Article

Demonstration of Three-Dimensional Indoor Visible Light Positioning with Multiple Photodiodes and Reinforcement Learning

Zhuo Zhang [1,†], Huayang Chen [1,†], Weikang Zeng [1], Xinlong Cao [1], Xuezhi Hong [1,*] and Jiajia Chen [1,2,*]

[1] Centre for Optical and Electromagnetic Research, South China Academy of Advanced Optoelectronics, South China Normal University, Guangzhou 510006, China; zhuo.zhang@coer-scnu.org (Z.Z.); huayang.chen@coer-scnu.org (H.C.); weikang.zeng@coer-scnu.org (W.Z.); xinlong.cao@coer-scnu.org (X.C.)

[2] Department of Electrical Engineering, Chalmers University of Technology, Hörsalsvägen 9, 41296 Gothenburg, Sweden

*   Correspondence: xuezhi.hong@coer-scnu.org (X.H.); jiajiac@chalmers.se (J.C.); Tel.: +86-2034725186 (X.H.); +46-317726034 (J.C.)

†   These authors contributed equally to this work.

check for updates

Abstract: To provide high-quality location-based services in the era of the Internet of Things, visible light positioning (VLP) is considered a promising technology for indoor positioning. In this paper, we study a multi-photodiodes (multi-PDs) three-dimensional (3D) indoor VLP system enhanced by reinforcement learning (RL), which can realize accurate positioning in the 3D space without any off-line training. The basic 3D positioning model is introduced, where without height information of the receiver, the initial height value is first estimated by exploring its relationship with the received signal strength (RSS), and then, the coordinates of the other two dimensions (i.e., X and Y in the horizontal plane) are calculated via trilateration based on the RSS. Two different RL processes, namely RL_1 and RL_2, are devised to form two methods that further improve horizontal and vertical positioning accuracy, respectively. A combination of RL_1 and RL_2 as the third proposed method enhances the overall 3D positioning accuracy. The positioning performance of the four presented 3D positioning methods, including the basic model without RL (i.e., Benchmark) and three RL based methods that run on top of the basic model, is evaluated experimentally. Experimental results verify that obviously higher 3D positioning accuracy is achieved by implementing any proposed RL based methods compared with the benchmark. The best performance is obtained when using the third RL based method that runs RL_2 and RL_1 sequentially. For the testbed that emulates a typical office environment with a height difference between the receiver and the transmitter ranging from 140 cm to 200 cm, an average 3D positioning error of 2.6 cm is reached by the best RL method, demonstrating at least 20% improvement compared to the basic model without performing RL.

Keywords: reinforcement learning; 3D indoor positioning; visible light positioning

1. Introduction

The developments of location-based mobile services and the Internet of Things urgently need stable and precise indoor positioning technologies [1]. As the widely deployed global positioning system (GPS) has poor coverage and accuracy in the indoor environment, indoor positioning systems (IPS) that employ alternative radio frequency (RF) technologies (e.g., Bluetooth, RFID, iBeacon, Wi-Fi, and near-field communication [2–4]) have been investigated. However, the RF-based IPS (e.g., 1–3/5–15/0.1–0.3 m with Bluetooth/Wi-Fi/UWB, respectively [5]) can be largely affected by

electromagnetic interference and multipath effect in a congested environment [6]. Compared with RF technologies, the visible light positioning (VLP) system has features with immunity to electromagnetic interference and high tolerance to multipath interference thanks to the domination of the LOS signal [7–11]. By simultaneously providing illumination and positioning services with the existing indoor lighting equipment (e.g., light-emitting diode (LED)), VLP with high positioning accuracy (e.g., in the order of centimeters [7]) is considered as one low cost and high energy efficiency solution for localization in the indoor environment.

Comparing with the two-dimensional (2D) positioning on a horizontal plane at a known height, three-dimensional (3D) positioning using the same setup is more challenging. In such a case, one needs to map the received signal to one more dimension (i.e., height) which increases the search space for the positioning process. To find the correct position in a larger searching space, the positioning algorithm and/or the hardware in 3D VLP systems are more complex than the 2D ones. From the hardware perspective, 3D VLP systems based on either a single photodiode (PD) or multiple PDs have been proposed. In the single-PD system, 3D positioning has been achieved by combining information from both the PD and the other hardware either at the receiver (e.g., accelerometer [12], rotatable platform [13]) or at the transmitter (e.g., steerable laser [14,15]), which is not necessarily simpler than the multiple-PD system from the system complexity perspective. 3D positioning based on a low complexity receiver with one PD only has also been proposed, which has additional requirements for the radiation patterns or geometric arrangement of LEDs to avoid ambiguity in height estimation [16,17]. 3D VLP systems using multiple PDs have also been proposed, in which the spatial or angular diversity of PDs are explored to estimate the 3D position of the receiver [18–20]. Though it needs more PDs at the receiver, it does not have any special requirement for the transmitter [16,17] and has shown the potential to reduce the number of LEDs for a simpler transmitter [13].

For the 3D positioning algorithm, trilateration and triangulation based methods are widely employed, in which the geometric relationship between the receiver and light sources (e.g., distance [19], incidence/irradiance angles [18,20]) is estimated from the received signal. One popular way to evolve from a 2D VLP algorithm to a 3D one is to conduct a brute-force search on several parallel 2D layers at various heights. After obtaining the horizontal positions on all candidate layers at a pre-defined height set, the estimated 3D position is determined as the one that most likely fulfills the constraint among the coordinates of the three dimensions [16,21,22]. To improve the efficiency in height estimation, a fast search method based on the golden section search (GSS) algorithm has been proposed which can significantly reduce the running time [16]. To further improve the positioning accuracy, machine learning (ML) techniques with outstanding nonlinear fitting capability have been introduced to the VLP systems. Supervised learning (SL) based VLP systems (e.g., neural network [23–25], random forest [26], and *K*-nearest neighbor [27]) have been proposed. However, the SL based VLP systems require sufficient training data to be prepared in advance, which increases the system complexity [27]. The performance of the SL positioning algorithms is also largely affected by the quality of the training. To avoid the above drawbacks, training-data-free ML techniques, such as reinforcement learning (RL), have been employed. Previous studies show that the application of RL in 2D VLP offers high and robust positioning accuracy [28,29]. Though the RL based 2D positioning algorithm has shown a higher tolerance to the error of a priori height information than the conventional one, the height of the object is still assumed to be known in advance under the 2D VLP framework. Moreover, in many applications with mobile devices, the exact height of the receiver is often unknown and could vary dynamically in a range much larger than the height error tolerance of the RL based 2D VLP algorithm.

In this paper, a 3D VLP system using multiple PDs and reinforcement learning is proposed which realizes high accuracy for 3D positioning without needs of data for off-line training. In the 3D VLP system, we first make a coarse estimation of the receiver height by exploring its relationship with the received signal strength (RSS) and then calculate the other two coordinates in the horizontal plane using trilateration [30]. To achieve high 3D positioning accuracy, three methods based on RL with different height update strategies are proposed. Experiments are carried out to evaluate the

performance of the proposed methods under different receiver sizes. The results show that when the height difference between the receiver and the transmitter is within [140, 200]-cm, compared with the case without machine learning (i.e., the Benchmark), all three proposed RL based methods can improve 3D positioning accuracy robustly. Unlike our previous 2D VLP work [28,29] that only estimates the position on a horizontal plane and still requires the height information as an input, this paper is an extension, include three new major contributions: (i) methods for 3D positioning are investigated that output coordinates in all three dimensions without a priori information about any dimensions; (ii) two novel reinforcement learning processes are devised specifically for 3D VLP, which target accuracy enhancement in the horizontal plane and the vertical dimension, respectively, and a combination of them offers the highest 3D positioning accuracy; (iii) the effectiveness of the proposed RL based 3D positioning methods are demonstrated experimentally.

The remainder of the paper is organized as follows. The operation principles of different 3D positioning methods, including the basic model and three RL based methods (i.e., Method 1/2/3) are explained in Section 2. Section 3 shows the experimental setup for 3D VLP, compares the performance of different positioning methods, and analyze the impact of the receiver size. Finally, Section 4 draws conclusions.

2. Operation Principle

A multi-PD VLP system with M ($M \geq 3$) LEDs at the same height on the ceiling is considered in this study. Figure 1 shows the considered 3D VLP system setup and the signal processing flow, including the basic model without RL (later referred to as the benchmark), and three proposed methods that employ RL to improve positioning accuracy.
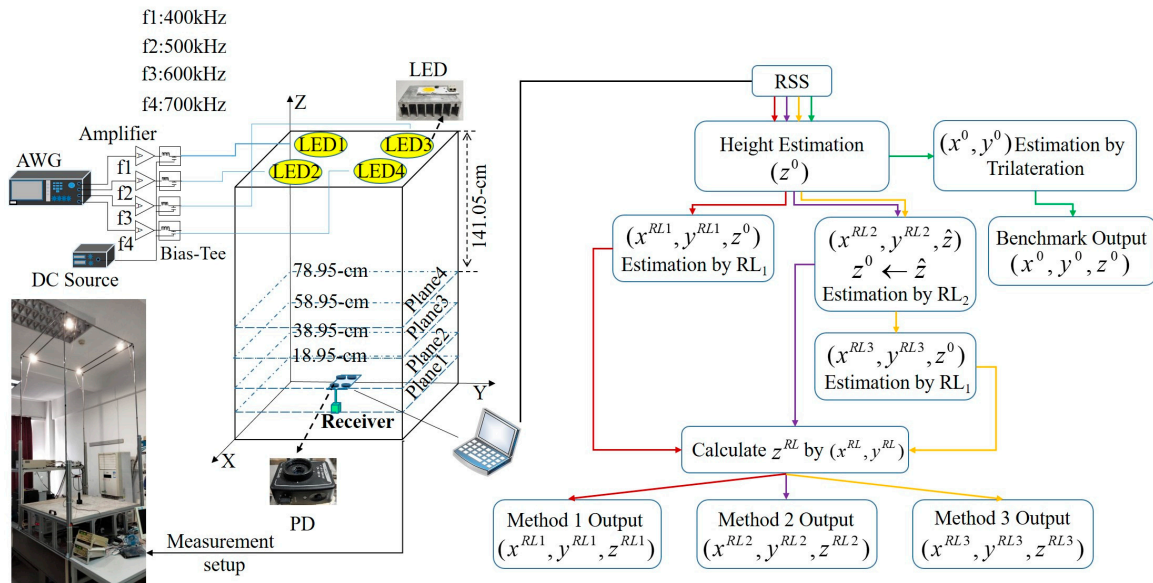
Figure 1. The 3D VLP system setup and the signal processing flow. The green, red, purple, and yellow lines represent the flow for the Benchmark, Method 1, Method 2, and Method 3, respectively. The inset shows the picture of our testbed.

The i-th ($i = 1, 2 \ldots M$) LED is located at $\left(L_i^x, L_i^y, L^z\right)$ and transmits a sinusoidal modulated signal with frequency f_i. At the receiver, N PDs are facing up at the same height, and the n-th ($n = 1, 2 \ldots N$) PD is located at (x_n, y_n, z). The received signal of the nth PD from all the LEDs is represented by

$s_n(t)$, whose power spectrum consists of M peak components at f_i ($i = 1, 2, \dots , M$) [31]. After Fourier transformation, it can be expressed as [31]:

$$S_n^{f_i} = \frac{(P_i)^2(m+1)^2 A^2 \beta^2 h^{2(m+m')}}{4\pi^2 dn, i^{2(2+m+m')}} \tag{1}$$

in which P_i is the transmitted optical power of the ith LED, A is the PD area, β is the PD responsivity, $h = L^z - z$ is the height difference between the receiver and LEDs, m (m') is the Lambertian radiation pattern order of the LED (PD) and $d_{n,i}$ is the distance between the nth PD and the ith LED. Note that the irradiance angle and incidence angle are assumed to be the same in (1) as the PDs (LEDs) are facing up (down). The RSS of these components obtained by the N PDs from the M LEDs can be represented by a vector.

$$\boldsymbol{Rec} = \left[S_1^{f_1}, \dots, S_1^{f_M}, \dots, S_N^{f_1}, \dots, S_N^{f_M}\right] \tag{2}$$

According to the location of the LEDs and PDs, we have:

$$\left(x_n - L_i^x\right)^2 + \left(y_n - L_i^y\right)^2 + h^2 = d_{n,i}^2 \tag{3}$$

2.1. Basic 3D Positioning Model

We first introduce a basic 3D positioning model, which is also referred to as benchmark later to show the accuracy improvement brought by the proposed reinforcement learning methods. Unlike the 2D VLP, the height of the receiver z, which equals to $L^z - h$, is unknown in the 3D VLP system and needs to be estimated. According to (1) and (3), the relationship between h and $S_n^{f_i}$ can be written as:

$$h \le dn, i = \left[\frac{C}{S_n^{f_i}}\left(\frac{h}{dn,i}\right)^{2(m+m')}\right]^{1/4} \le \left[\frac{C}{S_n^{f_i}}\right]^{1/4} \tag{4}$$

where $C = \frac{(P_i)^2(m+1)^2 A^2 \beta^2}{4\pi^2}$.

According to Equation (4), h is no more than the minimum value of $d_{n,i}$. We denote a coarse estimation of h as h_0, which equals d_{min} (i.e., the minimum value of the rightest term in Equation (4) among all possible combinations of N PDs and M LEDs) and is expressed as:

$$h0 = d_{min} = \text{mininum}\left[\frac{C}{S_n^{f_i}}\right]^{1/4} \quad \forall n \in [1, N], \forall i \in [1, M] \tag{5}$$

With h_0, the 2D coordinates on the horizontal plane of the N PDs can be estimated by the conventional trilateration method using (1) and (3). Specifically, we estimate the nth PD's 2D coordinates by solving the following equations:

$$\begin{cases} 2x_n(L_b^x - L_a^x) + 2y_n(L_b^y - L_a^y) = d_{n,a}^2 - d_{n,b}^2 + (L_b^x)^2 - (L_a^x)^2 + (L_b^y)^2 - (L_a^y)^2 \\ 2x_n(L_c^x - L_a^x) + 2y_n(L_c^y - L_a^y) = d_{n,a}^2 - d_{n,c}^2 + (L_c^x)^2 - (L_a^x)^2 + (L_c^y)^2 - (L_a^y)^2 \end{cases} \tag{6}$$

where $a/b/c$ are the indexes of three different LEDs. As there are M LEDs on the ceiling, C_M^3 different pairs of equations can be established [30]. The output of trilateration is obtained by averaging these estimations to mitigate the impact of noise. Our positioning target is the coordinate of the center of the receiver. Assuming the PDs locate symmetrically at the corners of the receiver, the receiver position is obtained by averaging the estimated locations of the N PDs. The above 3D VLP system is referred to as the benchmark, whose output is $(x^0, y^0, z^0 = L^z - h_0)$ (see Benchmark Output in Figure 1).

2.2. Reinforcement Learning To Enhance 3D Positioning Accuracy

h_0 derived from (5) is a coarse estimation of the actual height h. The difference between h_0 and h may not be minor, particularly when the PD is not right below any LED. Since the estimation of the other two coordinates requires the height information as an input, the coarse estimation of h causes error propagation in the benchmark, which results in low positioning accuracy in all three dimensions. Inspired by our previous study [29] that RL can offer high tolerance to inaccurate h in the 2D VLP system, we propose to use RL to improve the positioning accuracy of the 3D VLP system.

The RL mechanism is shown in Figure 2, in which the *Agent* learns knowledge in the action-evaluation *Environment* and improves the *Action* by adapting to the *Environment* [32]. In the 3D VLP system, if the RSS and height are free from the impact of noise in RSS or height estimation error, we can get the exact 3D coordinates by using trilateration. Therefore, the *Environment* to be learned in the 3D VLP system is the error in RSS measurement and the height estimation (see the red box in Figure 2). In other words, the aim of RL is to learn and compensate for the above errors contained in the *Environment* to get a better estimation of the receiver position. As we have multiple PDs available at the receiver, the relative distances between PDs are fixed and can be used to assess the positioning error for reward calculation in RL. The relative distance error vector $\mathbf{E}_{dis}$ is used by the *Agent* to evaluate the *State* of *Environment*, which is defined as:

$$\mathbf{E}_{dis} = \left\{ dis_{(i,j)} - \hat{dis}_{(i,j)} \left| i \neq j; i, j = 1, 2.., N \right. \right\} \tag{7}$$

The $dis_{(i,j)}(\hat{dis}_{(i,j)})$ in Equation (7) denotes the real (calculated) distance between the i-th and j-th PDs. The $dis_{(1,2)}$ of a four-PD receiver is shown in Figure 3 as an example. The *State* and *Reward* in the interaction between the *Agent* and *Environment* are defined as the maximum and average value of $\mathbf{E}_{dis}$, respectively:

$$State = \begin{cases} i, \text{if } \alpha_{i-1} < \max(\mathbf{E}_{dis}) \leq \alpha_i \text{ for } 1 \leq i < G \\ G, \text{if} \max(\mathbf{E}_{dis}) \geq \alpha_{G-1} \end{cases}, \tag{8}$$

$$Reward = \begin{cases} \frac{K-i}{K-1} * 100, \text{ if } r_{i-1} < \text{average}(\mathbf{E}_{dis}) \leq r_i \text{ for } 1 \leq i < K \\ 0, \text{if average}(\mathbf{E}_{dis}) \geq r_{K-1} \end{cases}, \tag{9}$$

where $(\alpha_0, \alpha_1, \dots, \alpha_{G-1})$ and $(r_0, r_1, \dots, r_{K-1})$ are pre-determined constants based on accuracy requirements, G and K are the numbers of possible values for the *State* and *Reward*, respectively. The learning process in RL uses an action-evaluation strategy, where the consequences of actions (i.e., *Reward*) is used as the metric to help find the optimal action at a certain *State* of *Environment*. If the current *State* is not the target state (e.g., 1 in our study), the *Agent* takes an action to adjust the RSS and height coordinate.

There are different ways to conduct 3D positioning incorporating the RL. Pseudocode 1 shows the pseudocode for two methods with different height update strategies, namely RL_1 and RL_2. The RL_1 is used in Method 1 that adjusts the RSS without changing h except for the last action in learning (i.e., h is fixed to be h_0 when adjusting the RSS and only gets updated after the final RSS is obtained), while the RL_2 is used in Method 2 that adjusts RSS and h sequentially in each action. Specifically, in Method 2, (x_n^{new}, y_n^{new}) is obtained by using the updated RSS and $\hat{d}_{n,i} = (Ch^{2(m+m')}/S_n^{f_i})^{\frac{1}{2(2+m+m')}}$ based on trilateration in Equation (3), and then height difference is updated as $\hat{h}$ by averaging the N height differences between each LED and the receiver's plane, which can be expressed as:

$$\hat{h} = \frac{1}{N} \sum_{n=1}^{N} \sqrt{\left(Ch^{2(m+m')}/S_n^{f_i}\right)^{\frac{1}{2+m+m'}} - \left(x_n^{new} - L_i^x\right)^2 - \left(y_n^{new} - L_i^y\right)^2} \tag{10}$$
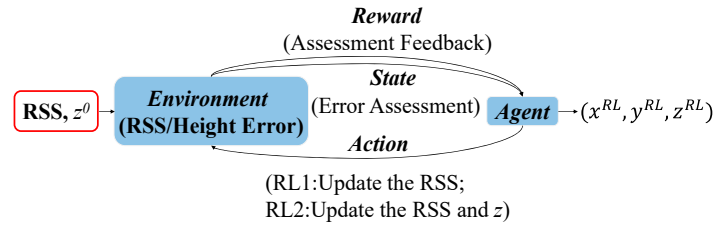
Figure 2. Schematic diagram of the reinforcement learning mechanism in the 3D VLP system.
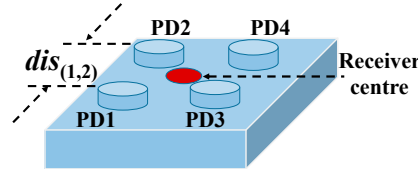
Figure 3. Receiver structure.

Pseudocode 1: Pseudocode for Method 1 and Method 2

---

1. **Input:** the RSS vector $\boldsymbol{Rec}$, the initial receiver height z^0
2. **Output:** the refined RSS vector $\boldsymbol{Rec}^{RL}$, the 3D coordinates of the receiver (x^{RL}, y^{RL}, z^{RL})
3. Calculate the coordinate of the nth PDs (x_n, y_n) with $\boldsymbol{Rec}$ and z^0 using trilateration
4. Obtain $\mathbf{E}_{dis}$ using (7)
5. Obtain the current *State* and *Reward*
6. $k = 0$
7. **for** $k <$ the maximum number of iterations **do**
8. $k \leftarrow k+1$
9. **for** *State* does not reach the target state and *Reward* is not less than the last one **do**
10. Obtain $2M{\times}N$ new $\boldsymbol{Rec}$ for $2M{\times}N$ actions (i.e., increase/decrease one of the $M{\times}N$ received signal strengths by a pre-defined step in one action)
11. **for** each new $\boldsymbol{Rec}$ **do**
12. Calculate the new coordinates of the nth PDs
13. Obtain $\mathbf{E}_{disr}$ using (7)
14. Update *State* and *Reward*
15. **end for**
16. Choose the *Action* with maximum *Reward*
17. Obtain $\boldsymbol{Rec}_{new}$ and update $\boldsymbol{Rec} \leftarrow \boldsymbol{Rec}_{new}$

　　　　　　　　　　　　　　　　　　　For Method 1 (Update the RSS)

18. Calculate the new coordinates of the nth PDs (x_n^{new}, y_n^{new})
19. Obtain $\hat{h}$ by (10) and update $z \leftarrow \hat{z} = L^z - \hat{h}$

　　　　　　　　　　　　　　　　　　　For Method 2 (Update the RSS and z)

20. **end for**
21. **end for**
22. Obtain finally $\boldsymbol{Rec}^{RL}$
23. Update z　$(z \leftarrow z^0$ in RL₁ and $z \leftarrow \hat{z}$ in RL₂$)$
24. Calculate the coordinate of the nth PDs (x_n^{RL}, y_n^{RL}) with $\boldsymbol{Rec}^{RL}$ and z
25. Obtain h_{RL} by (10) and update $z^{RL} \leftarrow L^z - h_{RL}$
26. Obtain the 3D coordinates of the receiver (x^{RL}, y^{RL}, z^{RL}) by averaging the coordinate of PDs

---

* z^{RL} corresponds to z^{RL1} and z^{RL2} for Method 1 and Method 2, respectively.

For the RSS adjustment in the RL_1/RL_2, each time one element of the RSS vector **Rec** is increased or decreased by step which is a minimum step to adjust the RSS values. After taking an action that modifies **Rec** (in RL_1 or RL_2) and h (in RL_2), the 3D coordinates of all PDs are obtained via trilateration and used to calculate its *Reward* based on a new $\mathbf{E}_{dis}$ according to Equation (9). The *Agent* chooses the *Action* with the maximum *Reward*, and update the *State* according to Equation (8).

Both methods continue the learning process until the target state or the maximum number of iterations. After learning, the estimated 3D coordinates of PDs after the last action in RL are saved. The receiver's 3D coordinates (i.e., $(x^{RL1}, y^{RL1}, z^{RL1}=L^z\text{-}h_{RL1})$ in Method 1 and $(x^{RL2}, y^{RL2}, z^{RL2}=L^z\text{-}h_{RL2})$ in Method 2) are obtained by averaging the coordinates of PDs and used as the final outputs (see Method 1 Output and Method 2 Output in Figure 1).

It is worth noting that the two methods concentrate on positioning accuracy improvement in the horizontal plane and height, respectively. In the RL_1, the learning process only puts the efforts to optimize the X and Y coordinates in the horizontal plane, while the RL_2 does one-step refinement for both the height and RSS in each action. It is also shown in the results (see Section 3), the two methods cannot achieve positioning accuracy improvement in all three dimensions simultaneously. Therefore, we combine the RL_1 and the RL_2, which is referred to as Method 3. Since our previous research in [29] shows that reinforcement learning can tolerate the inaccuracy of h to some extent, in Method 3 we use the RL_2 to update h and RSS, which are followed by the RL_1 to update the X and Y coordinates. Finally, the height estimation is refined according to Equation (10), and $(x^{RL3}, y^{RL3}, z^{RL3}=L^z\text{-}h_{RL3})$ is obtained (see Method 3 Output in Figure 1). The pseudocode for Method 3 is shown in Pseudocode 2.

Pseudocode 2: Pseudocode for Method 3

1. Input: the RSS vector *Rec*
2. Output: Coordinate of the receiver $(x^{RL3}, y^{RL3}, z^{RL3})$.
3. Estimate h_0 with (5) and $z^0=L^z-h_0$
4. Run RL_2 to obtain Rec^{RL} and $\hat{z}$
5. Update $Rec \leftarrow Rec^{RL}, z^0 \leftarrow \hat{z}$
6. Run RL_1 to obtain the 3D coordinate of the receiver (x^{RL3}, y^{RL3}, z^0)
7. Refine height z^{RL3}
8. Obtain the final 3D coordinates of the receiver $(x^{RL3}, y^{RL3}, z^{RL3})$

To better illustrate the RL processes in different 3D VLP methods, Table 1 summarizes the features of the three proposed methods. The RL-based 2D VLP method (i.e., PWRL in [29]) is also listed for comparison.

Table 1. Summary of different RL-based VLP methods.

Method RL Element	3D VLP	3D VLP	3D VLP	2D VLP
	Method 1	**Method 2**	**Method 3**	**PWRL [29]**
Input	(1) Measured RSS and (2) height estimated based on the basic 3D positioning model			(1) Measured RSS and (2) exact height
Environment	Errors in RSS measurement and height estimation			RSS error
Action	RSS adjustment under an estimated height(RL_1)	RSS and height adjustments (RL_2)	RSS adjustment (in both RL_1 and RL_2) and height adjustment (in RL_2)	Only RSS adjustment, where height is known.
State	Determined by the relative distance error with (8)			
Reward	Determined by the relative distance error with (9)			

3. Experiment Investigation

3.1. Experimental Setup

The performance of the proposed 3D VLP methods is investigated experimentally. Figure 1 shows the experimental setup. There are four LEDs (Cree CXA2435) on the ceiling with coordinates of (24.2, 19.8, 218.9), (83.5, 19.7, 218.9), (22.7, 78.1, 218.9), (82.6, 77.8, 218.9) in centimeter (cm), respectively. Four sinusoidal signals of frequency (400/500/600/700 kHz) from four signal generators are amplified and then combined with direct current (DC) signals via Bias-Tees (ZFBT-4R2GW+) to drive the four LEDs, respectively. As shown in Figure 3, the receiver consists of four PDs (PDA100A2) on the four corners. To ensure that the signal from all four LEDs can be received by the PD (field of view: ~60°) in the 120 cm × 120 cm area, the height difference between the PD and the LED of our test space should be larger than 71 cm. To investigate the impact of receiver size on the performance of the proposed 3D VLP methods, the distance between adjacent PDs is adjusted (i.e., $dis_{(1,2)}$ = 10/20/30/40 cm). In order to get ground truth locations of PDs and LEDs, we divide the area of a solid aluminium plate into many 10 cm × 10 cm grids with a ruler/tape measure which has the resolution of 1 mm and use the lower left side as the origin. The PD is mounted with an optical mounting post on a base which is moved on the grid to change the 2D coordinates on horizontal planes (see Figure 4a). The height of the PD is adjusted by changing the length of the optical mounting post on the base, and is measured manually with a ruler. The horizontal and height coordinates of LEDs are determined by finding their projections on the solid aluminium plate and their distance to this plate with the help of a plumb bob (see Figure 4b,c). To lower the measurement error, the averaged value of multiple measurements is used as ground truth locations. We take measurements at four test planes of different heights with 20 cm spacing, whose Z coordinates are 18.95/38.95/58.95/78.95 cm, corresponding to 199.95/ 179.95/159.95/139.95 cm for h, respectively. The height difference between the receiver and the ceiling in the testbed is about [140, 200]-cm, which emulates the cases of positioning a hand-held device in a typical office environment. Note that the tilt of a hand-held device could severely affect the positioning accuracy as Equation (1) no longer holds. As the average elbow height for a mixed male/female human population is 104.14 cm when he/she stands up [33], this offers about ± 30 cm margin for a room with a ceiling height of 270 cm. In case the height difference is larger, a stronger light source is needed to guarantee a reasonable signal-to-noise ratio (SNR) for high accuracy positioning [30]. For each test plane, four PDs are adjusted to the same height, and samples are taken at 49 uniformly distributed locations in the 120 cm × 120 cm area. The RSS at the receiver is measured using a spectrum analyzer (8593E, Agilent, Elgin, IL, USA) with a sweep time of 30 ms and averaged over 10 measurements. For example, the measured RSS in the center of Plane 4 are 0.354-μW, 0.292-μW, 0.309-μW, 0.319-μW for the sine wave signals from the four LEDs, respectively. For a practical receiver of small form factor, discrete Fourier transform of the temporal samples from an analog-to-digital converter can be conducted to measure the signal strength at different frequencies. The detailed parameters of the experimental setup are listed in Table 2.
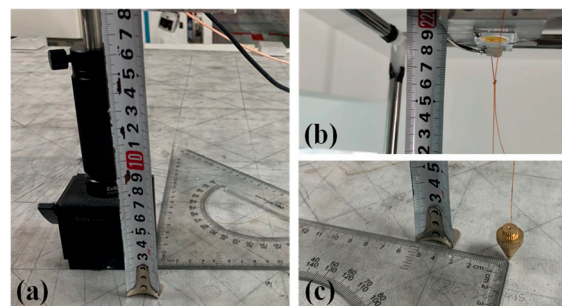

Figure 4. Measurements of the ground truth locations of (**a**) PD, (**b**) and (**c**) LED.

Table 2. Experimental parameters.

| Parameter	Value
Space size(length × width × height)	120 × 120 × 220 (cm)
Coordinates of LED1/LED2/LED3/LED4	(24.2, 19.8, 218.9)/ (83.5, 19.7, 218.9)/ (22.7, 78.1, 218.9)/ (82.6, 77.8, 218.9) (cm)
$f1/f2/f3/f4$	400/500/600/700 (kHz)
LED voltage	18.0 (V)
LED current	0.32 (A)
Lambertian order of LED (m)	1.78
Lambertian order of PD (m')	3.56
Distance between PD1 and PD2 ($dis_{(1,2)}$)	10/20/30/40 cm
Heights of Plane 1/2/3/4	18.95/38.95/58.95/78.95 (cm)
Height difference between receiver 1/2/3/4 to LEDs (h)	199.95/179.95/159.95/139.95 (cm)

To balance running time and positioning accuracy, we set $_\triangle step$ to 0.1 μw to adjust **Rec** in each action during the learning process. $K = G = 5$, $(\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4) = (0, 0.2, 0.5, 1, 2)$ in cm, and $(r_0, r_1, r_2, r_3, r_4) = (0, 0.05, 0.125, 0.25, 0.5)$ in cm. In general, the accuracy performance is improved when the number of iterations increases and exhibits a trend of convergence when the number of iterations exceeds a certain value. The number of iterations shall not be too small to achieve the state of convergence. On the other hand, since the processing time and computational complexity of the algorithm increase with a larger number of iterations, the number of iterations shall not be too large. Therefore, the maximum allowable number of iterations in RL based methods is set to 1000 empirically in this experiment to balance the complexity and positioning accuracy.

3.2. Performance Evaluation

We run Method 1/2/3 off-line with MatLab (MathWorks, Natick, MA, USA) on a desktop computer (i5 processor @2.29 GHz (Intel, Santa Clara, CA, USA) with 16 GB RAM) and the measured average processing time is 0.96/0.44/0.69-s, respectively. Figure 5 shows the spatial distribution of 3D positioning error for four different positioning methods (i.e., the benchmark and methods 1/2/3) when $dis_{(1,2)}$ equals to 40 cm. The 3D/2D positioning errors are the Euclidean distance between the real coordinates and the calculated coordinates of the receiver in the 3D/2D space, respectively. To illustrate the 3D positioning accuracy intuitively, we take the actual position of the sampling point as the center of the sphere and the 3D positioning error as the radius of the sphere. The radius r_{sphere} is defined as:

$$r_{sphere} = \sqrt{(x - x_{real})^2 + (y - y_{real})^2 + (z - z_{real})^2},$$ (11)

where (x, y, z) denotes the output of the positioning algorithms and $(x_{real}, y_{real}, z_{real})$ is the real coordinate of the receiver. The non-uniform distribution of errors is observed, which is the interplay of there location-dependent factors: (a) SNR which is higher at the center of test plane, (b) inaccurate a priori information about the VLP system (e.g., m and m' in (1)) that may cause significant overestimation or underestimation of the distance between PD and LED, and (c) the error in approximating the actual height difference with h_0 in Equation (5) which varies for different incidence/irradiance angles of the PD-LED pair used in the calculation of h_0.
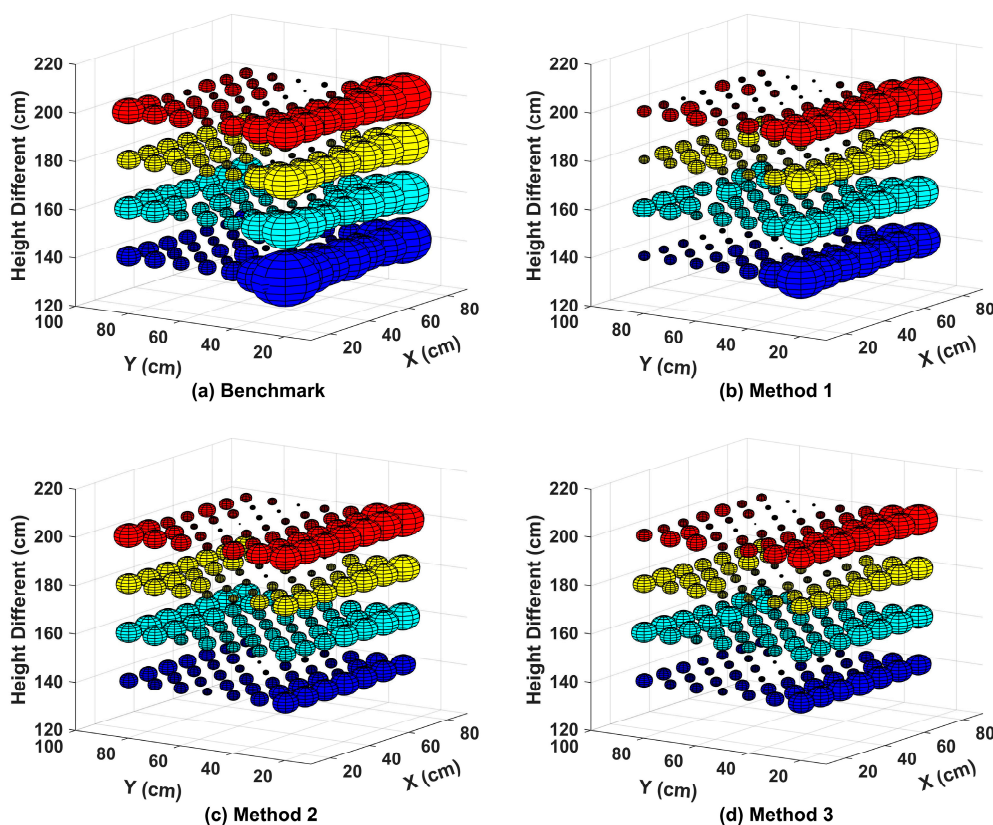
Figure 5. Spatial distribution of the 3D positioning error at different heights in the case of dis$_{(1,2)}$ = 40 cm for the (**a**) Benchmark, (**b**) Method 1, (**c**) Method 2, and (**d**) Method 3.

At the edges of test planes where the SNR is lower, the 3D positioning error is larger than that in the central of test planes with higher SNR. If the overestimation of the distance between the PD and LED happens (e.g., a result due to factor (b)), the approximation error in Equation (5) will be larger. For example, we find that the Lambertian model for LED1/LED2 with the parameters in Table 2 causes overestimation of the distance between LED and PD. This leads to significant larger positioning errors in the region with smaller Y which uses LED1/LED2 to calculate h_0. In general, all three RL based methods achieve higher 3D positioning accuracy than the benchmark in the test planes. Method 2 can reduce the error of some points to very small (e.g., test points on the left half of Figure 5c). However, the positioning error with Method 1 is more uniformly distributed in some planes (e.g., h = 139.95/159.95 cm in Figure 5b,c). Regardless of the height of the receiver's plane, Method 3 offers the best performance among the four methods.

To further analyze the impact of RL on positioning errors in different dimensions, Figure 6a–c give the cumulative distribution function (CDF) of height/2D/3D error, respectively. Here, 2D error represents the error in the horizontal plane. Plane 2 (i.e., h = 179.95 cm) is used as an example in Figure 6. As shown in Figure 6a–c, all three RL based methods can reduce the height/3D positioning error. For the height dimension, the improvement in Method 3 is most significant, which can reduce the height error from ~5.4 cm to ~3.5 cm for 90% of the test points. Thanks to the additional height update procedure, Method 2 outperforms Method 1 in terms of height estimation accuracy. As shown in Figure 6b, Methods 1 and 3 perform similarly and reduce the 2D positioning error significantly when compared with the Benchmark. For Method 2, though more points are having lower positioning error when compared with the Benchmark, the number of points with larger positioning error also increases. For example, the ratios of points with 2D positioning error of ≤1.76 cm (≥3.0 cm) are 71% and 57% (25% and 17%) for Method 2 and the Benchmark, respectively. This is consistent with the enhanced non-uniformity by Method 2 shown in Figure 5c. In Figure 6c, the 3D positioning error of

90% test points with the Benchmark is less than ~5.4 cm, which can be reduced to less than ~4.9 cm, ~4.0 cm, and ~3.6 cm by Methods 1–3, respectively. As Method 3 exhibits superior performance in both the height dimension and the XY plane, it offers the best 3D positioning performance among the four tested algorithms.
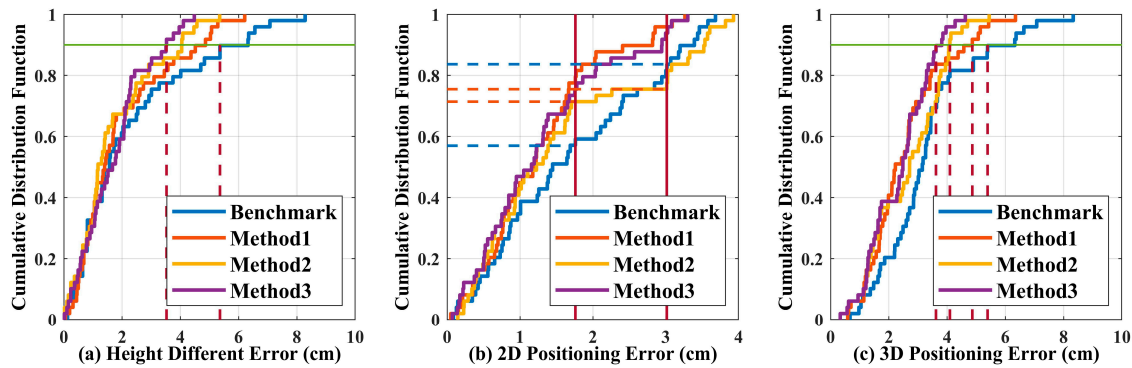
Figure 6. The cumulative distribution function of (**a**) height, (**b**) 2D, and (**c**) 3D positioning errors at Plane 2 in the case of $dis_{(1,2)} = 40$ cm.

Figure 6 implies that Method 2 outperforms Method 1 in the height dimension, while Method 1 outperforms Method 2 in the horizontal plane. Method 3 inherits the advantages of Method 1 and Method 2, performing the best in all dimensions. The performance superiority of the different RL based methods at different dimensions can be attributed to their unique learning mechanisms (see Figure 2 and Pseudocodes 1 and 2). The RL_1 focuses on optimization of the 2D positioning error, and updates the height estimation only at the end of the learning process, while the RL_2 updates the height estimation in each action, which improves the height estimation accuracy but no further optimization in the horizontal plane. For Method 3, it first uses RL_2 to get a better estimation of the height and then uses RL_1 to optimize the rest two coordinates (see Pseudocode 2). For the CDF, the tested error is a continuous random variable. As we keep each measured point as an individual test, there are always some steps in the CDF curves. As shown in Figure 6, we always give the upper bound of test errors for the proposed RL based algorithms (i.e., Methods 1–3) but the lower bound of test errors for the Benchmark. It is a conservative way to show the benefits brought by the RL. More test points might help to estimate more accurate improvement but would not make the concluding results not true.

Figure 7a shows the mean 3D positioning error obtained by the Benchmark and Method 3 for $dis_{(1,2)} = 10/20/30/40$ cm at different heights. 80% confidence intervals of the positioning error are also given in Figure 7a (i.e., the vertical bars). The improvement of 3D positioning accuracy with RL is obvious. The upper bounds of Method 3 are even smaller than the lower bounds of the Benchmark. Under different distances between adjacent PDs, Method 3 obtains a mean 3D positioning error below 3.2 cm regardless of the size of the receiver. The results also indicate that the performance of the two methods varies randomly in small ranges with respect to the height of test plane. Figure 7b shows the mean 3D positioning error obtained by the Benchmark and Method 3 for $dis_{(1,2)} = 10/20/30/40$ cm in the entire test space. The average 3D positioning errors with different receiver sizes are within [2.51, 2.69] cm and [3.15, 4.02] cm for Method 3 and the Benchmark, respectively, revealing an obvious reduction of the average 3D positioning error by at least 20%. Moreover, it also clearly indicates that the positioning performance is more stable (i.e., less variation of positioning errors) when the RL is implemented.
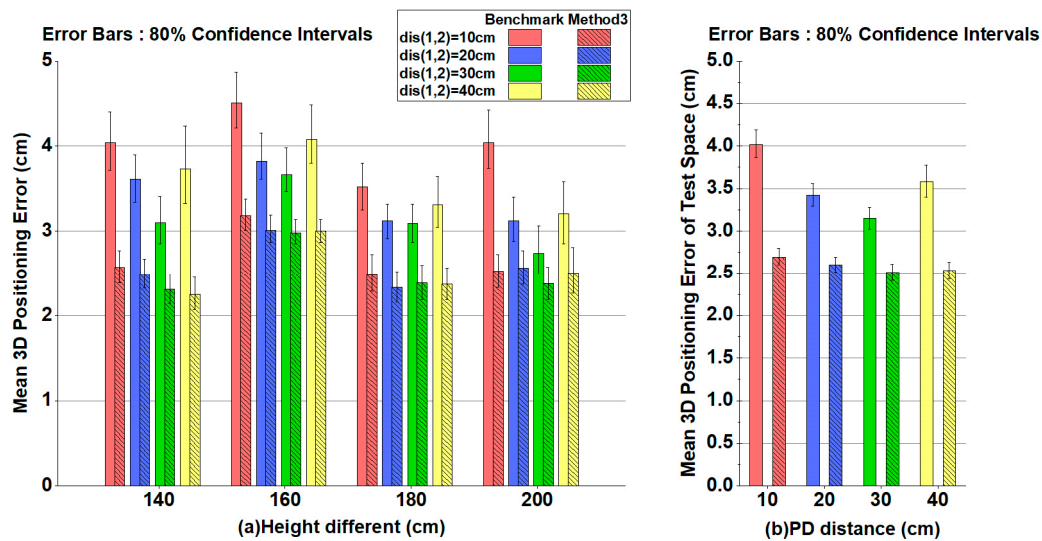
Figure 7. (**a**) Mean 3D positioning error at different heights for dis$_{(1,2)}$ = 10/20/30/40 cm with Benchmark/Method 3. (**b**) Mean 3D positioning error in the test space for dis$_{(1,2)}$ = 10/20/30/40 cm with Benchmark/Method 3.

4. Conclusions

A 3D indoor VLP system with reinforcement learning to enhance the positioning accuracy is proposed and experimentally investigated. The three proposed RL based methods share the *Agent-Environment* interaction framework with properly defined *State/Action/Reward*, but employ different height update strategies. The experimental results show that thanks to the learning process, all three RL based positioning methods outperform the Benchmark in terms of 3D positioning accuracy. The results also verify that Method 1 (Method 2) with RL_1 (RL_2) offers a significant improvement in the horizontal plane (height dimension) over the Benchmark. By combining RL_1 and RL_2, Method 3 offers the highest positioning accuracy not only in the 3D space but also in the height dimension and the horizontal plane, respectively. For the test planes with height difference from 140 cm to 200 cm, the mean 3D positioning error has been significantly improved (>20%) by Method 3 compared with the Benchmark. Moreover, the RL also reduces the variation of the 3D position error compared to the Benchmark with receivers of different sizes.

Author Contributions: Conceptualization, X.H. and J.C.; Data curation, W.Z. and X.C.; Funding acquisition, X.H. and J.C.; Writing—original draft, Z.Z. and H.C.; Writing—review & editing, X.H. and J.C. All authors have read and agreed to the published version of the manuscript.

References

1. Ling, R.W.C.; Gupta, A.; Vashistha, A.; Sharma, M.; Law, C.L. High precision UWB-IR indoor positioning system for IoT applications. In Proceedings of the 2018 IEEE 4th World Forum on Internet of Things (WF-IoT), Singapore, 5–8 February 2018; pp. 135–139.
2. Sthapit, P.; Gang, H.S.; Pyun, J.Y. Bluetooth based indoor positioning using machine learning algorithms. In Proceedings of the 2018 IEEE International Conference on Consumer Electronics—Asia (ICCE-Asia), Jeju, Korea, 24–26 June 2018; pp. 206–212.

3. Andrushchak, V.; Maksymyuk, T.; Klymash, M.; Ageyev, D. Development of the iBeacon's positioning algorithm for indoor scenarios. In Proceedings of the 2018 International Scientific-Practical Conference Problems of Infocommunications, Science and Technology (PIC S&T), Kharkiv, Ukraine, 9–12 October 2018; pp. 741–744.

4. Tang, S.; Tok, B.; Hanneghan, M. Passive indoor positioning system (PIPS) using near field communication (NFC) technology. In Proceedings of the 2015 International Conference on Developments of E-Systems Engineering (DeSE), Dubai, UAE, 12–15 December 2015; pp. 150–155.

5. Ultra-Wideband Technology for Indoor Positioning-by Infsoft. Available online: https://www.infsoft.com/technology/positioning-technologies/ultra-wideband (accessed on 30 October 2020).

6. Soner, B.; Ergen, S.C. Vehicular Visible Light Positioning with a Single Receiver. In Proceedings of the 2019 PIMRC, Istanbul, Turkey, 8–11 September 2019; p. 8904408.

7. Rahman, A.B.M.M.; Li, T.; Wang, Y. Recent advances in indoor localization via visible light: A survey. *Sensors* **2020**, *20*, 1382. [CrossRef] [PubMed]

8. Gu, W.; Aminikashani, M.; Kavehrad, M. Indoor visible light positioning system with multipath reflection analysis. In Proceedings of the 2016 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, 9–11 January 2016; pp. 89–92.

9. Wang, Q.; Luo, H.; Men, A.; Zhao, F.; Gao, X. Light positioning: A high-accuracy visible light indoor positioning system based on attitude identification and propagation model. *Int. J. Distrib. Sens. Netw.* **2018**, *14*, 1550147718758263. [CrossRef]

10. Komine, T.; Nakagawa, M. Fundamental analysis for visible-light communication system using LED lights. *IEEE Trans. Consum. Electron.* **2004**, *50*, 100–107. [CrossRef]

11. Lee, K.; Park, H.; Barry, J.R. Indoor channel characteristics for visible light communications. *IEEE Commun. Lett.* **2011**, *15*, 217–219. [CrossRef]

12. Yasir, M.; Ho, S.W.; Vellambi, B.N. Indoor positioning system using visible light and accelerometer. *J. Lightwave Technol.* **2014**, *32*, 3306–3316. [CrossRef]

13. Li, Q.L.; Wang, J.Y.; Huang, T.; Wang, Y. Three-dimensional indoor visible light positioning system with a single transmitter and a single tilted receiver. *Opt. Eng.* **2016**, *55*, 106103. [CrossRef]

14. Lam, E.W.; Little, T.D.C. Indoor 3D Location with Low-Cost LiFi Components. In Proceedings of the 2019 Global LIFI Congress (GLC), Paris, France, 12–13 June 2019; pp. 1–6.

15. Lam, E.W.; Little, T.D.C. Resolving height uncertainty in indoor visible light positioning using a steerable laser. In Proceedings of the 2018 IEEE International Conference on Communications Workshops, KansasCity, MO, USA, 20–24 May 2018; pp. 1–6.

16. Plets, D.; Almadani, Y.; Bastiaens, S.; Ljza, M.; Martens, L.; Joseph, W. Efficient 3D trilateration algorithm for visible light positioning. *J. Opt.* **2019**, *21*, 05LT01. [CrossRef]

17. Plets, D.; Bastiaens, S.; Ijza, M.; Almadani, Y.; Martens, L.; Reas, W.; Stevens, N.; Joseph, W. Three-dimensional visible light positioning: An Experimental Assessment of the Importance of the LEDs' Locations. In Proceedings of the 2019 International Conference on Indoor Positioning and Indoor Navigation (IPIN), Pisa, Italy, 30 September–3 October 2019; pp. 1–6.

18. Yang, S.H.; Kim, H.S.; Son, Y.H.; Han, S.K. Three-dimensional visible light indoor localization using AOA and RSS with multiple optical receivers. *J. Lightwave Technol.* **2014**, *32*, 2480–2485. [CrossRef]

19. Xu, Y.; Zhao, J.; Shi, J.; Chi, N. Reversed three-dimensional visible light indoor positioning utilizing annular receivers with multi-photodiodes. *Sensors* **2016**, *16*, 1245. [CrossRef] [PubMed]

20. Yasir, M.; Ho, S.W.; Vellambi, B.N. Indoor position tracking using multiple optical receivers. *J. Light. Technol.* **2016**, *34*, 1166–1176. [CrossRef]

21. Han, S.K.; Jeong, E.M.; Kim, D.R.; Son, Y.H.; Yang, S.H.; Kim, H.S. Indoor three-dimensional location estimation based on LED visible light communication. *Electron. Lett.* **2013**, *49*, 54–56.

22. Gu, W.; Kavehard, M.; Aminikashani, M. Three-dimensional indoor light positioning algorithm based on nonlinear estimation. In Proceedings of the SPIE OPTO, San Francisco, CA, USA, 13–18 February 2016; p. 97720V.

23. Liu, P.; Mao, T.; Ma, K.; Chen, J.; Wang, Z. Three-dimensional visible light positioning using regression neural network. In Proceedings of the 2019 15th International Wireless Communications and Mobile Computing Conference (IWCMC), Tangier, Morocco, 24–28 June 2019; pp. 156–160.

24. Alonso-González, I.; Sánchez-Rodríguez, D.; Ley-Bosch, C.; Quintana-Suárez, M. Discrete indoor three-dimensional localization system based on neural networks using visible light communication. *Sensors* **2018**, *18*, 1040. [CrossRef] [PubMed]

25. Zhang, S.; Du, P.; Chen, C.; Zhong, W.D. 3D indoor visible light positioning system using RSS ratio with neural network. In Proceedings of the 2018 23rd Opto-Electronics and Communications Conference (OECC), Jeju, Korea, 2–6 July 2018; pp. 1–2.

26. Jiang, J.; Guan, W.; Chen, Z.; Chen, Y. Indoor high-precision three-dimensional positioning algorithm based on visible light communication and fingerprinting using K-means and random forest. *Opt. Eng.* **2019**, *58*, 1. [CrossRef]

27. Xu, S.; Chen, C.C.; Wu, Y.; Wang, X.; Wei, F. Adaptive residual weighted K-nearest neighbor fingerprint positioning algorithm based on visible light communication. *Sensors* **2020**, *20*, 4432. [CrossRef] [PubMed]

28. Zhang, Z.; Chen, H.; Hong, X.; Chen, J. Accuracy enhancement of indoor visible light positioning using point-wise reinforcement learning. In Proceedings of the 2019 Optical Fiber Communications Conference and Exposition (OFC), San Diego, CA, USA, 3–7 March 2019.

29. Zhang, Z.; Zhu, Y.; Zhu, W.; Chen, H.; Hong, X.; Chen, J. Iterative point-wise reinforcement learning for highly accurate indoor visible light positioning. *Opt. Express* **2019**, *27*, 22161–22172. [CrossRef] [PubMed]

30. Zhang, W.; Chowdhury, M.I.S.; Kavehrad, M. Asynchronous indoor positioning system based on visible light communications. *Opt. Eng.* **2014**, *53*, 045105. [CrossRef]

31. Guo, X.; Shao, S.; Ansari, N.; Khreishah, A. Indoor localization using visible light via fusion of multiple classifiers. *IEEE Photonics J.* **2017**, *9*, 1–16. [CrossRef]

32. Milioris, D. Efficient indoor localization via reinforcement learning. In Proceedings of the 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 8350–8354.

33. Anthropometry | Sustainable Ergonomics Systems—Ergoweb. Available online: https://ergoweb.com/anthropometry/ (accessed on 30 October 2020).