Indoor Positioning by Deep Q-Network in VLC Environment

Sung Hyun Oh and Jeong Gon Kim*

Department of Electronic Engineering, Tech University of Korea, Si-heung, Republic of Korea
Email: osh119@tukorea.ac.kr (S.H.O.); jgkim@tukorea.ac.kr (J.G.K.)
*Corresponding author

***Abstract*—With the recent development of the Fourth Industrial Revolution, Internet of Things technology has been widely adopted. In addition, key technologies such as big data, artificial intelligence, and wireless communication are being combined. Positioning technology that uses these technologies is essential for locating human devices in modern industries. Although the Global Positioning System can provide relatively precise positioning outdoors, its performance is limited indoors due to propagation loss. Hence, various wireless signal-based indoor positioning technologies, such as WiFi, Bluetooth, ultra-wideband, and Visible Light Communication (VLC) are being studied. In this study, positioning in indoor VLC environments is analyzed using Deep Q-Network (DQN). Each element of reinforcement learning and the agent's action and reward function are set to increase positioning accuracy. Deep Q-Network (DQN) training is then performed to derive positioning performance. The simulation results show that the proposed model attains a positioning resolution of less than 15 cm and achieves a processing speed of less than 0.03 seconds to obtain the final position in the Visible Light Communication (VLC) environment.**

***Keywords*—Indoor Positioning System (IPS), Artificial Intelligence (AI), Deep Q-Network (DQN), Received Signal Strength (RSS), Visible Light Communication (VLC)**

I. Introduction

With the continued development of the Fourth Industrial Revolution, Internet of Things (IoT) technology is advancing rapidly. A combination of IoT, big data, Artificial Intelligence (AI), and wireless communication technologies has led to the emergence of Location-Based Services (LBS) in various fields, such as public services, daily life, commerce, and manufacturing processes. Although the Global Positioning System (GPS) provides relatively precise accuracy for outdoor positioning [1], its use for indoor positioning is limited due to challenges such as signal attenuation, multipath effects [2], and the absence of a clear Line of Sight (LOS) [3]. To overcome these limitations, researchers are exploring various technologies such as WiFi, Bluetooth [4], Ultra-WideBand (UWB) [5], and Visible Light Communication (VLC) [6] for their ability to achieve precise and reliable indoor positioning [7].

Indoor positioning techniques based on wireless communication technologies are widely used, and fingerprinting is the most popular for the positioning field due to its simplicity and high accuracy [8–10]. Other techniques include Time of Arrival (ToA), Time Difference of Arrival (TDoA), and Angle of Arrival (AoA) for indoor positioning. In recent studies, researchers have focused on improving the performance of indoor positioning systems. For example, Marina *et al.* [11] proposed a hybrid technique that used Received Signal Strength Indicator (RSSI)/TDOA to solve multipath interference problems and improve location tracking accuracy. Similarly, the Global best Local Neighborhood Particle Swarm Optimization (GbLN-PSO) algorithm was improved in [12] by Mozamir *et al.* conceded measurement errors and increasing the convergence speed of the algorithm. In the study by Wei *et al*. [13], an algorithm for automatically calibrating a large-scale indoor target signal positioning system was proposed to improve system performance, and Hashim *et al*. [14] introduced a method of distance estimation using a path loss model based on the PSO algorithm.

As previously mentioned, the fingerprinting, which is widely used due to its simplicity, generally consists of two steps. In the offline step, specific locations within the indoor environment are selected as Reference Points (RPs), and the Received Signal Strength (RSS) from all Access Points (APs) is measured at these locations. The collected RSS values are stored in a database and used to construct a fingerprint map. In the online step, the RSS that is measured by the User Equipment (UE) is matched with the fingerprint map; the result of this match determines the location of the UE [15]. Matching techniques such as cosine similarity, k-Nearest Neighbor (kNN), and Weighted k-Nearest Neighbor (WKNN) [16] are used in this step. In cosine similarity, the similarity between two vectors, which represent the fingerprint map and the RSS of the UE, is calculated using the cosine angle between the vectors. The value of the similarity is between –1 and 1, and when the value is close to 1, this means that the locations of the RP and the UE are very close each other. kNN is an unsupervised machine learning technique that uses Euclidean distance to calculate the proximity between a fingerprint map and the

Manuscript received March 3 2023; revised April 5 2023, accepted June 1 2023; published March 8, 2024.

UE. Localization is performed by determining the proximity between the RP and the UE; in this process, a smaller Euclidean distance indicates a higher proximity. The WKNN method improves localization precision by adding weights to the kNN.

However, the above matching techniques suffer from the problem of decreased processing speed as the fingerprint map grows larger. To solve this problem, we propose a reinforcement learning-based localization technique. Reinforcement learning is a method in which an agent performs actions in a given environment and then learns by observing changes in state and receiving rewards accordingly. In the proposed method, an agent, which is a reinforcement learning element, performs indoor localization based on the RSS of the UE. When the construction of the fingerprint map is complete, the location of the AP that transmits the highest RSS to the UE is designated as the initial prediction point for the agent. Subsequently, the agent moves the prediction point using one action of up, down, left, right, and stop, and it learns by receiving rewards based on proximity to the actual UE location. The trained reinforcement learning model achieves fast processing and identifies the RP closest to the UE.

The paper is structured as follows. In Section II, the system model used in this study is described. Section III presents a detailed explanation of the proposed positioning approach. In Section IV, the simulation parameters and results are presented. In Section V, the conclusion is summarized, and future work directions are explained in view of improving indoor localization.

II. System Model

This section provides a detailed description of the system model that is used to evaluate the proposed technique. First, the indoor VLC environment is described, including the size of the environment and the placement of the Light Emitting Diode (LED) APs. Next, the channel characteristics of the indoor VLC environment are analyzed, distinguishing between LOS and Non-Line of Sight (NLOS) signal paths. Finally, based on the analysis of the channel characteristics, the power distribution of the overall channel is presented.

A. Indoor Environment Configuration

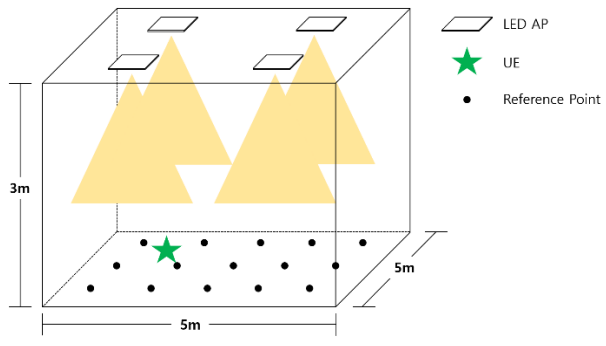

Fig. 1. Indoor VLC environment configuration.

The indoor VLC environment is shown in Fig. 1. As can be seen in the figure, the size of the indoor environment is $5m \times 5m \times 3m$, and it is assumed to be an empty space. Four LED APs are attached to the ceiling, 3 m above the floor. In this case, the transmission power of all LED APs is considered to be equivalent. In addition, the green stars in the figure represent UE and the black dots represent RP. The transmission characteristics of the LED APs and the receiving characteristics of the UEs used in the experiment will be explained in detail in the simulation section.

B. VLC Channel Characteristic Analysis

This subsection analyzes the VLC channel characteristics of signals transmitted from the LED AP. As mentioned earlier, signals transmitted from the AP reach the receiver through both LOS and NLOS paths. Fig. 2 illustrates the signal path for both LOS and NLOS scenarios.

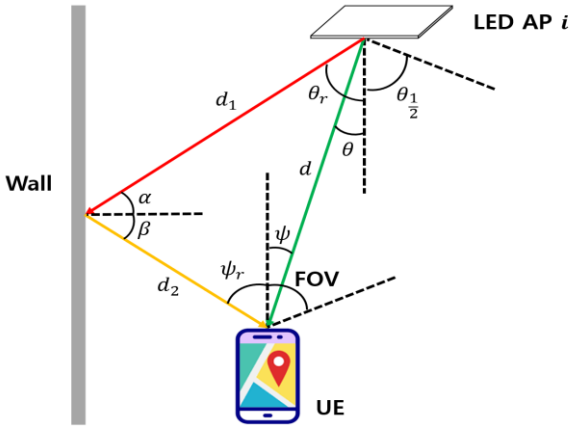

Fig. 2. Channel model for directed and non-directed path between LED AP and UE.

As shown in Fig. 2, in the LOS scenario, the optical signal from the LED AP travels directly to the UE. In the Fig. 2, d represents the distance between the LED AP and the UE, θ represents the irradiation angle of the LED AP, and ψ represents the reception angle of the UE. In the NLOS scenario, the optical signal from the LED AP travels to the UE indirectly by reflection from the wall. This is represented in the diagram, where, d_1 represents the distance between the LED AP and the wall, d_2 represents the distance between the wall and the UE, θ_r represents the irradiation angle from the LED AP to the wall, and ψ_r represents the reception angle at the UE of the signal from the wall. The RSS that the UE receives from the LED AP in these two scenarios can be expressed by the following equation.

$$h^b_{u,dir.} = P_t \frac{A(m+1)}{2\pi d^2} cos^m(\theta) T_s(\psi) C(\psi) \cos(\psi) \quad (1)$$

$$h^b_{u,non-dir.} = P_t \frac{A(m+1)}{2\pi d_1^2 d_2^2} \rho cos^m(\theta) dA_{wall} \cos(\alpha) \cos(\beta) T_s(\psi_r) C(\psi_r) \cos(\psi_r) \quad (2)$$

where, $h_{u,dir.}^{b}$ in Eq. (1) represents the RSS in the LOS scenario, and $h_{u,non-dir.}^{b}$ in Eq. (2) represents the RSS in the NLOS scenario. In the two equations above, P_t represents the transmission power of the LED AP. In Eq. (1), A represents the active area of the UE, m represents the Lambertian order, $T_s(\psi)$ represents the optical filter gain, and $C(\psi)$ represents the optical concentration gain. The value of ψ can be represented as $0 < \psi < \psi_c$, where ψ_c is the Field of View (FOV) of the UE. In Eq. (2), ρ represents the wall reflection coefficient, and dA_{wall} represents the surface element of the wall. The value of ψ_r can be represented as $0 < \psi_r < \psi_c$ in this Eq. (2).

Based on the RSS derived from the above equations, the overall channel RSS distribution can be obtained using Eq. (3).

$$h_{total} = h_{dir.} + h_{non-dir.} \tag{3}$$

Based on Eq. (3), the overall channel distribution can be represented as shown in Fig. 3.

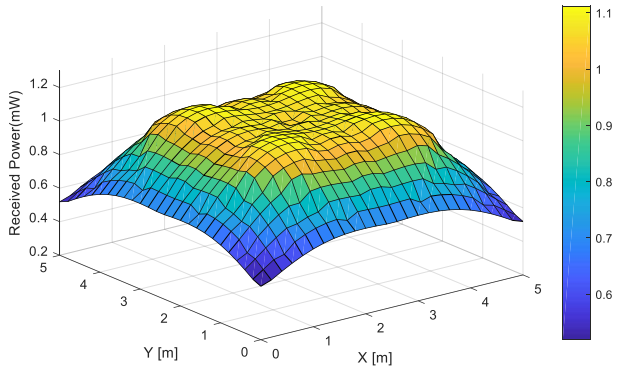

Fig. 3. Overall channel RSS distribution.

The next section discusses in detail the approach to estimating the location of the UE based on Deep Q-Network (DQN) in an indoor VLC environment with the channel distribution shown in Fig. 3.

III. Proposed Positioning Method

This section provides a detailed description of the proposed positioning method, which can be understood to comprise two techniques: fingerprinting and DQN. In indoor environments, the RSS is collected from each AP by selecting each RP. The collected RSS is stored in a fingerprint map and is later matched through the DQN agent. Then, the RSS for each AP is measured by the UE, and the DQN agent is initialized based on this RSS. Subsequently, the position of the UE is determined based on the actions and rewards of the DQN agent. A detailed explanation of each technique is provided in the following subsections.

A. Fingerprinting Technique

The fingerprinting technique is a promising solution for indoor positioning, and it offers high accuracy and simplicity when used in conjunction with RSS. The fingerprinting technique can be divided into two steps: the offline stage and the online stage. The offline stage involves measuring the RSS from each LED AP at a specific location, called the RP, in the indoor environment. The RPs are uniformly distributed in the environment. After the RSS measurements have been completed at all RPs, they are used to build a fingerprint database. The fingerprint database FP_{DB} can be represented as Eq. (4).

$$FP_{DB} = \begin{bmatrix} h_1^1 & \cdots & h_1^b & \cdots & h_1^B \\ \vdots & & \vdots & & \vdots \\ h_s^1 & \cdots & h_s^b & \cdots & h_s^B \\ \vdots & & \vdots & & \vdots \\ h_S^1 & \cdots & h_S^b & \cdots & h_S^B \end{bmatrix} \tag{4}$$

where, h_s^b represents the RSS between AP_b and RP_s; AP_b represents the bth AP; the total number of APs is B; RP_s represents the sth RP; the total number of RPs is S.

After the construction of the fingerprint map as described above, the offline step is complete and the online step begins. The online step is the process of location estimation based on the RSS of the UE. In this study, a deep reinforcement learning technique called DQN is used for location estimation. This technique will be explained in detail in the following subsection.

B. Deep Q-Network

This subsection explains the use of DQN for positioning. DQN is a Q-learning algorithm that predicts Q-values from input data and selects the optimal action based on these values to improving the stability and performance of reinforcement learning. Techniques such as experience replay, target network, and deep network are introduced to achieve this.

The study considers four reinforcement learning elements: Agent, State, Action, and Reward. The Agent is responsible for designating the predicted location based on the RSS of the UE. The State represents the fingerprint data for the predicted point location, which is determined based on the RSS of the UE. Action refers to the movement of the predicted point location—up, down, left, or right—and Reward represents the Agent's reward based on the Action. Next, additional explanations for State, Action, and Reward will be presented.

1) State

The state of the Agent refers to the ID number of the RP where the prediction point is located. Therefore, in this study, the size of the State is determined by the total number of RPs in the indoor environment. Using the number of RPs and the fingerprint map, the Agent can determine the RSS from each AP at the current prediction point. The initial State selects as the prediction point the location of the AP that sends the strongest signal of all the RSS received by the UE from each AP. This allows the Agent to deduce the location of the UE with minimal effort.

2) Action

The Action of the Agent refers to the movement of the prediction point. In this study, a total of five Actions were set. The configured Actions are as follows:

- Up = $P_E(t+1) = n_s + RP_x$
- Down = $P_E(t+1) = n_s - RP_x$
- Left = $P_E(t+1) = n_s - 1$
- Right = $P_E(t+1) = n_s + 1$
- Stop = $P_E(t+1) = P_E(t)$

where, P_E represents the prediction point location of the Agent, n_s refers to the current State, and RP_x refers to the number of RPs on the x-axis.

3) Reward

In this system, Rewards are calculated based on the Euclidean distance between the Agent's prediction point and the RSS of the UE. If the current state has a smaller Euclidean distance value than does the previous state, the action receives a reward of +1, and if the Euclidean distance value is increased, the action receives a penalty of –1.

4) Network configuration

This subsection explains the network architecture of the DQN model used in this study. This network architecture is shown in Fig. 4. The Agent's state is input into the model, and the model is designed to output one of five actions. There are two hidden layers, and the Rectified Linear Unit (ReLU) activation function is used. Furthermore, the size of the replay memory is set to 81,920, and the learning rate is set to 0.001.

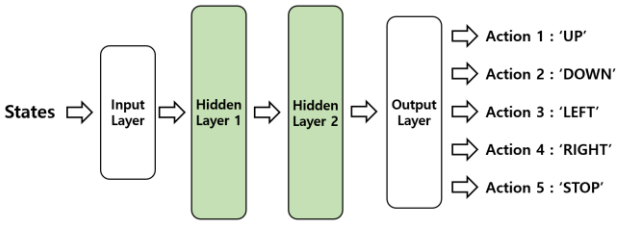

Fig. 4. DQN network configuration.

IV. Simulation Results

This section discusses the simulation parameters and the positioning results of the DQN model. Table I lists the parameters for the indoor environment characteristics, transmitter, receiver, and DQN model. The VLC scenario was implemented using MATLAB in the indoor environment, and the DQN model was designed using Python and implemented using the keras-rl2 library.

In Table I, 676 RPs are positioned such that the RPs have 20 cm intervals between two points from (0,0) [m] to (5,5) [m] considering room size. The update of the target network in DQN refers to the frequency of updating the target network across the total number of episodes. Hence, it means that the target network is updated once every 100 episodes. The epsilon-greedy policy is a commonly used exploration strategy in reinforcement learning. It balances the exploitation of actions known, to have high-reward and the exploration of unknown actions to find the optimal policy. At each step of the learning process, the agent chooses between exploring a new action with some probability epsilon (epsilon is usually a small value) or exploiting the action that currently has the highest estimated value. The results of the DQN model training, which uses these parameter are shown in Fig. 5.

TABLE I. Simulation Parameters

Parameter		Value
Indoor Environment	Room size	$5m \times 5m \times 3m$
	Number of APs	4
	Number of RPs	676
	Reflection coefficient	0.8
Transmitters	Transmission power	10 W
	Wavelength	420 nm
	Elevation	–90°
	Half power semi-angle	60°
Receiver	Field of view	60°
	Height	0.7 m (from bottom)
	Optical filter gain	1
	Optical concentrator gain	1.5
	Active area of UE	$1cm^2$
Deep Q-Network	Episode	50,000
	Max steps per episode	100
	Replay memory size	81,920
	Update of target network	100
	Policy	Epsilon-greedy policy

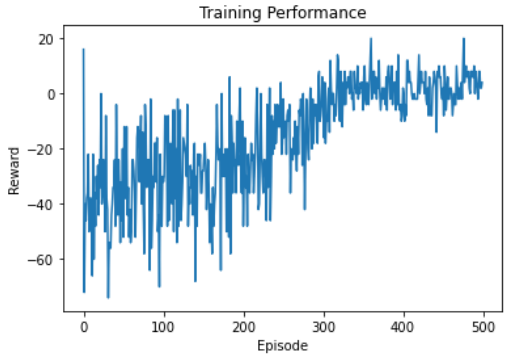

Fig. 5. Training performance of DQN.

Fig. 5 shows the reward per episode. The reward converges to a value greater than 0 as the episodes increase. This indicates that the Agent is finding the adjacent RPs to the UE more quickly as training progresses. Based on the trained model described above, a positioning test was carried out, and it was confirmed that the nearest reference point to the UE was identified in approximately 0.03 seconds. The positioning error was less than 15 cm.

Table II presents the results of a performance comparison between the proposed method and existing methods in an overall channel scenario. The schemes are shown in Table II; FP-based DQN is the method proposed in this paper. The triangulation method estimates the distance between the AP and the UE based on the received RSS from the UE. After it tracks the distance from each AP to UE, triangulation calculates the actual position of UE. The performance was evaluated in terms of processing time and positioning error. Because the triangulation method estimates the distance based on the RSS value and derives the positioning result, the processing time is short. However, the positioning error is

1.298m, and the positioning accuracy is low. On the other hand, the proposed method achieves a similar processing time to triangulation and has a smaller positioning error of 0.15 m.

TABLE II. COMPARISON BETWEEN PROPOSED AND CONVENTIONAL SCHEME

Performance	FP-Based DQN (Proposed)	Triangulation [17]
Processing Time [s]	0.03	0.00237
Positioning Error [m]	0.15	1.298

V. CONCLUSION

As indoor environments become larger and more complex, indoor positioning has emerged as an important research area. Hence, this paper investigates user positioning in indoor VLC environments through the DQN, which is an artificial intelligence technology. The objective of this study is to analyze the channel characteristics of the indoor VLC environment, build a fingerprint map, and use DQN to match the RSS of the UE. Traditional matching algorithms must manage an increase in processing time as the size of the fingerprint map increases. However, a trained DQN model can perform the matching quickly. This study designed a DQN model to perform optimal matching by a short processing time. To achieve this, the reinforcement learning elements of State, Action, and Reward were defined, and experiments were conducted to validate performance. The experimental results showed an positioning error of 15 cm within 0.03 seconds. Ultimately, the proposed model achieved fast position determination while increasing location accuracy by deploying a large number of RPs.

In future, we will improve the performance of the DQN model by adjusting its hyperparameters, and we will research policy optimization solutions such as Deep Deterministic Policy Gradient (DDPG). Additionally, current indoor environments are assumed to be small offices. Hence, there are plans to broaden the current assumption to include larger indoor environments to obtain the positioning of moving users and to verify the effectiveness of the proposed system in more realistic indoor environments.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Sung Hyun Oh and Jeong Gon Kim conceived the idea of this study. Sung Hyun Oh made substantial contributions to the system implementation and data analysis. Jeong Gon Kim contributed significantly to the interpretation of the results and supervised the conduct of this study. Sung Hyun Oh drafted the original manuscript. Jeong Gon Kim critically revised the manuscript for intellectual content. All authors approved the final version of the manuscript.

FUNDING

This work was supported by a National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NRF-2021R1F1A1063845).

REFERENCES

[1] A. Kumari and D. Bhatt, "Advanced system analysis and survey on the GPS receiver front end," *IEEE Access*, vol. 10, pp. 24611–24626, Mar. 2022.
[2] T. Antipas and M. Lawrence, "Accuracy and cluster analysis of 5.3 GHz indoor and 285 MHz semi-urban MIMO LOS and NLOS propagation multipaths," *Journal of Communications*, vol. 18, pp. 135–139, Feb. 2023.
[3] S. Sadowski and P. Spachos, "RSSI-based indoor localization with the Internet of Things," *IEEE Access*, pp. 30149–30161, June 2018.
[4] N. A. Khamla, E. N. Lukito, Widyawan, and H. Kazuhiko, "Indoor location tracking system based on Android application using Bluetooth low energy beacons for ubiquitous computing environment," *Journal of Communications*, vol. 15, pp. 572–576, July 2020.
[5] P. Krzysztof, G. Damian, T. Mateusz, and M. Artur, "UWB positioning system with the support of MEMS sensors for indoor and outdoor environment," *Journal of Communications*, vol. 15, pp. 511–518, June 2020.
[6] M. S. M Gismalla and M. F. L Abdullah, "Optimization of received power and SNR for an indoor attocells network in visible light communication," *Journal of Communications*, vol. 14, pp. 64–69, Jan. 2019.
[7] P. S. Farahsari, A. Farahzadi, J. Rezazadeh, and A. Bagheri, "A survey on indoor positioning systems for IoT-based applications," *IEEE Internet of Things Journal*, vol. 9, pp. 7680–7699, May 2022.
[8] Y. S. Eroglu, I. Guvenc, N. Pala, and M. Yuksel, "AOA-based localization and tracking in multi-element VLC systems," in *Proc. IEEE 16th Annual Wireless and Microwave Technology Conference (WAMICON)*, 2015, pp. 1–5.
[9] B. V. Krishnaveni, K. S. Reddy, and P. R. Reddy, "An Introduction to the TOA measurement for UWB indoor localization systems," in *Proc. 5th Conference on Information and Communication Technology (CICT)*, 2020, pp. 1–4.
[10] M. Han and G. Zeng, "Research on indoor radio frequency positioning algorithm based on TDOA," in *Proc. International Seminar on Computer Science and Engineering Technology (SCSET)*, 2022, pp. 89–92.
[11] K. Marina, E. S. Shaimaa, and Z. Abdelhalim, "Performance enhancement of an indoor localization system based on visible light communication using RSSI / TDOA hybrid technique," *Journal of Communications*, vol. 15, pp. 379–389, May 2020.
[12] M. S. Mozamir, R. B. Abu Bakar, W. I. Soffiah Wan Din, and Z. Binti Musa, "Improved GbLN-PSO algorithm for indoor localization in wireless sensor network," *Journal of Communications*, vol. 16, pp. 242–249, June 2021.
[13] Z. Zhitao, Y. Ye, W. Haoqian, and T. Wei, "An automatic calibration of large-scale indoor target signal positioning system," *Journal of Communications*, vol. 11, pp. 598–608, June 2016.
[14] H. A. Hashim, S. L. Mohammed, and S. K. Gharghan, "Path loss model-based PSO for accurate distance estimation in indoor environments," *Journal of Communications*, vo. 13, pp. 712–722, Dec. 2018.
[15] H. Q. Tran and C. Ha, "Fingerprint-based indoor positioning system using visible light communication — A novel method for multipath reflections," *Electronics*, vol. 8, pp. 63, Jan. 2019.
[16] M. T. Van, N. V. Tuan, T. T. Son, H. Le-Minh, and A. Burton, "Weighted k-nearest neighbor model for indoor VLC positioning," *IET Communication*, vol. 11, pp. 864–871, Mar. 2017.

[17] N. A. Mohammed and M. A. Elkarim, "Exploring the effect of diffuse reflection on indoor localization systems based on RSSI-VLC," *Optics Express*, vol. 23, pp. 20297–20313, July 2015.

Copyright © 2023 by the authors. This is an open access article distributed under the Creative Commons Attribution License (CC BY-NC-ND 4.0), which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.