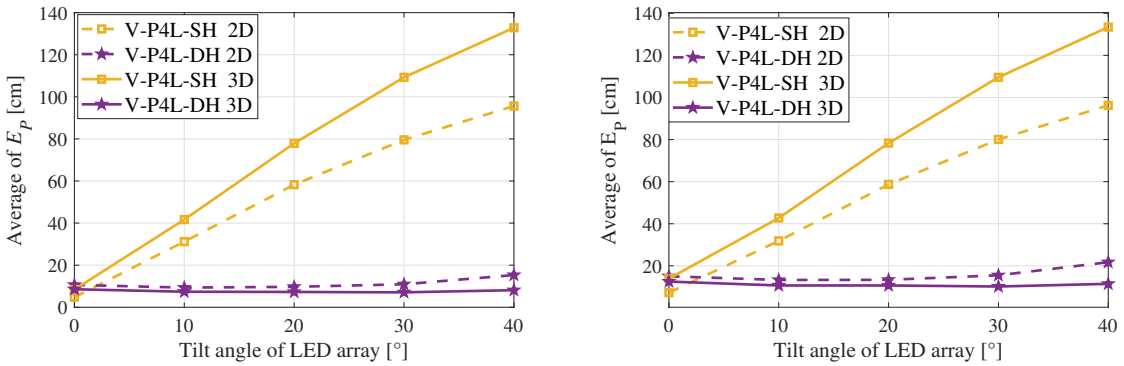
1) Effect of luminaire's tilted angle on accuracy performance: We first evaluate the effect of the tilted angle of the luminaire on localization accuracy of V-P4L for both 2D and 3D localization. This performance is represented by average E_P with the tilted angle of the luminaire varying from 0° to 40°. The width of the luminaire is 40 cm. As shown in Fig. 3(a), for 2D localization, when the tilted angle of the luminaire is 0°, i.e., LEDs have the same height, the basic V-P4L algorithm yields a slightly better performance than V-P4L-DH. This is because V-P4L-DH estimates the receiver position by optimization, while the basic V-P4L algorithm directly estimates the receiver position using geometric relations. However, as the tilted angle of the luminaire varies from 0° to 40°, the average E_P resulting from the basic V-P4L algorithm will increase from about 5 cm to about 98 cm. This is because the basic algorithm is tailored for scenarios in which the LEDs have the same height. For 3D localization, the average E_P of the basic V-P4L algorithm increases from about 10 cm to about 135 cm. In contrast, for all the tilted angles of the luminaire, the average E_P resulting from V-P4L-DH will be less than 17 cm for both 2D and 3D localization. In 3D localization, when searching for the optimal $\hat{t}_z$, the estimated 2D position and pose are optimized simultaneously. Therefore, the accuracy of 3D localization is higher than that of 2D localization.

Since the accuracy of pose estimation is also affected by the tilted angle of the luminaire, we then evaluate the effect of the tilted angle on the E_O resulting from V-P4L for both 2D and 3D localization. As shown in Fig. 4, the orientation errors along the X^{c}-axis, Y^{c}-axis and Z^{c}-axis are shown separately. When the tilted angle of the luminaire is 0°, the basic V-P4L algorithm can obtain better performance than V-P4L-DH. However, as the tilted angle of the luminaire varies from 0° to 40°, the average E_O resulting from the basic V-P4L algorithm increases from about 2° to about 20° when measured along the X^{c}-axis, Y^{c}-axis and Z^{c}-axis. In contrast, as the tilted angle of the luminaire varies from 0° to 40°, the average E_O resulting from V-P4L-DH 2D localization will be about 5° when measured along the X^{c}-axis and Z^{c}-axis, and will increase from 2° to 3° when measured along the Y^{c}-axis. Additionally, the average E_O resulting from V-P4L-DH 3D localization will be about 3° along the X^{c}-axis and Z^{c}-axis, and about 2° along the Y^{c}-axis. Therefore, for all the tilted angles of the luminaire, V-P4L-DH can consistently achieve a good performance.

We also evaluate the effect of partial occlusion on localization accuracy of V-P4L. Figure 3(a) shows the position estimation accuracy as the tilt angles of the luminaire change from 0° to 40° for a scenario without occlusion. In contrast, Fig. 3(b) shows the position estimation accuracy as the tilted angle of the luminaire changes from 0° to 40° when p_2 is blocked and is not on the image plane as shown in Fig. 2. As shown in Fig. 3(b), when the luminaire is partially blocked, V-P4L can still achieve high accuracy. In particular, the accuracy of the basic V-P4L algorithm is almost the same in Fig. 3(a) and Fig. 3(b). In addition, for V-P4L-DH, the accuracy of 2D-localization and 3D-localization will decrease by about 2 cm and 1 cm, respectively. Therefore, V-P4L is robust to partial occlusion as introduced in Section V.

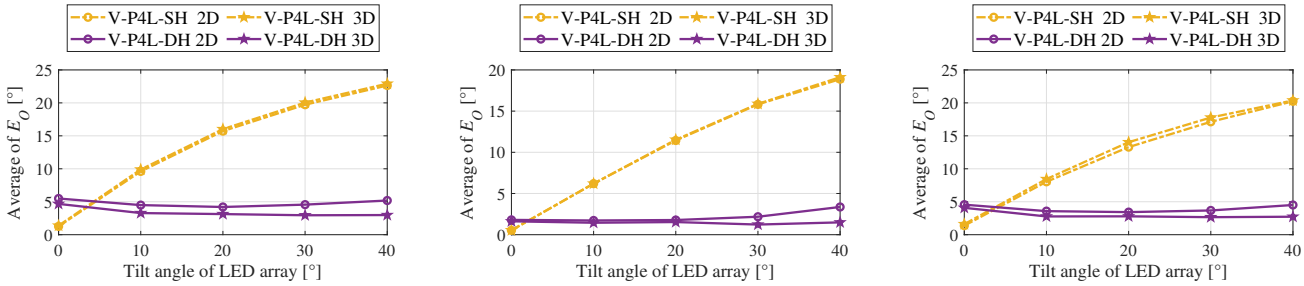
In this subsection, we have verified V-P4L can achieve high accuracy. Since 2D localization is the special case of 3D localization where the height of the receiver is known in advance, in the following subsections, we will only show the simulation results for 3D localization.

2) Effect of luminaire's width on accuracy performance: We now evaluate the effect of the luminaire's width on localization accuracy of V-P4L. This performance is captured by the average E_P as the width value varies from 20 cm to 100 cm. For scenarios in which the LEDs have different heights, the tilted angle of the luminaire is 20°. As shown in Fig. 5, under both similar and different heights for the LEDs, the proposed V-P4L is able to obtain the best performance among the three algorithms. This is because V-P4L can obtain the proper 3D-2D correspondences, and it does not rely on channel model.



(a) Without occlusion. (b) With occlusion.

Fig. 3. Position errors (E_P) as the tilted angle of the luminaire varies. In Fig. 3(b), the projection p_2 of P_2 is blocked by barriers and not on the image plane as shown in Fig. 2.



(a) Orientation errors along the X^c-axis. (b) Orientation errors along the Y^c-axis. (c) Orientation errors along the Z^c-axis.

Fig. 4. Orientation errors (E_O) as the tilt angle of the luminaire varies.

When LEDs have the same height, the average E_P resulting from the basic V-P4L algorithm will be below 19 cm. In contrast, for eCA-RSSR, this average will be around 70 cm as the width increases from 20 cm to 100 cm. Additionally, for the P4L algorithm, the average E_P decreases from over 100 cm to about 40 cm. When LEDs have different heights, the average E_P resulting from V-P4L-DH decreases from about 10 cm to 5 cm. In contrast, the average E_P resulting from both eCA-RSSR and the P4L algorithm decreases from over 100 cm to over 70 cm. As shown in Fig. 5, the localization accuracy increases with the increase of the width for V-P4L. However, the position error resulting from V-P4L will always be less than 15 cm when the width of the luminaire is longer than 20 cm regardless of the height differences among LEDs. Hence, V-P4L can be applied to popular indoor luminaires.

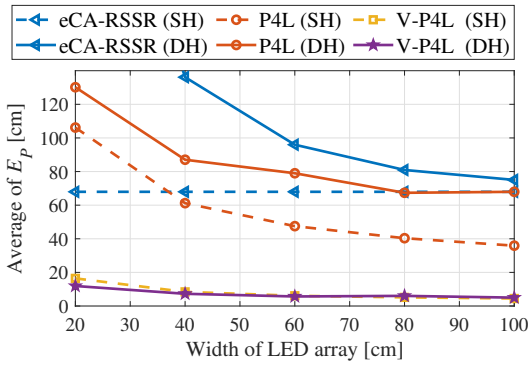
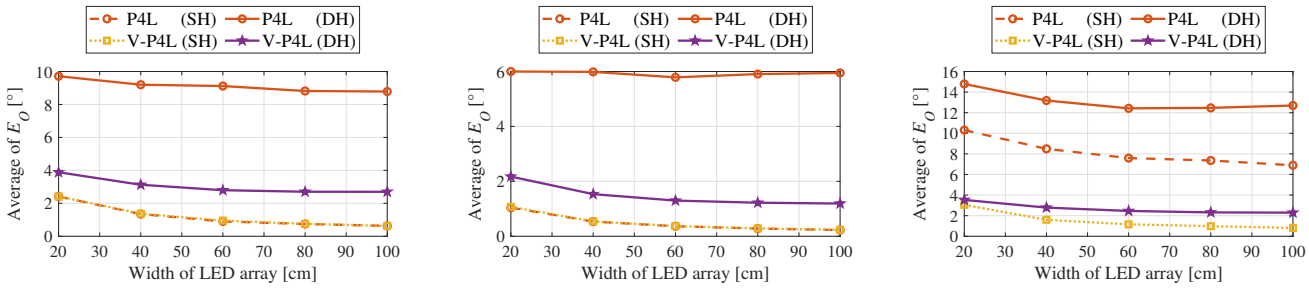


Fig. 5. Position errors (E_P) as the width of the luminaire varies.

Since the accuracy of pose estimation is also affected by the width of the LED luminaire, we now compare the orientation errors between V-P4L and P4L with varying widths of the luminaire. As shown in Fig. 6, for V-P4L, all the average orientation errors along the X^c-axis, Y^c-axis and Z^c-axis are less than 4° regardless of the height differences among the LEDs. When LEDs have the same height, the average E_O along the X^c-axis and Y^c-axis will decrease from about 2.5° to about 0.5° and from 1° to about 0.3°, respectively for both the basic V-P4L algorithm and the P4L algorithm. Additionally, the average E_O resulting from the basic V-P4L algorithm along the Z^c-axis decreases from about 3° to about 0.5°, which is over 5° better than the P4L algorithm. When LEDs have different heights, for V-P4L-DH, the average E_O along the X^c-axis, Y^c-axis and Z^c-axis decreases from 4° to 2.5°, from 2° to 1° and from 3.8° to 2°, respectively. In contrast, for the P4L algorithm, the average E_O along the X^c-axis, Y^c-axis and Z^c-axis will be about 9°, 6°, 13°, respectively. Therefore, compared with the P4L algorithm, V-P4L can obtain higher accuracy for pose estimation using popular indoor luminaire.

3) Effect of image noise on accuracy performance: Next, we evaluate the effect of image noise on the localization performance of V-P4L when the width of the luminaire is



(a) Orientation errors along the X^c-axis. (b) Orientation errors along the Y^c-axis. (c) Orientation errors along the Z^c-axis.

Fig. 6. Orientation errors (E_O) as the widths of the luminaire varies.

40 cm. When LEDs have different heights, the tilted angle of the luminaire is set to 20°. The image noise is modeled as a white Gaussian noise having an expectation of zero and a standard deviation, σ_n, ranging from 0 to 4 pixels [36]. Figure 7 shows the average position errors versus the image noise level. As shown in Fig. 7, when LEDs have the same height, the average E_P resulting from the basic V-P4L algorithm increases from 0 cm to 18 cm as the image noise increases from 0 to 4 pixels. In contrast, for the P4L algorithm, the average E_P increases from 30 cm to 40 cm. Additionally, for eCA-RSSR, the average E_P will be around 75 cm. When LEDs have different heights, the average E_P resulting from V-P4L-DH increases from 1 cm to 11 cm. In contrast, for the P4L algorithm, the average E_P is about 85 cm. Additionally, eCA-RSSR yields an average E_P of about 135 cm. Therefore, compared with the P4L algorithm and eCA-RSSR, V-P4L can obtain higher accuracy for position estimation. This is because the P4L algorithm relies on the given 3D-2D correspondences, and eCA-RSSR uses a perfect channel model. In contrast, V-P4L can obtain the 3D-2D correspondences by VLC and does not require a channel model.

Finally, we compare the accuracy of pose estimation between V-P4L and the P4L algorithm under different image noise level. Figure 8 show the average of the orientation errors along the X^c-axis, Y^c-axis, and Z^c-axis with the image noise ranging from 0 to 4 pixels. When LEDs have the same height, the average E_O resulting from the basic V-P4L algorithm along the X^c-axis, Y^c-axis, and Z^c-axis increases from 0° to 2.5°, from 0° to 1.2° and from 0° to 3.5°, respectively. In contrast, for the P4L algorithm, the average E_O along the X^c-axis and Y^c-axis increases from 0° to 1.2° and from 0° to 0.5°, respectively, which is a minor improvement over the basic V-P4L algorithm. However, the average E_O resulting from the P4L algorithm along the Z^c-axis increases from 6° to 8° which is over 4° worse than the basic V-P4L algorithm. When LEDs have different heights, for V-P4L-DH, the average E_O along the X^c-axis, Y^c-axis, and Z^c-axis increases from about 1° to less than 3.5°. In contrast, for the P4L algorithm, the average E_O along the X^c-axis, Y^c-axis, and Z^c-axis will be about 9°, 6°, 13°, respectively. Therefore, compared with the P4L algorithm, V-P4L can obtain more stable and accurate pose estimation regardless of the image noise due to the proper 3D-2D correspondences obtained by VLC.

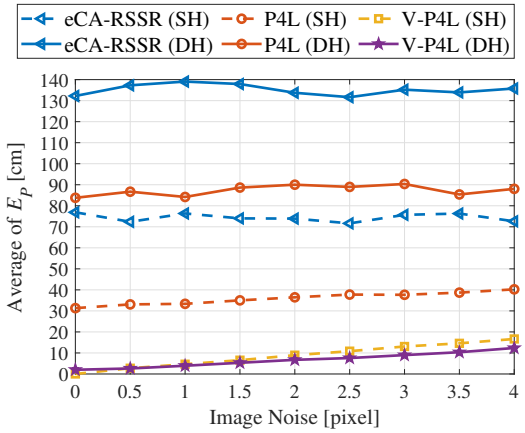


Fig. 7. Position errors (E_P) as the image noise varies.

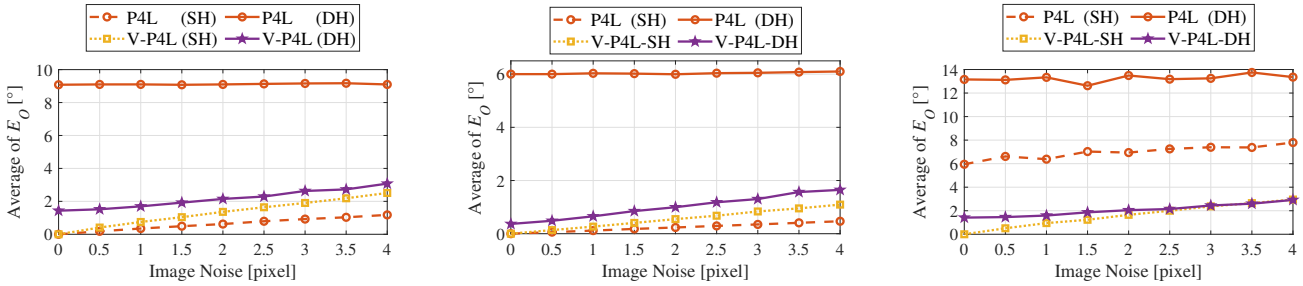
TABLE III
DEVICE SPECIFICATIONS.

Compoment/Paremeter		Model/Value
LED	Model	OSRAM GW PUSRA1.PM
	Semi-angle, $\Phi_{1/2}$	60°
	Transmitted Power, P_t	1-2 W
Camera	Model	TOSHIBA BU238M
	Intrinsic parameters	$f = 0.8$ cm
		$d_x = d_y = 5.84 \times 10^{-4}$ cm
		$(u_0, v_0) = (960, 600)$
	Resolution	1920×1200
	Exposure time	2.80 ms

C. Experimental Setup

An experimental prototype was set up to verify the efficiency of the proposed algorithm. The illustration of the experimental setup is shown in Fig. 9, and the specifications of devices are given in Table III.

At the transmitter side, a rectangular four-LED array mounted on the ceiling provides both illumination and location landmarks. Note that the proposed algorithm can be used for a single parallelogram luminaire, as well as point LED luminaires, and the models in these two applications are equivalent. The only difference between these two luminaire types is the type of the used LED luminaires. However, this has no effect on accuracy of V-P4L. Thus, we use a single



(a) Orientation errors along the X^c-axis. (b) Orientation errors along the Y^c-axis. (c) Orientation errors along the Z^c-axis.

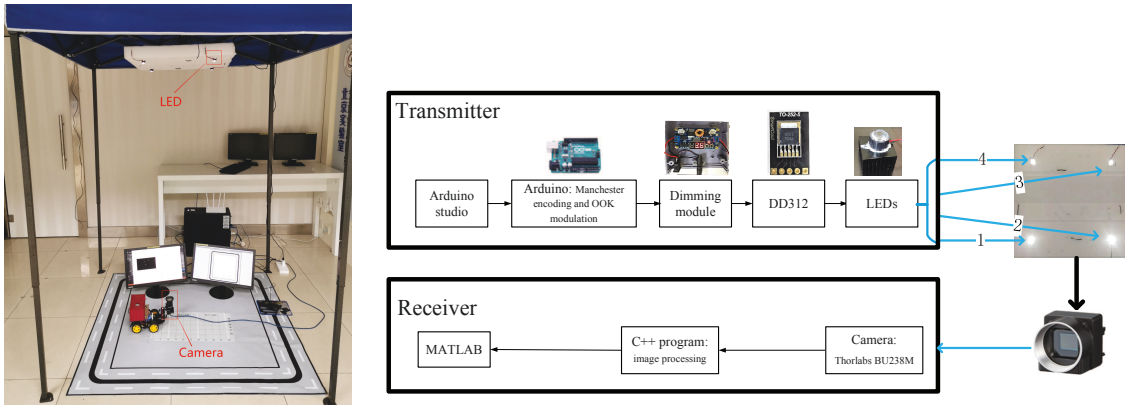
Fig. 8. Orientation errors (E_O) as the image noise varies.

parallelogram LED luminaire to introduce our system model without loss of generality. However, since we do not have a parallelogram luminaire, we use four point LED sources in experiment. The LED coordinates (in cm) are (0,0,175), (0,40,175), (40,40,175) and (40,0,175). One of the four LEDs transmits information which contains the coordinates of all LEDs. For the transmitter, Arduino Studio is first used to generate signals. Then, Arduino can achieve Manchester encoding and on-off keying (OOK) modulation. Manchester encoding can be employed to transform '0' to '01' and '1' to '10'. Then, the dimming module is used to control the transmitted power of the LED. Afterwards, a constant current LED driver, DD312, is used to avoid the flicking problem. Finally, the information of the transmitter can be broadcasted by the LED.

At the receiver side, a single camera is calibrated by a conventional method [35] to extract information from 2D images. Then, a proper exposure time is set to capture the LEDs. Once an image is obtained, image processing is implemented by a Visual C++ program to obtain the pixel coordinates of the LEDs' projections. The object detection is achieved by Hough Transform [39]. Afterwards, the pixel coordinates of the projections can be determined. The frame rate to scan the whole image is 165 fps. Then, to achieve high rate of data transmission, we exploit the region of interest function of this camera to scan the $1920\,\text{pixel} \times 60\,\text{pixel}$ regions which contain the projections of the LEDs with the frame rate of 1500 fps. Then, the coordinate information of the LEDs can be obtained by demodulating the images. Moreover, the pixel coordinates of the projections and the coordinate information of the LEDs are used as input for a MATLAB program. Finally, the camera position and pose can be estimated by implementing the subsequent processes in MATLAB.

D. Experimental Results

We evaluate the accuracy of the proposed algorithm through experiments. Since we do not have access to devices that can measure the accuracy of the rotation angles, we evaluate the position accuracy only. The proposed algorithm is implemented in an area of $50\,\text{cm} \times 50\,\text{cm} \times 175\,\text{cm}$. To verify the performance of the proposed algorithm, we estimate the position of 9 points when the receiver has random tilt angles. Figure 10 illustrates the localization results. As shown in Fig. 10, the red circles represent the positions of the reference points, while the blue crosses represent the positions of the estimated points. The 2D coordinates (in cm) of the reference points are $(5,5)$, $(5,25)$, $(5,45)$, $(25,5)$, $(25,25)$, $(25,45)$, $(45,5)$, $(45,25)$ and $(45,45)$. The height of the receiver is 27 cm. In particular, Fig. 10(a) shows the localization results of the basic V-P4L algorithm when LEDs have the same height. As shown in Fig. 10(a), the maximum position error is 4.56 cm, the minimum position error is 0.6 cm, and the average position error is 2.73 cm. Figure 10(b) shows the localization results of V-P4L-DH when LEDs have different heights. In this



(a) Testbed with a rectangular four-LED array. (b) Block diagram of experimental setup.

Fig. 9. Experimental setup of V-P4L.

experiment, we set the tilted angle of the LED plane is about 10°. We measure the tilted angle by gradienter. As shown in Fig. 10(b), the maximum position error is 4.42 cm, the minimum position error is 1.51 cm, and the average position error is 2.95 cm. From Fig. 10, we can observe that the position errors are less than 5 cm for both the center and the edges of the area regardless of the height differences among the LEDs. Therefore, the proposed V-P4L algorithm can achieve stable centimeter-level accuracy performance.

VII. Conclusion

In this paper, we have proposed a novel indoor localization algorithm named V-P4L that estimates the position and pose of the camera using a single, VLC-enabled LED luminaire. The camera is used to simultaneously capture the information in both time and space domains. Based on the information captured by the camera, V-P4L overcomes the limitation on the 3D-2D correspondences. Moreover, V-P4L can be implemented regardless of the height differences among LEDs. Therefore, V-P4L can achieve higher practicality and higher accuracy than eCA-RSSR and the conventional PnL algorithms. Simulation results have shown that for V-P4L the position error is always less than $15\,\mathrm{cm}$ and the orientation error is always less than 4° using popular indoor luminaires. Experimental results have shown that the average position error is less than 3 cm under both similar and different heights for the LEDs. Therefore, V-P4L is a promising indoor localization approach, which is suitable for common indoor scenarios.

References

[1] L. Bai, Y. Yang, C. Feng, C. Guo, and B. Jia, "Novel visible light communication assisted perspective-four-line algorithm for indoor localization," in *Proc. IEEE GLOBECOM (GC)*, Taipei, Taiwan, Jan. 2020, pp. 1–6.

[2] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE network*, vol. 34, no. 3, pp. 134–142, May/Jun. 2020.

[3] J. Luo, L. Fan, and H. Li, "Indoor positioning systems based on visible light communication: State of the art," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2871–2893, 4th Quart. 2017.

[4] Y. Zhuang, L. Hua, L. Qi, J. Yang, P. Cao, Y. Cao, Y. Wu, J. Thompson, and H. Haas, "A survey of positioning systems using visible led lights," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 1963–1988, 3rd Quart. 2018.

[5] M. F. Keskin, A. D. Sezer, and S. Gezici, "Localization via visible light systems," *Proc. IEEE*, vol. 106, no. 6, pp. 1063–1088, May 2018.

[6] Y. Yang, Z. Zeng, J. Cheng, C. Guo, and C. Feng, "A relay-assisted OFDM system for VLC uplink transmission," *IEEE Trans. Commun.*, vol. 67, no. 9, pp. 6268–6281, Sep. 2019.

[7] B. Zhou, A. Liu, and V. Lau, "Performance limits of visible light-based user position and orientation estimation using received signal strength under nlos propagation," *IEEE Trans. on Wireless Commun.*, vol. 18, no. 11, pp. 5227–5241, Nov. 2019.

[8] Y. Yang, Z. Zeng, J. Cheng, and C. Guo, "An enhanced DCO-OFDM scheme for dimming control in visible light communication systems," *IEEE Photon. J.*, vol. 8, no. 3, pp. 1–13, Jun. 2016.

[9] S.-Y. Hwang and J.-B. Song, "Monocular vision-based SLAM in indoor environment using corner, lamp, and door features from upward-looking camera," *IEEE Trans. Ind. Electron.*, vol. 58, no. 10, pp. 4804–4812, Oct. 2011.

[10] G. Feng, L. Ma, and X. Tan, "Line model-based drift estimation method for indoor monocular localization," in *Proc. 2018 IEEE VTC-Fall*, Chicago, IL, USA, Aug. 2018, pp. 1–5.

[11] C. Sertthin, T. Ohtsuki, and M. Nakagawa, "6-axis sensor assisted low complexity high accuracy-visible light communication based indoor positioning system," *IEICE Trans. Commun.*, vol. 93, no. 11, pp. 2879–2891, Nov. 2010.

[12] K. Qiu, F. Zhang, and L. Ming, "Let the light guide us: VLC-based localization," *IEEE Robot. Autom. Mag.*, vol. 23, no. 4, pp. 174–183, Dec. 2016.

[13] T. Q. Wang, Y. A. Sekercioglu, A. Neild, and J. Armstrong, "Position accuracy of time-of-arrival based ranging using visible light with application in indoor localization systems," *J. Lightw. Technol.*, vol. 31, no. 20, pp. 3302–3308, Oct. 2013.

[14] B. Zhu, J. Cheng, Y. Wang, J. Yan, and J. Wang, "Three-dimensional VLC positioning based on angle difference of arrival with arbitrary tilting angle of receiver," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 1, pp. 8–22, Jan. 2018.

[15] L. Bai, Y. Yang, C. Guo, C. Feng, and X. Xu, "Camera assisted received signal strength ratio algorithm for indoor visible light positioning," *IEEE Commun. Lett.*, vol. 23, no. 11, pp. 2022–2025, Nov. 2019.

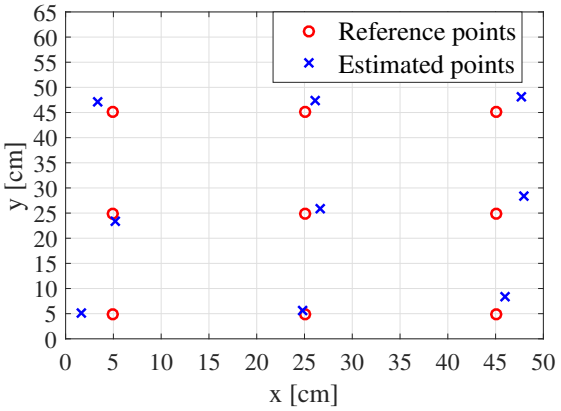
[16] L. Bai, Y. Yang, Z. Zhang, C. Feng, C. Guo, and J. Cheng, "A high-coverage camera assisted received signal strength ratio algorithm for indoor visible light positioning," *IEEE Trans. on Wireless Commun.*, pp. 1–1, Apr. 2021.

[17] B. Zhou, A. Liu, and V. Lau, "Joint user location and orientation estimation for visible light communication systems with unknown power emission," *IEEE Trans. on Wireless Commun.*, vol. 18, no. 11, pp. 5181–5195, Nov. 2019.

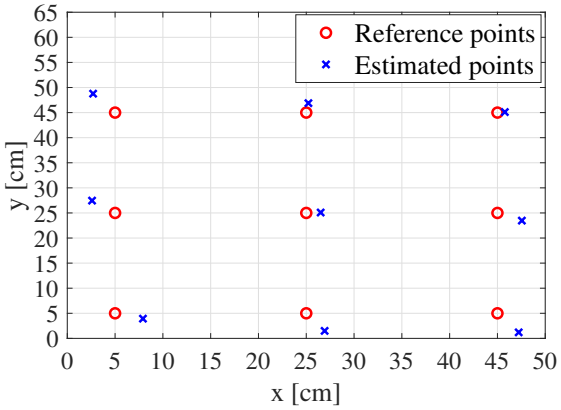
[18] Y. Li, Z. Ghassemlooy, X. Tang, B. Lin, and Y. Zhang, "A VLC smartphone camera based indoor positioning system," *IEEE Photon. Technol. Lett.*, vol. 30, no. 13, pp. 1171–1174, Jul. 2018.

[19] A.M. Vegni and M. Biagi, "An indoor localization algorithm in a small-cell led-based lighting system," in *Proc. IEEE Int. Conf. Indoor Position. Indoor Navig. (IPIN)*, Sydney, NSW, Australia, Nov. 2012, pp. 1–7.

[20] S.-H. Yang, H.-S. Kim, Y.-H. Son, and S.-K. Han, "Three-dimensional



(a) Localization results of V-P4L when LEDs have the same height.



(b) Localization results of V-P4L when LEDs have different heights.

Fig. 10. X-Y view of the localization results based on the proposed algorithm.

visible light indoor localization using aoa and rss with multiple optical receivers," *J. Lightw. Technol.*, vol. 32, no. 14, pp. 2480–2485, Jul. 2014.

[21] B. Zhu, J. Cheng, J. Yan, J. Wang, and Y. Wang, "Vlc positioning using cameras with unknown tilting angles," in *Proc. IEEE GLOBECOM (GC)*, Singapore.

[22] H. Zhao, J. Wang, and R. Liu, "High accuracy indoor visible light positioning considering the shapes of illuminators," in *Proc. IEEE Int. Symp. Broadband Multimedia Syst. Broadcast. (BMSB)*, Jeju, Korea (South), Jun. 2019, pp. 1–4.

[23] M. Biagi, S. Pergoloni, and A. M. Vegni, "LAST: A framework to localize, access, schedule, and transmit in indoor vlc systems," *J. Lightw. Technol.*, vol. 33, no. 9, pp. 1872–1887, May 2015.

[24] T.-H. Do and M. Yoo, "An in-depth survey of visible light communication based positioning systems," *Sensors*, vol. 16, no. 5, pp. 678, May 2016.

[25] Y.-S. Kuo, P. Pannuto, K.-J. Hsiao, and P. Dutta, "Luxapose: Indoor positioning with mobile phones and visible light," in *Proc. Annu. Int. Conf. Mobile Comput. Netw. (MobiCom)*, Maui, HI, USA., Sep. 2014, pp. 447–458.

[26] B. Zhou, V. Lau, Q. Chen, and Y. Cao, "Simultaneous positioning and orientating for visible light communications: Algorithm design and performance analysis," *IEEE Trans. Veh. Technol.*, vol. 67, no. 12, pp. 11790–11804, Dec. 2018.

[27] B. Zhou, A. Liu, and V. Lau, "Visible light-based user position, orientation and channel estimation using self-adaptive location-domain grid sampling," *IEEE Trans. on Wireless Commun.*, vol. 19, no. 7, pp. 5025–5039, Jul. 2020.

[28] F. Miramirkhani and M. Uysal, "Channel modeling and characterization for visible light communications," *IEEE Photonics J.*, vol. 7, no. 6, pp. 1–16, Dec. 2015.

[29] A. Cailean and M. Dimian, "Current challenges for visible light communications usage in vehicle applications: A survey," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2681–2703, 4th Quart. 2017.

[30] C. Xu, L. Zhang, L. Cheng, and R. Koch, "Pose estimation from line correspondences: A complete analysis and a series of solutions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1209–1222, Jun. 2017.

[31] B. Přibyl, P. Zemčík, and M. Čadík, "Absolute pose estimation from line correspondences using direct linear transformation," *Comput. Vis. Image Und.*, vol. 161, pp. 130–144, Aug. 2017.

[32] X. Wang, S. Pan, L. Qiu, L. S, and Z. Song, "Analytic algorithm of pose estimation based on two pairs of parallel lines," *Chin. J. Sci. Instrum.*, vol. 29, no. 3, pp. 600–604, Oct. 2008.

[33] T. Goto, S. Pathak, Y. Ji, H. Fujii, A. Yamashita, and H. Asama, "Line-based global localization of a spherical camera in manhattan worlds," in *Proc. IEEE Int. Conf. Robotics Automation (ICRA)*, Brisbane, QLD, Australia, May 2018, pp. 2296–2303.

[34] A. Masselli and A. Zell, "A new geometric approach for faster solving the perspective-three-point problem," in *Proc. Int. Conf. Pattern Recognit. (ICPR)*, Aug., Stockholm, Sweden 2014, pp. 2119–2124.

[35] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, pp. 1330–1334, Dec. 2000.

[36] L. Zhou, Y. Yang, M. Abello, and M. Kaess, "A robust and efficient algorithm for the PnL problem using algebraic distance to approximate the reprojection distance," in *Proc. AAAI Conf. Artif. Intell.*, Honolulu, Hawaii, USA, Jul. 2019, vol. 33, pp. 9307–9315.

[37] Arthur M. L., "On the calculation of Euler angles from a rotation matrix," *IEEE Trans. on Commun.*, vol. 17, no. 3, pp. 335–337, Aug. 1986.

[38] Philips, "LED-tubes," http://www.lighting.philips.com.cn/prof/led-lamps-and-systems/led-tubes#pfpath=0-LED_GR, 2020, [Online, accessed 8-Apr.-2020].

[39] D. H. Ballard, "Generalizing the hough transform to detect arbitrary shapes," *Pattern Recognition*, vol. 13, no. 2, pp. 111–122, Sept. 1981.