High-Precision Camera-enabled Visible Light Positioning System with Enhanced LED Recognition

Bugao Liu[1], Dingyu Ge[1], Zhigang Ma[1], Xiaodong Liu[1,*], Zhenghai Wang[1] and Xun Zhang[2]

[1]Nanchang University, 999 Xuefu Avenue, Honggutan District, Nanchang, 330031, Jiangxi, China
[2]Institut Supérieur D'électronique de Paris, 10 rue de vanves, Issy-les-moulineaux, 92130, France

Abstract

Visible light positioning (VLP) technology has attracted widespread research interest owing to its high-precision positioning, immunity to electromagnetic interference, abundant spectral resources, and cost-effectiveness. However, the region of interest (ROI) in existing camera-enabled VLP systems is highly susceptible to environmental noise, which degrades LED beacon recognition accuracy and complicates the simultaneous realization of precise positioning and stable lighting. To tackle the challenges, a novel indoor camera-enabled VLP systems is proposed. Specifically, a superglue-based tracking detection and dynamic positioning algorithm is proposed to replace the traditional LED area detection method relying on pixel intensity, thereby significantly enhancing the robustness of the VLP system while ensuring real-time performance and high accuracy. Moreover, a VLP testbench is developed to evaluate the positioning performance. Experimental results demonstrate that within a test area of $4 \times 4 \times 3\mathrm{m}^3$, 80% of the positioning errors are within 8.2cm. It indicates that the proposed system exhibits superior positioning accuracy compared to existing positioning methods.

Keywords

Visible light positioning, image sensor, region of interest, feature extraction, superglue

1. Introduction

With the rapid advancement of the Internet of Things (IoT) and the proliferation of smart devices, coupled with the fact that a significant proportion of human activities occur indoors, indoor positioning has emerged as one of the core enabling technologies, attracting increasing attention from both academia and industry [1]. In response to the growing demand for accurate and reliable indoor positioning, a variety of technologies have been developed. In this context, traditional positioning systems, including Bluetooth, Wireless Fidelity (Wi-Fi), and ultra-wideband (UWB) have exhibited considerable potential and broad applicability [2]. Nevertheless, these radio frequency (RF)-based positioning systems face challenges in balancing positioning accuracy with hardware cost and are susceptible to electromagnetic interference, particularly in scenarios involving dense terminal deployments in indoor environments [3]. Therefore, it is imperative to explore complementary technologies to address the growing demand for wide-area and high-precision positioning in complex indoor environments.

Visible light positioning has emerged as a promising solution by harnessing existing light emitting diode (LED) infrastructure and abundant visible light spectrum to support wide-area positioning services with cost-effective deployment, while its short wavelength enables centimeter-level positioning accuracy [4, 5]. Moreover, VLP is suitable for environments with severe RF interference or sensitive to electromagnetic interference due to the inherent electromagnetic interference immunity of visible light[6].

IPIN-WCAL 2025: Workshop for Computing & Advanced Localization at the Fifteenth International Conference on Indoor Positioning and Indoor Navigation, September 15–18, 2025, Tampere, Finland
*Corresponding author.
liubugao@email.ncu.edu.cn (B. Liu); Gedingyv@email.ncu.edu.cn (D. Ge); mazhigang@ncu.edu.cn (Z. Ma); xiaodongliu@ncu.edu.cn (X. Liu); zhenghai.wang@ncu.edu.cn (Z. Wang); xun.zhang@isep.fr (X. Zhang)
https://github.com/fskdnlkdsf (B. Liu); https://ieeexplore.ieee.org/author/837364500966238 (D. Ge); https://teacher.ncu.edu.cn/publish/mazhigang/ (Z. Ma); https://ieeexplore.ieee.org/author/37086532699 (X. Liu); https://ieeexplore.ieee.org/author/37088444786 (Z. Wang); https://perso.isep.fr/xzhang/ (X. Zhang)
0009-0005-6387-7517 (B. Liu); 0009-0000-7035-4443 (D. Ge); 0009-0007-8038-6298 (Z. Ma); 0000-0003-4112-9604 (X. Liu); 0009-0008-4507-5758 (Z. Wang); 0000-0002-8501-1969 (X. Zhang)

Compared to photodiode (PD)-based VLP systems, camera-based VLP systems offer several distinct advantages, including higher spatial resolution through detailed image capture, the ability to simultaneously track multiple targets without requiring array deployment, and the integration of rich environmental data for enhanced accuracy and robustness[7]. Although the camera-based positioning method demonstrates greater adaptability to ambient light interference compared to PD-based systems, it still exhibits significant limitations in dynamic environments, particularly in terms of insufficient robustness and suboptimal real-time positioning performance [8]. Considering this facts, the authors in [9] designed and implemented a robust and flexible indoor VLP system based on collaborative LEDs and edge computing, which addressed the failure of existing collaborative LED positioning algorithms under smartphone receiver rotation or tilt and achieved an effective balance between bandwidth and computing resources. Moreover, a loosely coupled fusion method integrating rolling shutter camera-enabled VLP with inertial data from an inertial measurement unit (IMU) was proposed to significantly enhance the localization robustness of VLP systems under adverse operational conditions, such as insufficiency LED transmission power, dynamic LED state switching, and ambient light interference [10]. Besides addressing the positioning errors induced by terminal angle rotation and noise interference in the aforementioned works, the recognition of regions of interest and the extraction of LED beacon identifiers (IDs) present another significant technical challenge. The authors proposed an image tracking algorithm based on threshold segmentation and geometric feature analysis of LED arrays. By constructing candidate ROIs and selecting the near-rectangular structure closest to the image center as the optimal ROI, the algorithm effectively mitigates natural light interference and achieves real-time and accurate tracking and positioning of LED arrays [11]. It should be noted that this approach requires the LED array to be deployed in a rectangular form. However, in many practical scenarios, the deployment of LEDs is irregular, which poses a significant challenge to the applicability of the method.

To address the aforementioned challenges, this paper proposes a superglue-based positioning method aimed at enhancing the anti-interference capability of indoor CMOS cameras-enabled VLP systems. The performance of the proposed system is validated through a VLP testbench. The main contributions of this paper are summarized as follows.

1) A novel dual-LED VLP method is proposed in this paper. Specifically, the camera-based receiver simultaneously captures the IDs of two LED beacons and estimates its own location information based on the LED location information embedded in the IDs. To accurately detect the ROI of LED, an improved superglue graph neural network (GNN) algorithm is introduced, replacing the conventional threshold-based ROI detection method to enhance LED recognition accuracy. Meanwhile, an adaptive threshold mechanism is designed to filter out high-intensity pixel clusters in the binary image. Moreover, ambient light interference is mitigated through bidirectional row and column edge detection combined with inter-row pixel difference analysis.
2) A semi-physical VLP experimental testbed is constructed. Specifically, the LED beacon employs Manchester coding and on-off keying (OOK) modulation to transmit ID information, while the CMOS-based receiver is mounted on a two-dimensional mobile platform. Experimental results demonstrate that the detection performance of the proposed modified superglue detection method is significantly better than the baseline solution. Based on this, the positioning performance of the VLP system proposed in this paper can reach the centimeter level, 80% of the positioning errors are within $8.2\mathrm{cm}$ in a $4 \times 4 \times 3\mathrm{m}^3$ test area.

The remainder of this paper is organized as follows. Section II presents the VLP system framework. The superglue-based positioning method is proposed in Section III. Section IV details the implementation of FPGA-based VLP testbench and provides comprehensive analysis of positioning performance evaluation. Finally, Section V concludes the paper.

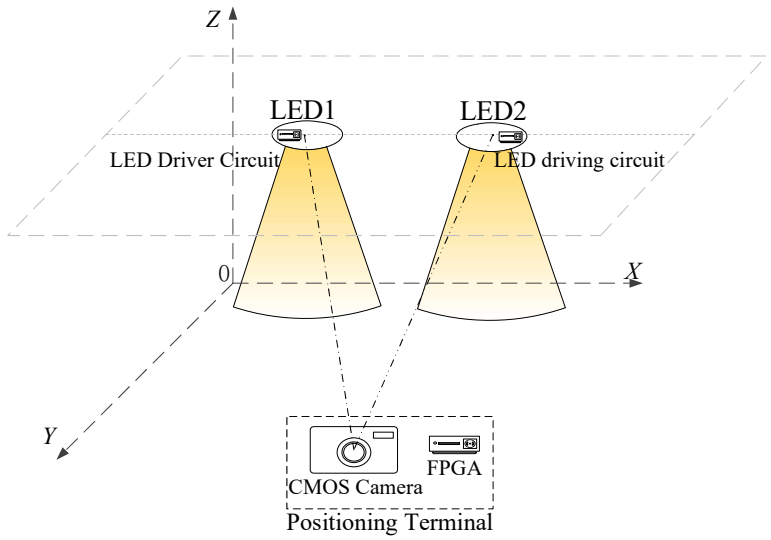

Figure 1: The considered VLP system model.

2. System Model and Positioning Principle

2.1. System Model

In a typical indoor environment, VLP systems must ensure uniform lighting while providing high-precision positioning services . To achieve this, an even number of lamps are often deployed uniformly indoors. As shown in Fig. 1, two LED lamps, serving as positioning transmitting beacons, are evenly distributed in the considered VLP system. It is important to note that the three-dimensional (3D) coordinates of the LED beacon corresponding to each ID are unique. At the receiving end, a common CMOS image sensor is employed to capture beacon information. This system requires only two LED beacons for 3D positioning. By simultaneously capturing their signals, the receiver's location is estimated based on the known beacon positions (c.f. ID) and the tranceiver distance.

2.2. Positioning principle

Different from conventional positioning methods that rely on three or more LED beacons for position estimation, the proposed VLP system achieves accurate positioning using only two LED beacons. The principle of the proposed VLP system is elucidated through geometric analysis.

As illustrated in Fig. 2, the world coordinates of the two LED beacons in the proposed VLP system are defined as (x_1, y_1, z_1) and (x_2, y_2, z_2). Their projected coordinates $P_k(i_k, j_k), k = 1, 2$ on the CMOS-based receiver plane are obtained through the perspective projection of the LED optical centers via the imaging lens. For deployment cost and aesthetic considerations, the LED beacons are typically installed at the same horizontal height. In other words, the Z-axis coordinate of the LED beacons in the world coordinate system is same, i.e., $(z_1 = z_2)$. Therefore, the Euclidean distance d_{LED} between two LED beacons in the world coordinate system can be determined as follows.

$$d_{\mathrm{LED}} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}. \tag{1}$$

It is assumed that the center point of the camera lens, denoted (x_p, y_p, z_p) and referred to as 'lens' in Fig. 2, serves as the receiver end location. Moreover, the projection coordinates of the 'lens' on the LED beacon plane and the imaging plane are defined as $A(x_p, y_p, z_p)$ and $A^{'}(0, 0)$, respectively. Note that the coordinates $A^{'}$ is the the pixel coordinates and is set to $(0, 0)$ for simplicity. Thus, the distances d_{kp} from point A to the two LED beacons are expressed as

$$d_{A,k} = \sqrt{(x_p - x_k)^2 + (y_p - y_k)^2}, \tag{2}$$

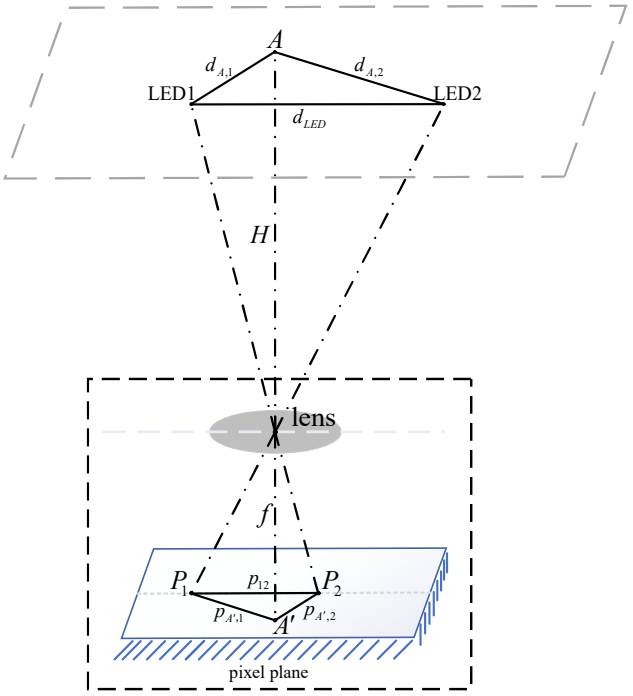

Figure 2: The principle of the proposed VLP method.

where $k = 1, 2$ is the k-th LED beacon. Moreover, the distance from the projection point of the LED beacon to the centroid $A^{'}$ of the LED pixel area can be calculated as

$$p_{A',k} = \sqrt{i_k^2 + j_k^2}. \tag{3}$$

Then, the Euclidean distance between the two LED beacons projected in the imaging plane is given as

$$p_{12} = \sqrt{(i_1 - i_2)^2 + (j_1 - j_2)^2}. \tag{4}$$

Since the plane between the image sensor and the LED beacon is parallel, the vertical height H between the image sensor and the LED beacon can be derived using the principle of similar triangles as

$$H = f \times \frac{d_{\text{LED}}}{p_{12}}, \tag{5}$$

Thus, the z_p of the receiver terminal coordinate can be determined as $z_p = z_1 - H$. Given that the imaging plane and the world coordinate system are parallel with coincident axis orientations, the following relationship can be derived based on the geometric similarity of triangles

$$\frac{d_{A,k}}{H} = \frac{p_{A',k}}{f}. \tag{6}$$

The values x_p and y_p of can be solved based on the aforementioned geometric relationships (2) .

3. Superglue-enabled ROI Detection Method

Based on the positioning principle described in the previous section, the proposed VLP method lies in accurately finding the projection points P_1 and P_2 of the LED on the imaging plane. However, since the signal power attenuation is inversely proportional to the fourth power of distance, the presence of ambient light and noise further degrades the beacon detection performance in the imaging system.

Therefore, it is essential to design an ROI detection algorithm for the LED beacons. Considering that the illumination of the VLP system varies with environmental conditions, and some image frames captured by the camera are blurred due to human movement or shadow effects, the superglue-based graph neural network (GNN) is introduced to tackle these issue. It employs an attention mechanism to learn illumination-invariant features and utilizes its global context modeling capacity to match LED with time series or multi-view information to solve the problem of single-frame image blur[12]. In this context, since superglue is fundamentally a general feature matching network, it requires specific modifications to achieve efficient ROI detection of LEDs. Specifically, the LED beacons exhibit significantly higher pixel intensity in camera-captured images, the algorithm can leverage these high-intensity pixels as primary feature matching indicators. This approach enables targeted feature matching for LED detection while maintaining real-time performance.

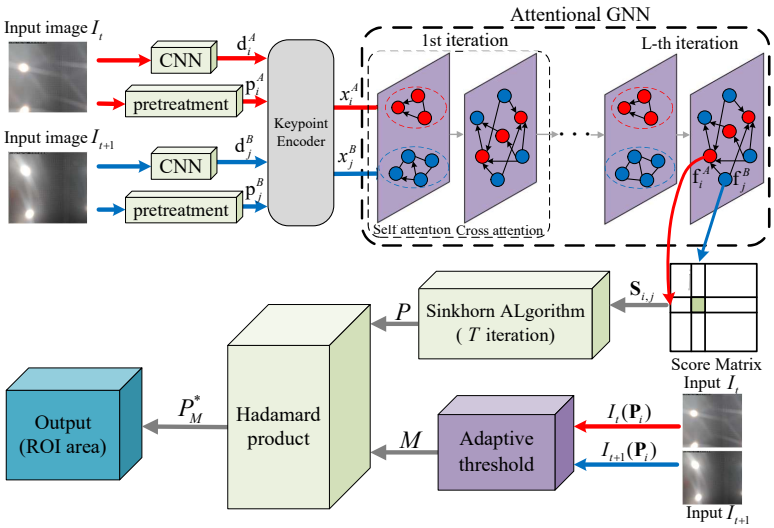

Figure 3: Superglue-based LED'S ROI detection

The basic principle of the proposed method is that the key points within image typically represent projections of salient 3D points, and the ROI detection of LEDs is achieved by partially matching local features between the two images. In other words, a valid LED beacon is successfully detected if region of interest (ROI) area matching degree between the two images is the highest. Therefore, ROI detection can be formulated as an optimization problem, which involves predicting the matching matrix $\mathbf{P}$ from two sets of local features using a GNN-based model. In this context, the structure of the modified superglue-based GNN feature matching method is shown in Fig. 3.

The model inputs two consecutive images A and B, and then outputs a set of key point positions p_i and their associated visual descriptors d_i through image preprocessing and CNN extraction, where $p_i \triangleq (x; y; c)_i$ consists of image coordinates (x, y) and detection confidence c. Images A and B can obtain M and N sets of local features $(p_i; d_i)$, respectively. Then, a keypoint encoder sub-model is employed to map keypoint positions p_i and their visual descriptors d_i into a single vector $\mathbf{x}_i$. Moreover, these vectors are input into the attention GNN network consisting of L layers of self-attention and cross-attention modules to simultaneously capture the spatial relationship and visual features of key points and to obtain a more powerful matching descriptors f_i. Note that the matching process is formulated as:

$$\begin{aligned} f_i^A &= F\left(\mathbf{x}_i\right), \\ f_j^B &= F\left(\mathbf{x}_j\right), \end{aligned} \tag{7}$$

where $F(\cdot)$ represents function of the attention GNN. Based on these f_i^{A} and f_i^{B}, a score matrix is constructed and the relationship is described as follows.

$$\mathbf{S}_{i,j} = \langle f_i^A, f_j^B \rangle, \quad \forall (i,j) \in M \times N, \tag{8}$$

where $\langle \cdot, \cdot \rangle$ is the inner product. Note that the magnitude of the matching descriptors varies based on the features and is adjusted during training to reflect the prediction confidence. Then, the optimal matching matrix $\mathbf{P}$ is obtained via differentiable optimal transport and is given as

$$\mathbf{P} = \underset{\mathbf{P} \in \mathbb{P}}{\operatorname{argmin}} \sum_{i,j} S_{i,j} p_{i,j} + \lambda H(\mathbf{P}), \tag{9a}$$

$$H(\mathbf{P}) = -\sum_{i,j} p_{i,j} (\log P_{i,j} - 1), \tag{9b}$$

where $\mathbb{P}$ represents the set of matching matrices. $\lambda H(\mathbf{P})$ represents the entropy regularization term, which smooths the optimization problem and mitigates overfitting to local optimal solutions. It is solved by the Sinkhorn algorithm.

To address ambient lighting variations and improve the robustness of LED detection, a dynamic adaptive threshold γ_{auto} is introduced to replace the fixed intensity threshold and it is given as

$$\gamma_{\text{auto}} = \max(\mu_S + 3\sigma_S,\, 230) \tag{10}$$

where μ_S represents mean intensity of S and σ_S is standard deviation of S. Note that the $S = A(p_i) \cup B(p_j)$ is the union of pixel intensities from consecutive images A and B. Moreover, the constant 230 serves as an empirical lower bound to ensure detectability in extreme low-light scenarios. Then, the subset $\mathcal{M}$ of features greater than the feature threshold γ_{auto} is given as

$$\mathcal{M} = \left\{ (i,j) \mid \frac{A(\mathbf{p}_i) + B(\mathbf{p}_j)}{2} \geq \tau_{\text{auto}} \right\} \tag{11}$$

Finally, the match results are obtained by

$$P^*_{\mathcal{M}} = P^* \odot M, \quad M_{i,j} = \begin{cases} 1 & (i,j) \in \mathcal{M} \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

where $\odot$ resprents the Hadamard product.

4. Experimental Testbed and Result Analysis

4.1. Semi-physical experimental testbed

As shown in Fig. 4, a semi-physical experimental testbed is constructed in a typical indoor scenario with a size of $4 \times 4 \times 3\text{m}^3$ scenario to validate the effectiveness of the proposed positioning method. Specifically, two commercial LED emitters are installed on the ceiling with a 2.5m spacing centered in the room. To ensure uniform lighting and stable signal transmission, the transmitter employs Manchester coding and OOK modulation technology to encode and modulate the beacon ID. The signal operates within the $1\tilde{}10\text{kHz}$ frequency range, and the LED beacon is driven by a constant-current source. The modulated optical signal is captured by an OV5640 CMOS camera (operating at 60 fps with 640×480 resolution) mounted on a ground robotic platform, with the camera constrained to an unwavering vertical orientation facing upward. The acquired data undergoes offline MATLAB processing for feature extraction and precision position estimation.

4.2. Results analysis

In order to validate the effectiveness and superiority of the proposed ROI detection method, the receiver is located directly below one LED beacon and the performance of ROI detection methods is firstly compared. Specifically, Fig. 5 illustrates the detection accuracy of the proposed scheme, Gaussian enhanced (GE) detection, and conventional regional pixel threshold segmentation (RPTS) detection scheme at varying transceiver heights. It is apparent from Fig. 5 that the ROI detection

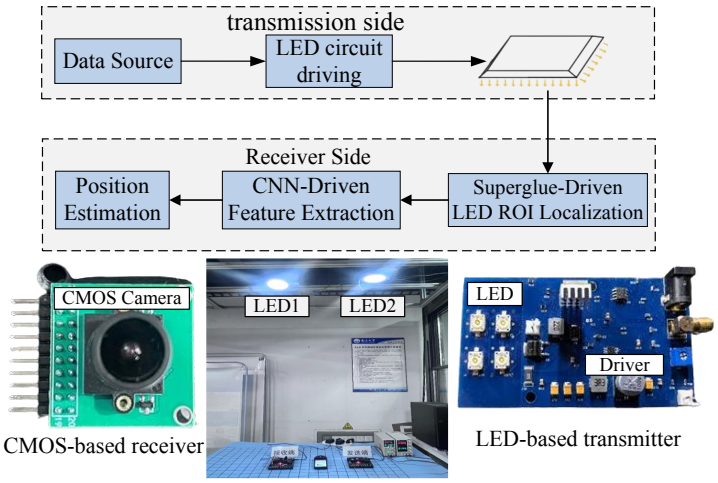

Figure 4: The structure of the semi-physical experimental testbed.

accuracy decreases as the transceiver distance increases. This is attributed to the more severe distance attenuation at greater distances, which degrades the image quality. Moreover, it can be seen that the proposed scheme significantly outperforms the two baseline schemes, achieving a detection accuracy exceeding 80% at a distance of $2.5\mathrm{m}$, which is a more common indoor positioning distance. It is worth noting that the proposed method could support longer distances if a higher-performance camera is utilized.

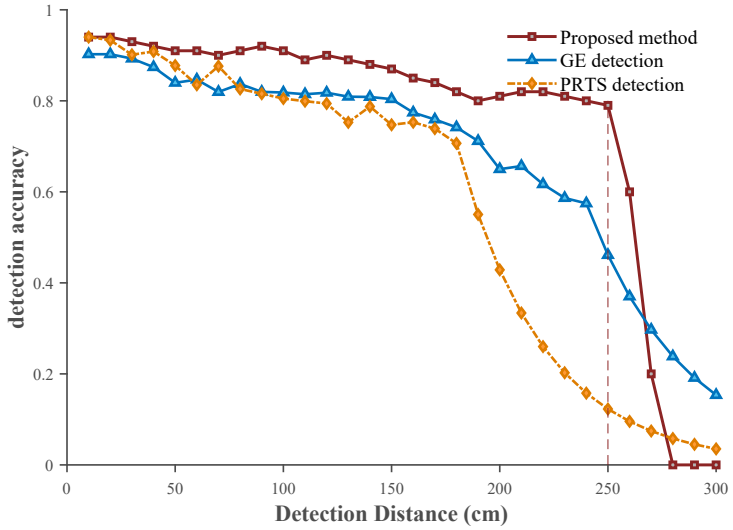

Figure 5: The ROI detection accuracy versus different detection distance via different methods

Secondly, beacon signals are collected at 256 evenly distributed grid points within the indoor area. To ensure reliability of the result, each location is measured 20 times(400ms per trial) and the average value is calculated, resulting in a total of 5120 measurement instances for analysis. As shown in Fig. 5, the positioning results of 256 discrete measurement points under the aforementioned setup are compared with the real position, and a two-dimensional positioning error density map smoothed by a Gaussian kernel is utilized to quantify the spatial error characteristics. It is evident from Fig. 5 that the positioning results near the center of the area closely match the real position, while those in the edge regions exhibit certain deviations, and the worst-case error reaching approximately 0.08 meters. This is due to the limited field of view of the transceiver device, which results in partial image distortion captured by the camera at the edge point. However, the overall positioning performance achieves centimeter-level

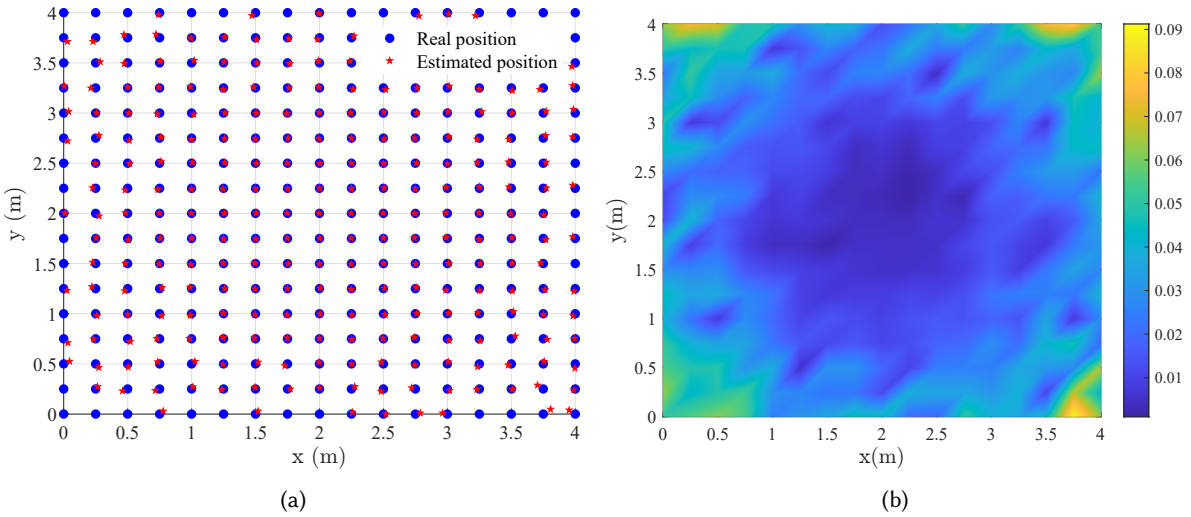

Figure 6: The (a) comparison of measurement point positioning results with real position and (b) two-dimensional positioning error density distribution.

accuracy.

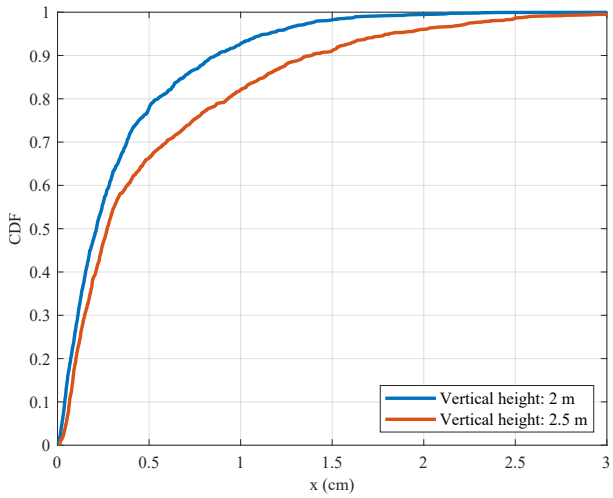

Figure 7: CDF curves of positioning errors in different vertical height scenarios

Finally, the cumulative distribution function (CDF) of positioning errors at two different LED beacon heights are compared and analyzed to accurately evaluate the positioning performance of the proposed VLP system. Fig. 7 illustrates the CDF curves of positioning errors at vertical heights of $2.0\mathrm{m}$ and $2.5\mathrm{m}$. As depicted in the figure, the positioning error of the VLP system increases with height. Specifically, the proposed system achieves a positioning error of $5.2\mathrm{cm}$ at an 80% confidence level for a height of $2.0\mathrm{m}$ outperforming the $8.2\mathrm{cm}$ error observed at a height of $2.5\mathrm{m}$. Notably, the proposed positioning method achieves centimeter-level positioning accuracy using only two LED beacons.

5. Conclusion

This study proposes a high-precision indoor positioning method based on SuperGlue. The method requires the camera to capture only two LED beacons and performs positioning estimation using an ID-geolocation database and geometric relationships. To address the challenge of LED detection and recognition being susceptible to ambient light interference, a GNN-based modified superglue method is designed to accurately detect the ROI of LED beacons in the image. Moreover, an adaptive threshold

detection module is introduced to enhance the robustness of the method. A semi-physical VLP testbed is constructed to evaluate the effectiveness of the proposed approach. Experimental evaluations conducted in a typical indoor scenario area demonstrate that the proposed method achieves significantly more accurate ROI detection compared to the baseline solution. Furthermore, 80% of the positioning errors of the proposed VLP system are within 8.2cm.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (Grant 62561039 and 62161023); in part by the Young Natural Science Foundation of Jiangxi Province (Grant 20224BAB212004); and in part by the University-Industry Collaborative Education Program (Grant 231002455265614).

Declaration on Generative AI

This paper have not employed any Generative AI tools.

References

[1] J. Singh, N. Tyagi, S. Singh, F. Ali, and D. Kwak, "A systematic review of contemporary indoor positioning systems: Taxonomy, techniques, and algorithms," *IEEE Internet Things J.*, vol. 11, no. 27, pp. 34717-34733, Nov. 2024.

[2] P. S. Farahsari, A. Farahzadi, J. Rezazadeh, and A. Bagheri, "A survey on indoor positioning systems for IoT-based applications," *IEEE Internet Things J.*, vol. 9, no. 10, pp. 7680-7699, May. 2022.

[3] S. Bastiaens, M. Alijani, W. Joseph, and D. Plets, "Visible light positioning as a next-generation indoor positioning technology: A Tutorial," *IEEE Commun. Surveys Tut.*, vol. 26, no. 4, pp. 2867-2913, 2024.

[4] Y. Wei, X. Liu, Y. Wang, S. Ma, F. Zhou, and O. A. Dobre, "Channel and location estimation enabled a novel BDCNP network for massive MIMO VLCP systems," *IEEE Wireless Commun. Lett.*, vol. 13, no. 1, pp. 218-222, Jan. 2024.

[5] Z. Wang, X. Huang, X. Chen, M. Xu, X. Liu, Y. Wang, and X. Zhang, "PoE-enabled visible light positioning network with low bandwidth requirement and high precision pulse reconstruction," *IEEE J. Indoor and Seamless Positioning and Navigation*, vol. 2, pp. 1-11, Jan. 2024.

[6] X. Liu, Z. Chen, Y. Wang, F. Zhou, Y. Luo, and R. Q. Hu, "BER analysis of NOMA-Enabled visible light communication systems with different modulations," *IEEE Trans. Veh. Technol.*, vol. 68, no. 11, pp. 10807-10821, Nov. 2019.

[7] J. Chen, D. Zeng, C. Yang, and W. Guan, "High accuracy, 6-DoF simultaneous localization and calibration using visible light positioning," *J. Lightw. Technol.*, vol. 40, no. 21, pp. 7039-7047, Nov. 2022.

[8] X. Yang, Y. Zhuang, M. Shi, X. Sun, X. Cao, and B. Zhou, "RatioVLP: Ambient light noise evaluation and suppression in the visible light positioning system," *IEEE Trans. Mobile Comput.*, vol. 23, no. 5, pp. 5755-5769, May. 2024.

[9] X. Liu, L. Guo, H. Yang, and X. Wei, "Visible light positioning based on collaborative LEDs and edge computing," *IEEE Trans. Comput. Social Syst.*, vol. 9, no. 1, pp. 324-335, Feb. 2022.

[10] W. Guan, L. Huang, B. Hussain, and C. P. Yue, "Robust robotic localization using visible light positioning and inertial fusion," *IEEE Sensors J.*, vol. 22, no. 6, pp. 4882-4892, Mar. 2022.

[11] L. Chen, L. Feng, J. Wang, T. Zhang, H. Zhang, and Z. Xue, "Hybrid indoor positioning system based on LED array," *IEEE Photon. J.*, vol. 17, no. 2, pp. 1-11, Apr. 2025.

[12] P. -E. Sarlin, D. DeTone, T. Malisiewicz, and A. Rabinovich, "SuperGlue: Learning feature matching with graph neural networks," in *2020 IEEE/CVF Conf. on Comput. Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, 2020, pp. 4937-4946.