Simultaneous Position and Orientation Estimation in Optical Camera Communication Based Indoor Localization System

Yitong Wang
School of Engineering
RMIT University
Melbourne, Australia
s3587959@student.rmit.edu.au

Kandeepan Sithamparanathan
School of Engineering
RMIT University
Melbourne, Australia
kandeepan.sithamparanathan@rmit.edu.au

Ke Wang
School of Engineering
RMIT University
Melbourne, Australia
ke.wang@rmit.edu.au

***Abstract*—The simultaneous position and orientation estimation is achieved in an optical camera communication-based indoor localization system. By employing the dual camera receiver and CNN-based processor, an average error of 0.0571m and 0.5038° is achieved respectively.**

***Index Terms*—optical camera communication, indoor localization, dual camera receiver, convolutional neural network**

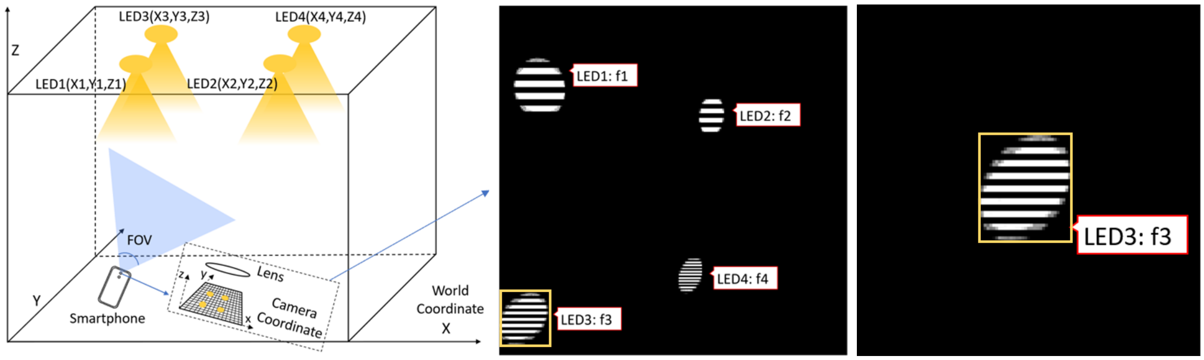
I. Introduction

Recently, optical wireless communication (OWC) has attracted significant interest in both industry and academia. As a technology exploring the optical band, it can effectively relieve the pressure on the highly congested RF spectrum [1]. Among different OWC technologies, optical camera communication (OCC) adopts the off-the-shelf light emitting diode (LED) as transmitter and camera as receiver, which is more cost effective and practical. OCC can realize up to tens of megabit-per-second (Mbps) wireless communication and provide a range of indoor applications including internet of things (IoT), indoor navigation and smart healthcare [2]. In addition to the communication capability, OCC can also provide indoor localization function, where centimeter-level accuracy has been achieved in typical indoor environment [3]. However, most recent OCC localization studies assume the receiver faces straight upward to the ceiling and ignore the critical tilt angle [4]. In reality, the orientation of the receiver in OCC systems has a great impact on the imaging position of the transmitter in the field of view (FOV), and hence, it fundamentally limits the indoor positioning system's accuracy. On the other hand, whilst stereo vision based method can provide orientation estimation, it requires at least three LEDs captured in each camera [5]. Due to the finite FOV of receiver, capturing at least three LEDs is challenging, limiting the working area of the method.

In this work, we propose for the first time a joint position and orientation estimation scheme with high precision in OCC system leveraging the dual camera receiver and CNN. Nowadays, the availability of dual camera receiver on smartphones provides a benefit on OCC based localization to estimate 3D information by obtaining depth information through disparity [2]. CNN can help to extract features from the captured images to determine the 6-degree of freedom (DOF) position and orientation information for localization. Results show that the proposed method can realize an average position and orientation estimation error of 0.0571m and 0.5038° respectively in a room with a size of 5m×5m×3m. In addition, even if less than three LEDs are captured, results show that the simultaneous position and orientation estimation can still be achieved at slightly worse precision. Hence, the proposed scheme can also improve the coverage area in practical applications.

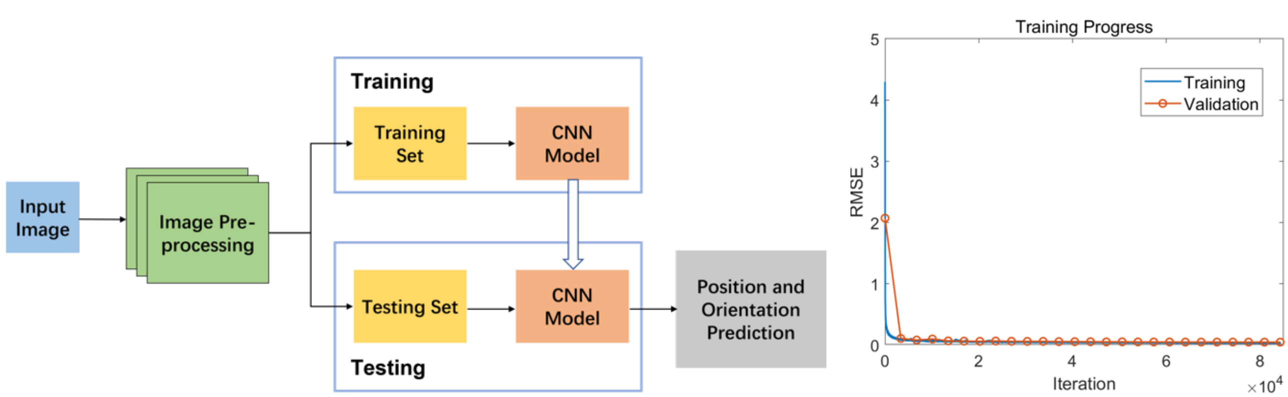
II. OCC-Based Indoor Localization System

Our proposed OCC based indoor localization system is shown in Fig.1a. Four LEDs are mounted on the ceiling of a room with a size of 5m×5m×3m. Each LED transmits the unique LED-IDs (location coordinates) modulated at different frequencies. The dual cameras on a smartphone with rolling shutter mode are used as the receiver to capture images. By comparing the disparity of corresponding projection point from each camera, the distance from the LED to the receiver can be deducted, providing additional information to achieve 3D localization. The superposition image from dual camera receiver as shown in Fig.1b contains rich information as well, including: (*i*) the LED-IDs are represented by the width of bright and dark stripes on LEDs' projection; (*ii*) the position and ellipse shape of LEDs' projection on captured images reflect geometric relationship between the transmitters and the receiver, including the rotation information; (*iii*) the size and intensity distribution of the projection areas among pixels, as well as the disparity from dual cameras provide the estimation of the distance between the transmitter and the receiver. In Fig.1c, the LED3 is zoomed in to clearly show the details mentioned above. To effectively explore these features from



(a) The localization system setup (b) The superposition image (c) The zoomed in image of LED3

Fig. 1: The proposed localization system: a) the localization system setup, b) the captured image by dual camera receiver, c) the zoomed in image of LED3



(a) The proposed CNN model (b) The training progress

Fig. 2: Overview of the proposed CNN model: a) the proposed CNN model, b) the training progress

images, we propose a novel scheme by combining dual camera receiver and CNN for simultaneous position and orientation estimation.

However, since we need to obtain information in 6-dimensions (3D location and rotation angles in 3 dimensions) information, training such a network effectively is costly and time consuming. To solve this issue, here we adopt the transfer learning principle, which converges rapidly during the training process and performs better by leveraging existing knowledge such as features and weights from a pre-trained neural network. GoogLeNet [6] is a 22-layer deep CNN for image classification. In our CNN model, we have to modify the structure of GoogLeNet trained on the ImageNet dataset to fit our regression purpose. The proposed CNN model and the training progress are shown in Fig.2. For the input as shown in Fig.2a, each input instance consists of three 224×224 grayscale images, which are the two images from the two cameras and the superposition image. We also convert the 3D orientation to 4D quaternion due to its better performance on regression [7], leading to a 7D output (3D position and 4D quaternion). In addition, the dropout layer and three softmax classifiers in the original GoogLeNet are replaced by a 7D fully connected layer and regression layer in our model, to generate the 7D output. We use a dataset of size 40000 and use 90% as the training set. The loss function is mean square error (MSE). A batch size of 128 and the Adam optimizer are used for training. We also choose an initial learning rate of 0.0001 with a 0.8 drop factor per 20 epochs. There is no overfitting observed during the training process, as shown in Fig.2b.

III. Results and Discussion

Fig.3 and Fig.4 show the cumulative distribution functions (CDFs) of position and orientation estimation errors on the test set. It can be seen that 90% of position errors in each direction are smaller than 0.0630m, and 90% of orientation errors in each angle are less than 0.85°. We also calculate the average Euclidean distance and quaternion angle error, which is 0.0571m and 0.5038° respectively. To the best of our knowledge, this is the first time achieving high accuracy simultaneous estimation of position and orientation in OCC based indoor localization system. In addition, we can observe that the accuracy on Z direction is better than the other two directions, which is due to the fact that rich distance information is available from LED's projection, leading to more accurate estimation of the height information.

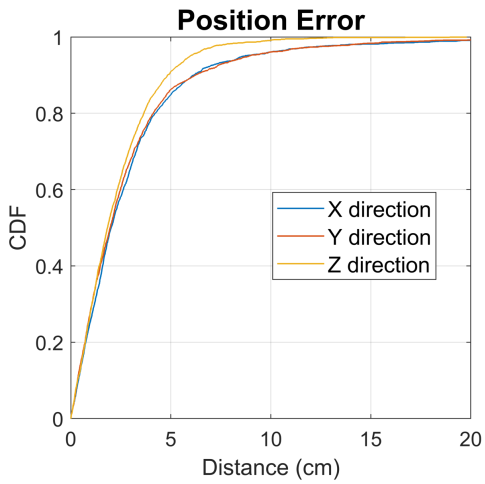


Fig. 3: CDF of position estimation error

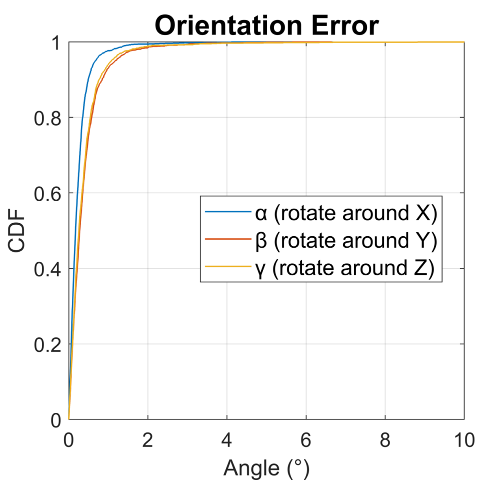


Fig. 4: CDF of orientation estimation error

Furthermore, we compare the mean errors when less than three LEDs are captured in the images in Table.I. It is clear that very similar performance on each dimension is achieved for both cases with two LEDs captured and three LEDs captured, which is also comparable to the performance of the complete test set. Even when only one LED is captured, simultaneous position and orientation estimation can still be achieved with slightly worse accuracy. It is demonstrated that our proposed method can reduce the number of LEDs needed in captured images, thereby expanding the coverage area of the localization system.

TABLE I: Mean Error Comparison

	X(m)	Y(m)	Z(m)	$\alpha(°)$	$\beta(°)$	$\gamma(°)$
Test Set	0.0313	0.0296	0.0233	0.2489	0.4007	0.3889
One LED	0.0972	0.1169	0.0387	3.0745	2.9863	8.8484
Two LEDs	0.0294	0.0307	0.0234	0.9597	0.8150	1.2319
Three LEDs	0.0273	0.0273	0.0235	0.8235	0.6708	0.9303

IV. Conclusion

In this paper, a novel OCC based indoor localization scheme to jointly estimate position and orientation has been proposed and demonstrated. By combining the dual camera receiver and CNN, an average position and orientation estimation error of 0.0571m and 0.5038° is achieved, respectively. In addition, results also show that similar position and orientation estimation errors have been achieved even when less then three transmitters are captured, and hence, the coverage area of localization system has been improved effectively.

References

[1] Haas Harald, Elmirghani Jaafar and White Ian 2020 Optical wireless communicationPhil. Trans. R. Soc. A.378:20200051.20200051

[2] W. A. Cahyadi, Y. H. Chung, Z. Ghassemlooy, and N. B. Hassan, ”Optical Camera Communications: Principles, Modulations, Potential and Challenges,” Electronics, vol. 9, no. 9.

[3] J. Luo, L. Fan and H. Li, ”Indoor Positioning Systems Based on Visible Light Communication: State of the Art,” in IEEE Communications Surveys & Tutorials, vol. 19, no. 4, pp. 2871-2893, Fourthquarter 2017.

[4] Y. Xie, Y. Wang, S. Kandeepan, and K. Wang, ”Machine Learning Applications for Short Reach Optical Communication,” Photonics, vol. 9, no. 1, p. 30, 2022. [Online]. Available: https://www.mdpi.com/2304-6732/9/1/30.

[5] Z. Zhu, Y. Yang, C. Guo, C. Feng and B. Jia, ”Visible Light Communication Assisted Perspective Circle and Arc Algorithm for Indoor Positioning,” ICC 2022 - IEEE International Conference on Communications, Seoul, Korea, Republic of, 2022, pp. 3832-3837.

[6] Szegedy, Christian & Liu, Wei & Jia, Yangqing & Sermanet, Pierre & Reed, Scott & Anguelov, Dragomir & Erhan, Dumitru & Vanhoucke, Vincent & Rabinovich, Andrew. (2014). Going Deeper with Convolutions.

[7] T. Hu and Q. Liao, ”Real-Time Camera Localization with Deep Learning and Sensor Fusion,” ICC 2021 - IEEE International Conference on Communications, Montreal, QC, Canada, 2021, pp. 1-7.