



Iterative point-wise reinforcement learning for highly accurate indoor visible light positioning

Downloaded from: <https://research.chalmers.se>, 2026-04-04 03:48 UTC

Citation for the original published paper (version of record):

Zhang, Z., Zhu, Y., Zhu, W. et al (2019). Iterative point-wise reinforcement learning for highly accurate indoor visible light positioning. *Optics Express*, 27(16): 22161-22172.

<http://dx.doi.org/10.1364/OE.27.022161>

N.B. When citing this work, cite the original published paper.



Iterative point-wise reinforcement learning for highly accurate indoor visible light positioning

ZHUO ZHANG,¹ YAGUANG ZHU,¹ WENTAO ZHU,¹ HUAYANG CHEN,¹ XUEZHI HONG,^{1,4} AND JIAJIA CHEN^{1,2,3,5}

¹South China Academy of Advanced Optoelectronics, South China Normal University, Guangzhou 510006, China

²KTH Royal Institute of Technology, Isafjordsgatan 22, Electrum 229, 164 40 Kista, Sweden

³Chalmers University of Technology, Hörsalsvägen 9, 412 96 Gothenburg, Sweden

⁴xuezhi.hong@coer-scnu.org

⁵jjaiac@chalmers.se

Abstract: Iterative point-wise reinforcement learning (IPWRL) is proposed for highly accurate indoor visible light positioning (VLP). By properly updating the height information in an iterative fashion, the IPWRL not only effectively mitigates the impact of non-deterministic noise but also exhibits excellent tolerance to deterministic errors caused by the inaccurate *a priori* height information. The principle of the IPWRL is explained, and the performance of the IPWRL is experimentally evaluated in a received signal strength (RSS) based VLP system and compared with other positioning algorithms, including the conventional RSS algorithm, the *k*-nearest neighbors (KNN) algorithm and the PWRL algorithm where iterations exclude. Unlike the supervised machine learning method, e.g., the KNN, whose performance is highly dependent on the training process, the proposed IPWRL does not require training and demonstrates robust positioning performance for the entire tested area. Experimental results also show that when a large height information mismatch occurs, the IPWRL is able to first correct the height information and then offers robust positioning results with a rather low positioning error, while the positioning errors caused by the other algorithms are significantly higher.

© 2019 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

1. Introduction

The indoor positioning system (IPS) offers localization service complementing the global positioning system, which is often unavailable inside the building. The conventional IPS uses radio frequency (RF) technologies, such as RFID, Wi-Fi, ZigBee and Bluetooth [1–4], which generally have relatively low positioning accuracy (e.g., few meters in [4]) due to the carrier fading and are vulnerable to electromagnetic interference (EMI). On the other hand, accurate positioning is critical for location based services, such as navigation or location-based advertising on mobile devices, particularly for the indoor case. In this regard, visible light positioning (VLP) using optical carriers from the ubiquitous lighting systems (e.g., light emitting diodes LEDs) is an attractive solution for the IPS. The VLP overcomes the disadvantages of the RF technologies offering relatively high positioning accuracy [5] and is immune to the EMI [6].

In a VLP system, the light sources are pre-installed for illumination serving as beacons [7]. The receiver, which consists of one detector [8] or multiple detectors [9], converts the optical signal from the beacons into the electrical signal and estimates its position according to certain positioning algorithms. The VLP can be considered as a special application of the visible light communication, where the positioning algorithms translate the information in the received signal (e.g., received signal strength RSS [10]) into the position of the receiver by exploring the uniqueness of the optical channel between the transmitter (i.e., beacon) and

receiver (i.e., detector). Therefore, the positioning accuracy is affected by two types of errors: 1) the noise/interference during the signal measurements (i.e., non-deterministic error) and 2) the inaccurate signal-to-position interpretation (i.e., deterministic error). It has been found that a receiver based on multiple detectors outperforms the one based on a single detector in terms of inter-cell interference mitigation [11]. Moreover, to improve the positioning accuracy, machine learning (ML) algorithms, especially supervised learning (SL), have been introduced to the VLP [12], such as k -nearest neighbors (KNN) [13], back-propagation [14], random forest based classifiers and adaBoost based classifiers [15]. However, performance of the SL assisted VLP systems is largely affected by the training data. For instance, the number of offline training samples and the spatial distribution of the sampling points [16], etc. may significantly impact the positioning results. To get rid of the sophisticated training phase, reinforcement learning (RL), which maximizes the expected benefits by emphasizing how should the *Agent* acts based on the *Environment* knowledge [17], has been introduced to the VLP system [18,19]. In our previous study, a point-wise reinforcement learning (PWRL) based VLP system has been demonstrated [20], which reduces the non-deterministic noise by the RL point by point with the *Agent*.

In this paper, we extend the work presented in [20], and propose iterative point-wise reinforcement learning (IPWRL) for further improvement of accuracy in the VLP. The IPWRL is designed to compensate not only the non-deterministic noise, as that is already done in the PWRL, but also the deterministic noise caused by inaccurate *a priori* information of system parameters. By implementing the PWRL in an iterative fashion and updating the inaccurate parameters, i.e., the height difference of the receiver and LEDs, properly, the positioning error can be reduced significantly. Experimental investigations are conducted in a VLP system that is able to measure RSS to evaluate the positioning performance of the IPWRL. A comparison among the proposed IPWRL, the conventional RSS [10], the KNN [13], and the PWRL [20], is carried out. Our results reveal that the IPWRL is able to maintain the low positioning errors even when a large height difference is introduced, showing excellent robustness against deterministic errors, while the positioning errors of the other methods increase sharply.

2. Operation principle

A multi-detector VLP system is considered having a receiver with N detectors and M ($M \geq 3$) LEDs that are all on the ceiling and hence are assumed in the same height. The i th LED is located at position (L_i^x, L_i^y, L^z) and transmits sinusoidal modulated signal with frequency f_i . The RSS at frequency f_i is used to estimate the distance between the detector and the i th LED. The received signal $s_n(t)$ of the n th detector at $(x_n, y_n, L^z - h)$ from all the LEDs can be expressed as [10,21]:

$$s_n(t) = \sum_{i=1}^M \frac{(m+1)A}{2\pi d_{n,i}^2} \cos^m(\varphi) \cos^{m'}(\psi) \beta p_i(t - \tau_i) + w(t), \quad (1)$$

in which A is the detector area, $d_{n,i}$ is the distance between the i th LED and the n th detector, β is the detector responsivity, $w(t)$ denotes the noise, m (m') is the Lambertian radiation pattern order of the LED (detector), φ and ψ are the radiation angle and incidence angle, respectively. $p_i(t)$ is the direct-current (DC) biased and windowed sinusoid waveform, where the time delay $\tau_i = d_{n,i}/c$ and c is the speed of light in vacuum.

The power spectrum of $s_n(t)$ consists of M peak components at f_i ($i = 1, 2, \dots, M$). Assuming the detector is facing up $\cos(\varphi) = \cos(\psi) = h / d_{n,i}$, the RSS of these components obtained by the N detectors can be represented by a $M \times N$ vector **Rec**:

$$\mathbf{Rec} = \{\text{peaks of } \mathcal{F}(s_n(t))\}_{n=1,2,\dots,N} = [S_1(f_1), \dots, S_1(f_M), \dots, S_N(f_1), \dots, S_N(f_M)], \quad (2)$$

where

$$S_n(f_i) = \frac{(m+1)^2 A^2 \beta^2 h^{2(m+m')}}{4\pi^2 d_{n,i}^{2(2+m+m')}} \quad (3)$$

and $\mathcal{F}(\cdot)$ denotes the Fourier transform. Each RSS $S_n(f_i)$ at frequency f_i is determined by $d_{n,i}$ according to Eq. (3). According to the location of the LED and detector, we have

$$(x_n - L_i^x)^2 + (y_n - L_i^y)^2 + h^2 = d_{n,i}^2. \quad (4)$$

In the conventional RSS positioning algorithm [10], the positions of the detectors are determined by trilateration with Eqs. (2)-(4) and the least square (LS) estimation [22]. The receiver position is obtained by averaging the estimated detectors' locations.

2.1 Point-wise reinforcement learning

The accuracy of the above positioning process is affected by both deterministic and non-deterministic errors. The deterministic one reflects the impact of inaccurate information of system parameters, e.g., h , A , L_i^x , L_i^y , $\phi_{1/2}$ and $\psi_{1/2}$. These parameters can be obtained from either the datasheet of the device, e.g., the detector area A or additional measurements before the positioning process, e.g., the height difference between the receiver and LEDs h , the half power angle of the LEDs (detectors) $\phi_{1/2}$ ($\psi_{1/2}$), the location of LED i (L_i^x , L_i^y , L_i^z). As can be seen from Eqs. (1) and (3), the inaccurate *a priori* information might cause errors when estimating the distance between the detector and LEDs d , which in turn might impact the positioning results. The non-deterministic one is mainly referred to as shot noise and thermal noise, which exists in any practical VLP system. The noise included in **Rec** varies for different points. Assuming that accurate system parameters are provided, a point-wise reinforcement learning algorithm has been devised to mitigate the impact of non-deterministic noise in the RSS measurements [20]. Figure 1 shows the schematic diagram of the PWRL algorithm. The *Environment*, i.e., unknown RSS error, is learned by the *Agent* via interactions without training, where the *Environment* rewards and stimulates the *Agent* at a certain state to take the right action.



Fig. 1. Schematic diagrams of point-wise reinforcement learning.

To appropriately define the states and rewards, we first denote the actual (calculated) $N(N-1)/2$ relative distances among N detectors as \mathbf{dis}_{real} (\mathbf{dis}_{calc}), which can be expressed as

$$\begin{aligned} \mathbf{dis}_{real} &= (dis_{real12}, dis_{real13}, \dots, dis_{real1N}, \dots, dis_{real(N-1)N}) \\ \mathbf{dis}_{calc} &= (dis_{calc12}, dis_{calc13}, \dots, dis_{calc1N}, \dots, dis_{calc(N-1)N}), \end{aligned} \quad (5)$$

where \mathbf{dis}_{realij} (\mathbf{dis}_{calcij}) denotes the real (calculated) distance between the i th and j th detectors. The relative distance error vector \mathbf{dis}_{error} is defined as the difference between \mathbf{dis}_{real} and \mathbf{dis}_{calc} :

$$\mathbf{dis}_{error} = \mathbf{dis}_{real} - \mathbf{dis}_{calc}. \quad (6)$$

Then the G states and K rewards are defined according to the maximum and average values of dis_{error} , respectively:

$$State = i, \text{ if } \alpha_{i-1} < \max(dis_{error}) \leq \alpha_i \quad \text{for } 1 \leq i \leq G, \quad (7)$$

$$Reward = \frac{K-i}{K-1}, \text{ if } r_{i-1} < \text{average}(dis_{error}) \leq r_i \quad \text{for } 1 \leq i \leq K, \quad (8)$$

in which $(\alpha_1, \alpha_2, \dots, \alpha_G)$ and (r_1, r_2, \dots, r_K) are predefined constants. The *Agent* judges whether the dis_{error} is in the target state (e.g., $State = 1$). If not, the *Agent* increases/decreases the i th RSS element of Rec in Eq. (3) by a certain $step$, which is denoted as the $(2i-1)$ -th and $(2i)$ -th actions, respectively. For example, the output of the first and second of the $2M \times N$ actions are:

$$Rec_{new_1} = [S_1(f_1) + step, S_1(f_2), \dots, S_1(f_M), S_2(f_1), \dots, S_2(f_M), \dots, S_N(f_1), \dots, S_N(f_M)], \quad (9)$$

$$Rec_{new_2} = [S_1(f_1) - step, S_1(f_2), \dots, S_1(f_M), S_2(f_1), \dots, S_2(f_M), \dots, S_N(f_1), \dots, S_N(f_M)], \quad (10)$$

With RSS in Eqs. (9) or (10), new dis_{calc} is obtained and the dis_{error} is updated according to Eq. (6). The new states and instant rewards are then obtained according to Eqs. (7) and (8), respectively. After testing all possible actions, the *Agent* chooses the action with the maximum reward, which completes an episode of the RL. The learning process continues until the target state or the upper limit of the cycles. After learning, a finalized RSS vector Rec^{PWRL} is obtained and used to calculate the position of detectors/receiver by trilateration.

2.2 Iterative point-wise reinforcement learning

Previous results show that the PWRL is effective in mitigating the influence of non-deterministic noise when the given system parameters are accurate [20]. Nevertheless, the impact of deterministic error on positioning accuracy has not been tackled. One widely existing factor contributed to the deterministic error in the practical VLP systems is the uncertainty of height. For example, the height of receiver on handheld devices might vary when the user changes the height of hands unconsciously or purposely. In other words, in many cases the height difference between the detector and the LEDs could randomly change within a certain range rather than being a fixed value, which degrades the positioning accuracy. To compensate for such deterministic errors caused by *a priori* height information, we propose to use the PWRL iteratively. The schematic diagrams of the proposed iterative point-by-point reinforcement learning algorithm (IPWRL) is shown in Fig. 2.

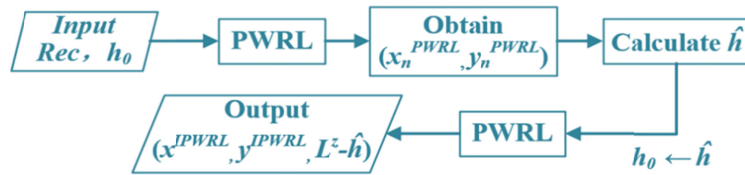


Fig. 2. Schematic diagrams of iterative point-wise reinforcement learning.

Assuming that the exact height of the receiver is unknown, the height difference between the receiver and LEDs is set to be h_0 in the initial stage as the input. With the PWRL algorithm, the projected 2-D position of the n th detector is estimated as (x_n^{PWRL}, y_n^{PWRL}) . The height difference between the i th LED and the n th detector is then updated as $h_{n,i}$ according to the following equation:

$$h_{n,i} = \sqrt{\hat{d}_{n,i}^2 - (x_n^{PWRL} - L_i^x)^2 - (y_n^{PWRL} - L_i^y)^2}, \quad (11)$$

where $d_{n,i}$ can be obtained with $S_n(f_i)$ as:

$$\hat{d}_{n,i}^2 = \left[\frac{(m+1)^2 A^2 \beta^2 h_0^{2(m+m')}}{4\pi^2 S_n(f_i)} \right]^{\frac{1}{(2+m+m')}} \quad (12)$$

A set of $M \times N$ equations can be established according to Eq. (11). To get an updated height difference between the receiver and LEDs \hat{h} , two different methods are proposed. The first method separates the $M \times N$ equations into N subsets on a per-detector basis by assuming $h_{n,i} = h_n$. After obtaining the estimation of h_n in each subsets with LS estimation, \hat{h} is calculated as the averaged value of the N estimations, i.e., $\hat{h} = \frac{1}{N} \sum_{n=1}^N h_n$. The second method assumes $h_{n,i} = \hat{h}$, and uses LS estimation based on the $M \times N$ equations to get \hat{h} . Hereafter, we refer IPWRL1 (IPWRL2) to as the IPWRL employing the first (second) method to get \hat{h} .

By replacing h_0 with \hat{h} , we employ the PWRL algorithm again and obtain a updated position of n th detector as $(x_n^{IPWRL}, y_n^{IPWRL})$. The position of receiver is then calculated by averaging the estimated coordinates of the detectors $(x^{IPWRL} = \frac{1}{N} \sum_{n=1}^N x_n^{IPWRL}, y^{IPWRL} = \frac{1}{N} \sum_{n=1}^N y_n^{IPWRL}, L^z - \hat{h})$. The pseudocode of the IPWRL algorithm is shown in Table 1.

Table 1. Pseudocode for the IPWRL algorithm.

Initialization:	16.	Obtain dis_{error} using Eq. (6)
1. Set the target state	17.	Update the state and reward
2. Set G states according to Eq. (7)	18.	end for
3. Set K rewards according to Eq. (8)	19.	Calculate the projected 2-D coordinate of N detectors (x_n^{PWRL}, y_n^{PWRL}) and corresponding Rec^{PWRL}
Main:		
4. Input: the RSS vector Rec, h_0		
5. Output: Coordinate of receiver $(x^{IPWRL}, y^{IPWRL}, L^z - \hat{h})$.	20.	Obtain the coordinate of receiver $(x^{PWRL}, y^{PWRL}, L^z - h_0)$ by averaging the coordinate of detectors
7. Obtain dis_{error} using (6)	21.	Calculate \hat{h} with Eqs. (11) and (12) for IPWRL1 (IPWRL2)
8. Obtain the current state and reward		
9. $k = 0$	22.	Update the difference in height $h_0 \leftarrow \hat{h}$
10. for $k <$ the upper limit of the cycles do	23.	Reuse PWRL algorithm (i.e., run Steps 7-18)
11.	$k \leftarrow k + 1$	24. Calculate the projected 2-D coordinate of N detectors $(x_n^{IPWRL}, y_n^{IPWRL})$ and corresponding Rec^{IPWRL}
12.	For the current state does not reach the target state and the reward is not less than the last one do	
13. Obtain $2M \times N$ new RSS vectors Rec for $2M \times N$ actions	25.	Obtain the coordinate of receiver $(x^{IPWRL}, y^{IPWRL}, L^z - \hat{h})$ by averaging the coordinate of detectors
14.	for each new Rec do	
15.	Calculate the new coordinate of the n th detectors	26. Return $(x^{IPWRL}, y^{IPWRL}, L^z - \hat{h})$

3. Experimental investigation

3.1 Experiment setup

The experimental setup of the investigated VLP system and the data processing flow are shown in Fig. 3. The overall size of our experimental platform is $120\text{ cm} \times 120\text{ cm} \times 120\text{ cm}$, where four sinusoid signals of different frequencies (400/500/600/700 kHz) are first generated by four signal generators, and then combined with the DC signals by bias-tees, respectively. For simplicity, the signal from the four LEDs are distinguished by the signal frequency. The four LEDs are on the ceiling at (21.9, 20.8, 120), (76.9, 18.4, 120), (20.1, 80.5, 120), (81.6, 79.2, 120) in cm, respectively. The considered receiver module has 4 detectors (PDA100A2) situated in four corners of a square, where the edge (i.e., dis_{real12}) can be adjusted. Note due to limited conditions for experimental setup, the size of the used detectors is relatively large, so dis_{real12} is difficult to set less than 10 cm. For practice, it may fit applications with large-size user equipment, such as a tablet, low-speed indoor vehicles. The height of the receiving plane is set to 17.95 cm, which means that the real height difference between transmitter and receiver is 102.05 cm. We used the method in [10] to experimentally measure the half power angle of the LEDs (detectors), and obtained the values of m (m') as 1.68 (3.57).

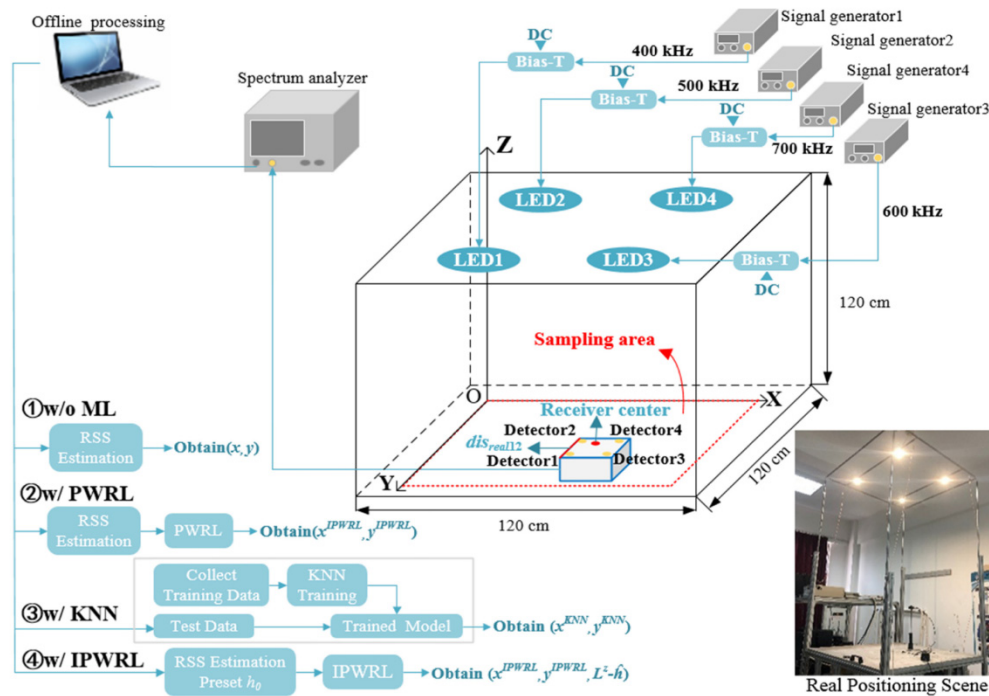


Fig. 3. Experimental setup of the VLP system and the corresponding data processing flow.

Taking into account the size of the receiver module cannot be ignored, the sampling area is set to $70\text{ cm} \times 70\text{ cm}$. A spectrum analyzer is used to measure the RSS vector at 49 different points (i.e., Fig. 4(a) shows the sampling points of Detector 1), which is used as input to the IPWRL algorithm. For comparison, the above test samples are also the input into the conventional RSS algorithm [10], the PWRL algorithm [20], and the KNN algorithm [13]. In order to collect training data for the KNN algorithm, we take 49 points at the same height for three times, 25 of which coincide with the points corresponding to test samples (i.e., 25 red points shown in Fig. 4(b)).

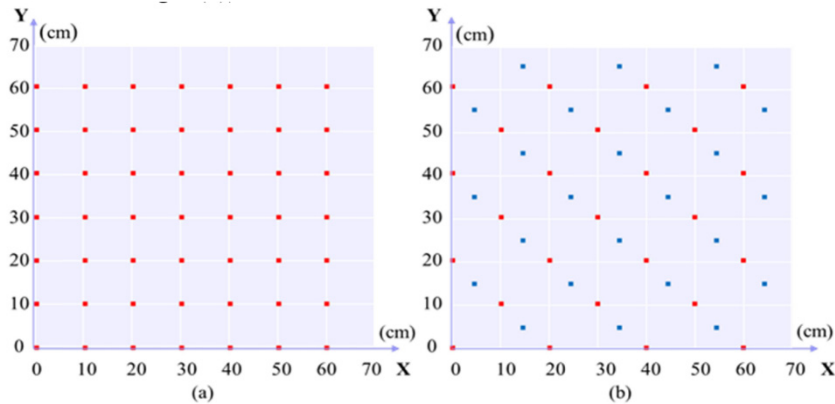


Fig. 4. Sampling points of Detector 1: (a) input for testing different positioning algorithms and (b) training data for the KNN. The 25 red points in (b) coincide with some test points shown in (a).

3.2 Performance investigation

Figure 5(a) shows the mean positioning error with the conventional RSS algorithm (i.e., without any ML algorithm) and with the PWRL, KNN, IPWRL1 and IPWRL2 as a function of dis_{real12} (i.e., 10/20/30/40 cm) when the height information is considered accurate. In Fig. 5, h_0 in the IPWRL is 102.05 cm. The gain achieved by using reinforcement learning is obvious, regardless of the distance between detectors. When dis_{real12} is 10/20/30/40 cm, the mean positioning error is reduced from 2.34/2.46/2.65/2.75 cm to 2.24/2.07/1.94/2.01 cm by replacing the conventional RSS algorithm with the PWRL. In contrast, the performance of the KNN is not as robust as that of the PWRL or IPWRL, although is better than that of the conventional RSS algorithm at $dis_{real12} = 30/40$ cm. This may be due to fact that the samples obtained by the closely located detectors (i.e., dis_{real12} of 10/20 cm) are more correlated. The reduction in the diversity of samples makes it more difficult to find the correct samples in the KNN. The two IPWRL algorithms are better than the KNN and the conventional RSS, but slightly worse than the PWRL. For both the PWRL and IPWRL, the improvement over the conventional algorithm becomes more significant when increasing dis_{real12} . The performance gap between the IPWRL and PWRL shrinks for a larger value of dis_{real12} and becomes negligible when $dis_{real12} = 40$ cm.

Figure 5(b) is the cumulative distribution function (CDF) of the positioning error when $dis_{real12} = 40$ cm. The corresponding spatial distribution is shown in Fig. 5(c). Without any ML algorithm the mean positioning error is 2.75 cm, and 80% of samples have errors within [2.66 cm, 2.85 cm]. When the PWRL is employed, 80% of samples have errors within [1.85 cm, 2.16 cm] and the mean positioning error is reduced to 2.01 cm, leading to a reduction of 27%. The performance of the two IPWRL algorithms is similar as that of the PWRL: the corresponding mean positioning errors are 2.02 cm and 2.03 cm, and 80% of the sample errors are within the range of [1.89 cm, 2.16 cm] and [1.90 cm, 2.17 cm], respectively. Though in Fig. 5(b) a considerable number of points can get zero error by using the KNN algorithm, the results in Fig. 5(c) show that zero error can only be achieved when some of the test points coincide with the samples used in the training phase (i.e., red points in Fig. 4(b)). In contrast, the positioning error of those points that do not coincide with sampling points in the training phase are significantly higher. Therefore, the performance of the KNN is largely dependent on the training samples. It requires enough training data captured in the points that are the same or close to the test points, resulting in a higher implementation complexity. Compared with the KNN, the PWRL and IPWRL algorithms do not require any training process and show robust performance across the whole tested area.

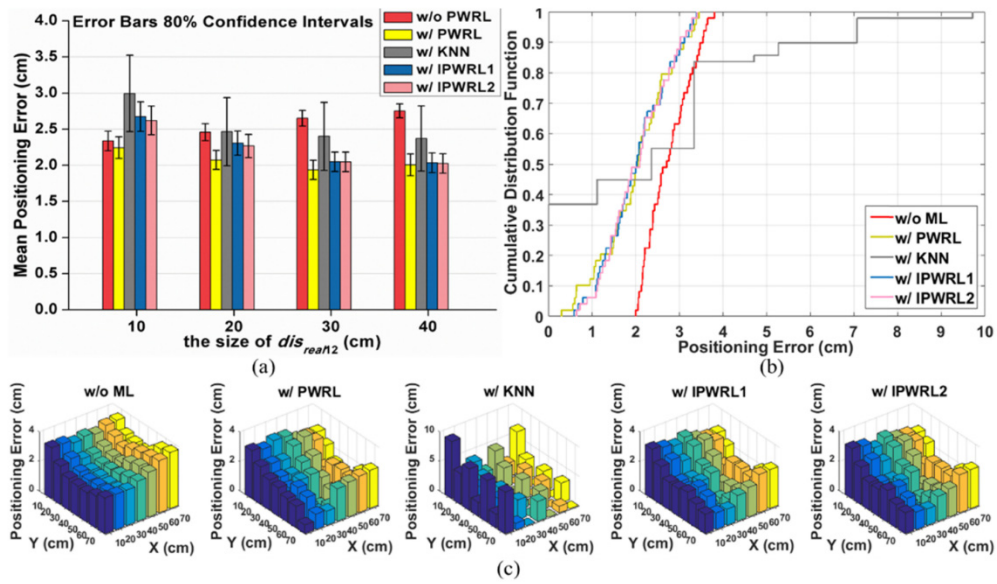


Fig. 5. (a) Positioning error versus dis_{rea12} (cm) for different positioning algorithms; (b) the cumulative distribution function and (c) spatial distribution of the positioning error ($dis_{rea12} = 40$ cm).

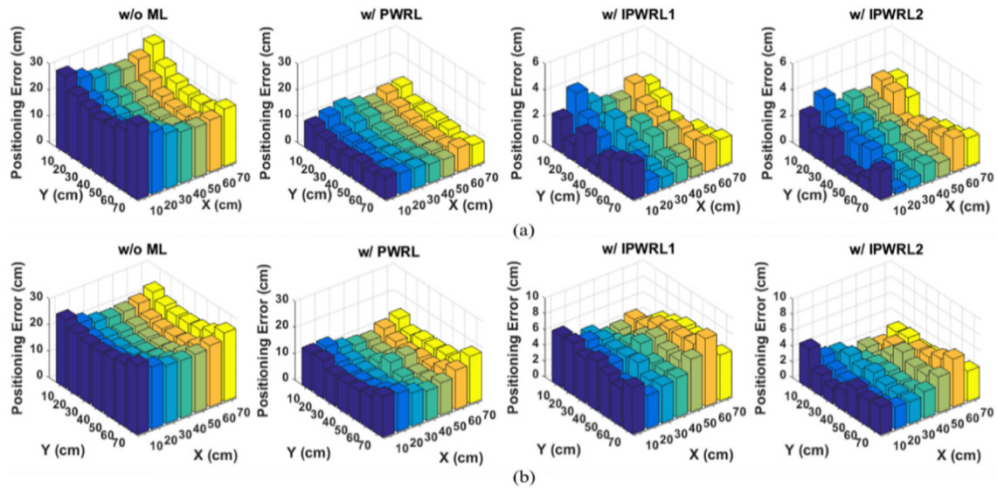


Fig. 6. Spatial distribution of the positioning error when $dis_{rea12} = 40$ cm with (a) $h_0 = 51.05$ cm and (b) $h_0 = 153.05$ cm.

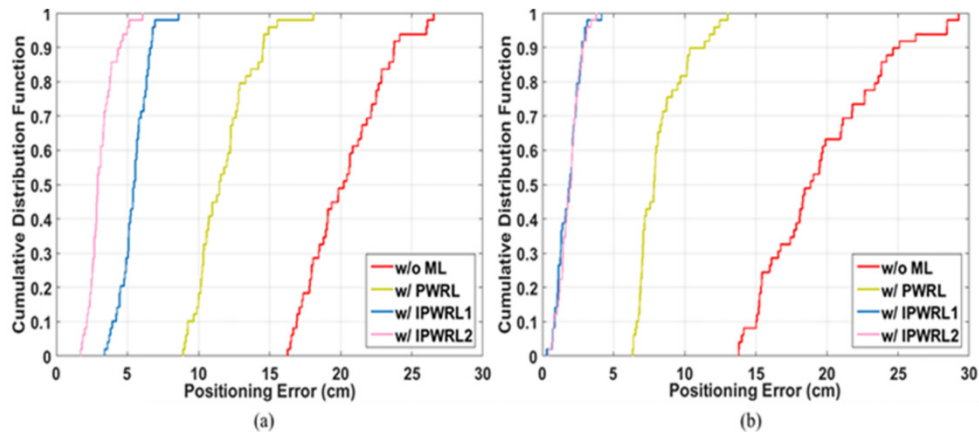


Fig. 7. Cumulative distribution function of the positioning error when $dis_{real/12} = 40$ cm with (a) $h_0 = 51.05$ cm and (b) $h_0 = 153.05$ cm.

It is also expected that the IPWRL performs similarly as the PWRL when the system parameters are accurate. However, such a conclusion no longer holds when there is some uncertainty existing in height information. We first assumed the input height difference is 51.05 cm or 153.05 cm instead of the correct value of 102.05 cm. The spatial distribution and CDF of positioning error of the conventional RSS algorithm, PWRL and (IPWRL12) are shown in Figs. 6 and 7, when the h_0 is set to 51 cm larger (i.e., Figs. 6(a) and 7(a)) or smaller (i.e., Figs. 6(b) and 7(b)) than the actual height difference h_{real} . The KNN algorithm does not take into account the height for positioning and cannot be robust to height difference. Therefore, we exclude it for comparison here. The results in Figs. 6 and 7 reveal that the positioning accuracy decreases sharply for the conventional algorithm. In contrast, the PWRL algorithm largely reduces the positioning error due to mismatched height information. Specifically, when h_0 is 153.05 cm (51.05 cm), the PWRL can improve the mean positioning error from 19.52 cm (20.34 cm) to 8.24 cm (11.76 cm). It is obvious that the IPWRL offers even better performance than that of the PWRL. Figure 7 further shows that the results of the IPWRL1 and IPWRL2 algorithms are similar for $h_0 = 153.05$ cm, which can reduce the mean positioning error to 1.84 cm and 1.90 cm, respectively. For $h_0 = 51.05$ cm, the IPWRL2 shows better performance, which reduces the mean positioning error to 3.14 cm, and 80% of the positioning error is less than 3.8 cm. While the IPWRL1 can reduce the mean positioning error to 5.45 cm, and 80% of the positioning error is less than 6.4 cm.

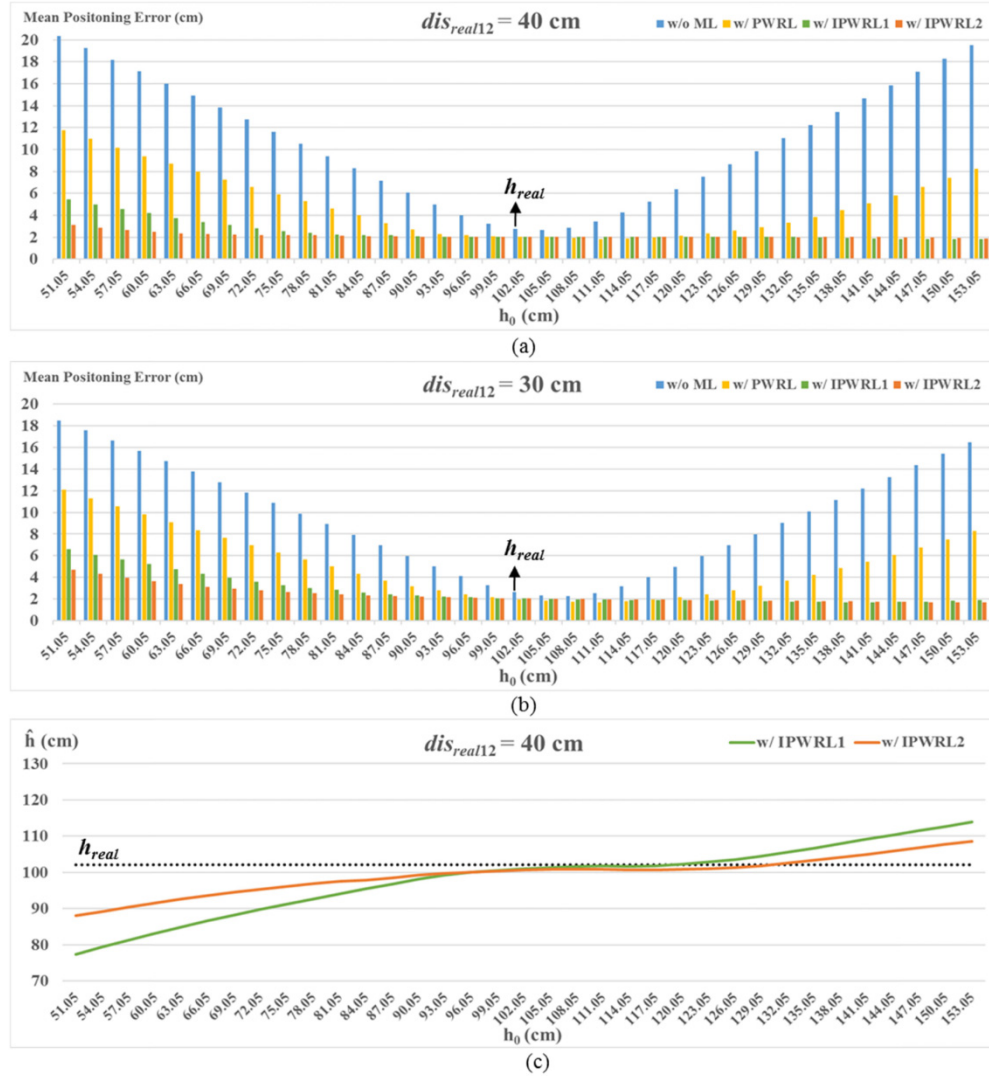


Fig. 8. Positioning error versus errors of height difference for the dis_{real12} of (a) 40 cm and (b) 30 cm, and (c) the estimated \hat{h} by the two IPWRL algorithms versus different h_0 for the dis_{real12} of 40 cm.

Without loss of generality, we further compare the performance of different positioning algorithms by adjusting h_0 in a step of 3 cm within the range of [51.05 cm, 153.05 cm]. The measured mean positioning error are shown in Figs. 8(a) and 8(b) for $dis_{real12} = 40$ cm and $dis_{real12} = 30$ cm, respectively. Figure 8(c) shows the estimated \hat{h} by the two IPWRL algorithms versus different h_0 for $dis_{real12} = 40$ cm. Both the PWRL and IPWRL algorithms offer higher positioning accuracy than the conventional one for all cases. The enhancement of accuracy is more significant for a larger mismatch between h_0 and h_{real} . When h_0 is close to h_{real} , both IPWRL and PWRL can achieve a mean positioning error of ~ 2 cm regardless of the value of dis_{real12} . Increasing the gap between h_0 and h_{real} , the performance of the IPWRL is obviously better than that of the PWRL. Within the whole tested area, the mean positioning error can be reduced to about 5 cm by both of the IPWRL algorithms, while the performance of the PWRL degrades quickly at the boundary of the tested range. The positioning error of

the PWRL exceeds 5 cm for $h_0 > 141.05$ cm (141.05 cm) and $h_0 < 78.05$ cm (81.05 cm) for the dis_{real12} of 40 cm (30 cm). Therefore, with the help of the iteration, the IPWRL offers obviously higher tolerance to the height information mismatch. The IPWRL2 has almost the same performance as the IPWRL1 when the assumed height difference in the range [93.05 cm, 153.05cm], but outperforms when the input height difference in the range [51.05 cm, 93.05 cm]. This can be explained by the fact that the estimated \hat{h} by using the IPWRL2 is closer the real height difference h_{real} than the IPWRL1, which is clearly shown in the Fig. 8(c). The advantage of the IPWRL2 is more obvious when the difference between h_{real} and h_0 is larger.

The IPWRL can be implemented by running the PWRL twice in a iterative fashion, where some parameters in the first iteration are updated and then employed in the second one. It is obvious the iterations in IPWRL linearly increase the total running time. It should be noted that the additional running time compared to the PWRL may decrease when a larger inaccuracy is introduced in height information. The additional complexity can be further reduced, which calls for a future study of algorithm optimization.

4. Conclusion

In this paper, we have proposed iterative point-wise reinforcement learning for high-accuracy indoor VLP systems. By using the PWRL twice in an iterative fashion, the IPWRL is able to compensate the positioning errors caused by the inaccurate height information as well as shot noise and thermal noise. Experimental results verify that the IPWRL inherits the advantage of the PWRL that outperforms the conventional RSS algorithm in terms of positioning accuracy, and the KNN algorithm in terms of robust performance without the need of training data. The results also show that when the height information mismatch is large, the proposed IPWRL maintains the mean positioning error low (~ 5 cm), $\sim 75\%$ and $\sim 58\%$ lower than that achieved by the conventional RSS algorithm and PWRL algorithm, respectively.

Through simulations, it is found that the inaccurate LEDs' locations could introduce impact similar as that did by the height difference. In the future, to make IPWRL suitable for abundant scenarios, we will enhance the IPWRL to address the accuracy issues caused by other deterministic errors (e.g., the inaccurate LEDs' locations), while improving computational complexity in order to adapt to rapidly changed parameters. In addition, we have verified by simulation that the gain still exists regardless how short the distance between detectors is and will carry out a future work to further validate this finding by experiments.

Funding

The Swedish Foundation for Strategic Research, the Göran Gustafsson Stiftelse, the Swedish Research Council, the Swedish ICT-TNG, National Natural Science Foundation of China (NSFC) (61605047, 61671212, 61550110240), and the Natural Science Foundation of Guangdong Province (2016A030313438).

References

1. P. Bahl and V. N. Padmanabhan, "RADAR: An in-building RF-based user location and tracking system," in *Proceedings of IEEE INFOCOM*, 775–784 (2000).
2. Y. Zhuang, Z. Syed, Y. Li, and N. El-Sheimy, "Evaluation of two WiFi positioning systems based on autonomous crowdsourcing of handheld devices for indoor navigation," *IEEE Trans. Mobile Comput.* **15**(8), 1982–1995 (2016).
3. S. Fang, C. Wang, T. Huang, C. Yang, and Y. Chen, "An enhanced ZigBee indoor positioning system with an ensemble approach," *IEEE Commun. Lett.* **16**(4), 564–567 (2012).
4. Y. Zhuang, J. Yang, Y. Li, L. Qi, and N. El-Sheimy, "Smartphone-based indoor localization with Bluetooth low energy beacons," *Sensors (Basel)* **16**(5), 596 (2016).
5. H. Hosseinianfar, M. Noshad, and M. Brandt-Pearce, "Positioning for visible light communication system exploiting multipath reflections," in *Proceedings of 2017 IEEE International Conference on Communications (ICC)*, Paris, 1–6 (2017).

6. H. Burchardt, N. Serafimovski, D. Tsonev, S. Videv, and H. Haas, "VLC: beyond point-to-point communication," *IEEE Commun. Mag.* **52**(7), 98–105 (2014).
7. J. Luo, L. Fan, and H. Li, "Indoor positioning systems based on visible light communication: state of the art," *IEEE Comm. Surv. and Tutor.* **19**(4), 2871–2893 (2017).
8. F. Seguel, N. Krommenacker, P. Charpentier, and I. Soto, "Visible light positioning based on architecture information: method and performance," *IET Commun.* **13**(7), 848–856 (2019).
9. Y. Liu, K. Park, B. S. Ooi, and M. Alouini, "Indoor localization using three dimensional multi-PDs receiver based on RSS," in *2018 IEEE Globecom Workshops*, Abu Dhabi, United Arab Emirates, 1–6 (2018).
10. Y. Zhuang, L. Hua, L. Qi, J. Yang, P. Cao, Y. Cao, Y. Wu, J. Thompson, and H. Haas, "A survey of positioning systems using visible LED lights," *IEEE Comm. Surv. and Tutor.* **20**(3), 1963–1988 (2018).
11. M. Yasir, S. Ho, and B. N. Vellambi, "Indoor position tracking using multiple optical receivers," *J. Lightwave Technol.* **34**(4), 1166–1176 (2016).
12. X. Li, Y. Cao, and C. Chen, "Machine learning based high accuracy indoor visible light location algorithm," in *2018 IEEE International Conference on Smart Internet of Things (SmartIoT)*, Xi'an, 198–203 (2018).
13. M. T. Van, N. V. Tuan, T. T. Son, H. Le-Minh, and A. Burton, "Weighted k-nearest neighbour model for indoor VLC positioning," *IET Commun.* **11**(6), 864–871 (2017).
14. C. Hsu, S. Liu, F. Lu, C. Chow, C. Yeh, and G. Chang, "Accurate indoor visible light positioning system utilizing machine learning technique with height tolerance," in *2018 Optical Fiber Communications Conference and Exposition (OFC)*, San Diego, California, 1–3 (2018).
15. X. Guo, N. Ansari, L. Li, and H. Li, "Indoor localization by fusing a group of fingerprints based on random forests," *IEEE Internet of Things Journal* **5**(6), 4686–4698 (2018).
16. J. He, C. Hsu², Q. Zhou, M. Tang, S. Fu, D. Liu, L. Deng, and G. Chang, "Demonstration of high precision 3D indoor positioning system based on two-layer ANN machine learning technique," in *2019 Optical Fiber Communications Conference and Exposition (OFC)*, San Diego, California, 1–3 (2019).
17. P. Wawrzyński, "Reinforcement learning with experience replay for model-free humanoid walking optimization," *International Journal of Humanoid Robotics* **11**(3), 137 (2014).
18. E. Bejar and A. Moran, "Deep reinforcement learning based neuro-control for a two-dimensional magnetic positioning system," in *2018 4th International Conference on Control, Automation and Robotics (ICCAR)*, Auckland, 268–273 (2018).
19. D. Milioris, "Efficient indoor localization via reinforcement learning," in *2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, United Kingdom, 8350–8354 (2019).
20. Z. Zhang, H. Chen, X. Hong, and J. Chen, "Accuracy enhancement of indoor visible light positioning using point-wise reinforcement learning," in *2019 Optical Fiber Communications Conference and Exposition (OFC)*, San Diego, California, 1–3 (2019).
21. X. Guo, S. Shao, N. Ansari, and A. Khreishah, "Indoor localization using visible light via fusion of multiple classifiers," *IEEE Photonics J.* **9**(6), 1–16 (2017).
22. W. Zhang, M. I. S. Chowdhury, and M. Kavehrad, "Asynchronous indoor positioning system based on visible light communications," *Opt. Eng.* **53**(4), 045105 (2014).