IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 29, 2020, accepted June 11, 2020, date of publication June 22, 2020, date of current version July 1, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3003977

High Precision Weighted Optimum K-Nearest Neighbors Algorithm for Indoor Visible Light Positioning Applications

HUY QUANG TRAN AND CHEOLKEUN HA, (Member, IEEE)
Robotics and Mechatronics Laboratory, University of Ulsan, Ulsan 44610, South Korea
Corresponding author: Cheolkeun Ha (cheolkeun@gmail.com)

This work was supported by the Research Fund of University of Ulsan under Grant 2020-0631.

ABSTRACT Enhancing the accuracy of indoor visible light positioning systems with simple, real-time, and stable methods is one of the interesting challenges in recent research. In this paper, a relatively minor mean positioning error of 8 mm and a 42-52% improvement in computational time could be achieved within a real space of 1.2 m × 1.2 m × 1.2 m by transcending the serious limitations of the traditional k-nearest neighbors (KNN) algorithm. These disadvantages (slow execution time, high error formation) are a result of finding the nearest neighbors from all the fingerprints, averaging the Euclidean distances, and the excessive passivity of the K value. To overcome the above limitations of KNN, we proposed a maximum received signal strength recognition (MRR) technique and weighted optimum KNN (WOKNN) algorithm, which is a combination of optimum KNN (OKNN) and weighted KNN (WKNN). While MRR was used to reduce the computational time, WOKNN was used to solve the remaining problems. Specifically, OKNN was used to automatically determine the best number of nearest neighbors for each position in the area under consideration, and WKNN helped improve the errors that come from the Euclidean distance averaging process. Based on positive experimental results and a meaningful comparison with various versions of KNN, we demonstrated that the improved conventional KNN algorithm can achieve very high positioning accuracy and is totally suitable for several specific 2-D indoor positioning applications.

INDEX TERMS Indoor positioning, k-nearest neighbors, visible light positioning, weighted k-nearest neighbors.

I. INTRODUCTION

Unlike the outdoor environment, where the acceptable positioning accuracy can range from a few dozen centimeters to a few meters with the Global Positioning System [1], indoor positioning activities require a much higher quality positioning accuracy, even to the millimeter level. A distinguishing feature of this type of environment is that objects to be located often either move in narrow spaces such as corridors, stairs, aisles, and specialized rooms or stay in a position with many surrounding obstacles that remain out of sight. These characteristics make the enhancement of the positioning performance for indoor environment challenging.

Recently, a promising visible light-based positioning system (VLPS) has attracted a great deal of attention because of the extremely low positioning error that this system can provide. VLPS continues to emerge as a leading candidate in the field of indoor positioning because of its advantages of accuracy, cost effectiveness, and simplicity even though it faces some difficulties, including multipath reflection, light interference between Light Emitting Diode (LED) lights, and noises from sunlight and artificial light sources [2], [3].

To implement positioning operations, some common methods such as angle of arrival [4], time of arrival [5], [6], time difference of arrival [7], phase difference of arrival (PDOA) [8], received signal strength (RSS) [9], [10], and a hybrid of these techniques [11], [12] have been applied. Each method has its own advantages and disadvantages, of which RSS is considered as one of the cheapest and simplest solutions [13]. This solution measures the power strength of the LED light thanks to a photodetector (PD) placed on the object. Based on these data, the distance from the transmitter to the receiver can be estimated by converting the RSS data into the length between LEDs, or estimated

The associate editor coordinating the review of this manuscript and approving it for publication was Tyson Brooks.

locations can be inferred by combining RSS data and the fingerprinting method [2], [14]. The RSS is a popular, simple approach that provides acceptable positioning error for indoor positioning applications [13]; This approach is also applied for other types of signals such as WIFI [15], [16], Bluetooth [17], [18], Lorawan [13], and Zigbee [19], [20]. In the VLPS field, recent research based on the RSS platform have continued to improve [9], [11], [21]. In [9], the authors presented a Metropolis-Hastings algorithm-based Bayesian to improve the positioning capability in a 3-D indoor environment via computer simulations. In a similar environment, Zhang *et al.* [11] improved the robustness of VLPS by creating a hybrid of RSS and PDOA based on artificial neural networks. Using simulations, the authors demonstrated that their approach achieved robust positioning outcomes under various levels of RSS. In a more realistic approach, the authors in [21] established two real experimental spaces for their VLPS. By experimenting with multiple scenarios, median localization errors of 1.9 cm and 16.1 cm could be achieved thanks to the connection between RSS and the spring relaxation algorithm.

The combination of RSS data and machine learning has been a growing trend in the era of artificial intelligence. Using machine learning algorithms, many different approaches have been proposed with the overall goal of optimizing the location accuracy, complexity, and practical applicability of these techniques [22]–[24]. Among them, KNN is a preferred option [25]–[28]. The authors in [25] performed a field experiment for 3-D space based on RSS, fingerprint and KNN. The experimental results proved that the above combination and weighted average method improved the positioning accuracy up to 0-6 cm. In [26], Fakhrul *et al.* also designed a practical VLPS by connecting RSS and the weighted KNN (WKNN) method. Moreover, an important improvement from this paper was the reduction in the number of required fingerprint databases, which was realized using a fast calibration method. This approach showed that although the number of fingerprints is reduced, the positioning accuracy is relatively high with a mean error of 2.7 cm when applying to only 12 offline measurements. With a similar method, Wenge *et al.* [27] applied a square wave modulation model to detect the luminaire from each LED using a Fast Fourier Transform and then compared the difference between ID vectors in the online mode and the offline mode, resulting in a mean error of 1.39 cm. In addition to using KNN for the purpose of estimating location, the authors in [28] used it as a tool to support the later positioning process. As is well-known, KNN is a simple machine learning algorithm with two functions of classification and regression. Taking advantage of the classification function of KNN and two robust fusion algorithms, Guo *et al.* demonstrated that their technique can achieve a mean square positioning error of less than 5 cm. The common feature of these articles is based on the traditional KNN algorithm and an improved version of KNN, namely WKNN. WKNN is a considerable improvement of KNN and can achieve a better positioning accuracy. However, the full potential of KNN has not been realized due to its own weaknesses. First, the execution time of the system is directly proportional to the calculation time of the Euclidean distances between the considering point and all the fingerprints. More fingerprints mean a slower execution. Normally this method is very time-consuming. Secondly, averaging the coordinates of all the nearest neighbors, regardless of the difference in each Euclidean distance, causes an error in the final estimation result. Finally, the number of nearest neighbors (K) is usually chosen manually and a unique K value is fixed for all fingerprints. This is also a significant drawback because each fingerprint corresponds to different optimal K value in reality. There are some points that have the same K value, but not all points have the same optimal K. In this paper, we address the above limitations of KNN to achieve better positioning quality with a simple and stable method.

Our main contributions are as follows:

1) We first propose a Maximum Received Signal Strength Recognition (MRR) technique to reduce the computational time when using the KNN algorithm. RSS intensity depends on the distance between the transmitter and the receiver. There is always a maximum RSS value at any position (including the center of the floor) in a practical VLPS, and the maximum RSS value belongs to the LED that has the closest distance to the PD. In this method, the peaks of received signals that come from four LED lights are captured by the receiver at all reference points. Then, we can divide the entire testbed floor into 4 specific areas according to the maximum RSS data. During a real positioning estimation process, the point under consideration can only be in one of the four classified areas. Therefore, the process of finding nearest points only takes place in a small area instead of the entire floor. This greatly reduces the execution time for the KNN algorithm.
2) We then designed the weighted optimum KNN (WOKNN) algorithm to improve the positioning inaccuracy caused by averaging Euclidean distances and fixing the number of nearest neighbors. WOKNN is a combination of optimum KNN (OKNN) and WKNN, where OKNN automatically optimizes the K value at each position thanks to a multi-layer perceptron classifier (MLP) model. Then, each Euclidean distance has an applied weighted factor that is inversely proportional to the distance obtained by employing the WKNN method.

The paper is organized as follows. In Section II, the overview of the proposed visible light system and our solution are presented. The experimental results and discussion are given in Section III, and some conclusions are drawn in Section IV.

II. PROPOSED POSITIONING ALGORITHM

To provide a comprehensive overview of the VLPS as well as our proposed solutions, we first present RSS measurements

in the visible light positioning systems. Then, the WOKNN algorithm is illustrated in more detail.

A. RECEIVED SIGNAL STRENGTH MEASUREMENT IN A VISIBLE LIGHT POSITIONING SYSTEM

In this experimental system, we only consider the line-of-sight channel without the non-line-of-sight channel due to the specific characteristics of the experimental model. In this model, there is a single wall near the experimental area, but the distance from the wall to this area is quite far (1.2 m), and there are no obstructions in the experimental area itself. Therefore, the negative influence of multipath reflection can be ignored. For the directed optical channel, the received optical power at the receiver can be expressed as [29]

$$P_R^i = P_T^i H_{DC}^i \tag{1}$$

where i is the number of LED lights, P_T^i is the transmitted optical power, H_{DC}^i is the channel DC gain, which is defined as

$$H_{DC} = \frac{\mathrm{R}_T(\varphi)\,\mathrm{A}_E(\delta)}{d^2}, \quad 0 \le \delta \le \delta_C \tag{2}$$

Here, φ is the irradiance angle, δ and δ_C are the incidence angle and the receiver field of view, respectively, and d presents the distance between receiver and transmitter. The Lambertian radiant intensity $\mathrm{R}_T(\varphi)$ and effective signal-collection area $\mathrm{A}_E(\delta)$ are defined as [30]

$$\mathrm{R}_T(\varphi) = \frac{n+1}{2\pi}\cos^n(\varphi) \tag{3}$$

$$\mathrm{A}_E(\delta) = A\cos^n(\delta)\,T_F(\delta)\,G(\delta), \quad 0 \le \delta \le \delta_C \tag{4}$$

where A is the detector physical area and $T_F(\delta)$ is the gain of the optical filter. The gain of the optical concentrator $G(\delta)$ and the Lambertian order n are presented as [29], [30]

$$G(\delta) = \left(\frac{\eta}{\sin\delta_C}\right)^2 \tag{5}$$

$$n = \frac{-\ln(2)}{\ln(\cos(\varphi_{1/2}))} \tag{6}$$

where η is the refractive index and $\varphi_{1/2}$ is the semi-angle of transmitter

To gather the RSS data set for the training process in the next part, we set up a 10 cm × 10 cm grid covering the entire floor of the test space as depicted in Fig. 1. In this paper, 169 fingerprint points were considered. After collecting all the data, the RSS dataset is expressed as

$$RSS = \begin{bmatrix} R_{11} & R_{12} & \dots & R_{1n} \\ R_{21} & R_{22} & \dots & R_{2n} \\ \dots & \dots & \dots & \dots \\ R_{m1} & R_{m2} & \dots & R_{mn} \end{bmatrix} \tag{7}$$

where $m = 13$ and $n = 13$ are the number of rows and columns of the grid, respectively.

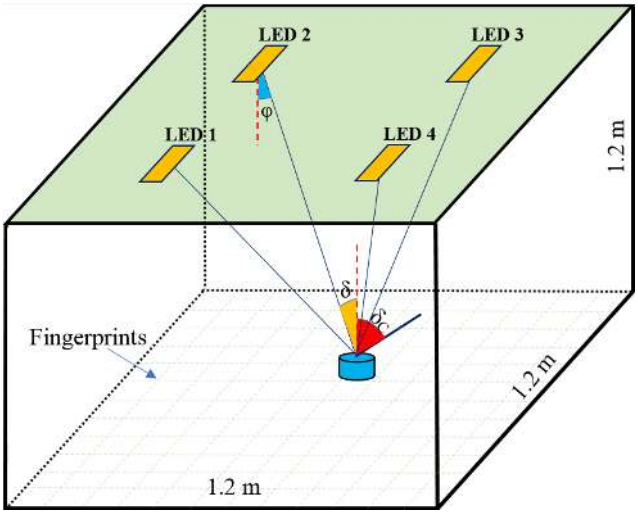

FIGURE 1. Proposed scenario of visible light positioning system.

B. INDOOR POSITIONING WITH WOKNN ALGORITHM

As discussed in Section I, KNN is a simple, stable and popular technique used for visible light-based indoor positioning applications. In this paper, we have unleashed the unexplored potential of the traditional KNN algorithm by fully exploiting the considerable advantages of the algorithm and overcoming the inherent limitations of KNN (i.e., problems associated with time-consuming execution, error formation due to averaging the Euclidean distances, and the passivity of the K value). Specifically, we proposed the MRR technique and WOKNN algorithm, where the MRR technique is used to minimize the computational time and WOKNN maximizes the positioning accuracy.

1) METHOD OVERVIEW

As illustrated in Fig. 2, the proposed VLPS has three main categories, which are the optical transmitter, optical receiver, and positioning process, respectively. In the optical transmitter section, we installed four LEDs on the top of the testbed frame. Each LED is arranged in a fixed position. However, a slight deviation in the LED position does not affect the final positioning result according to the RSS-based fingerprint method. It is important that LEDs are kept constant throughout the experiment to ensure that stable RSS values are obtained. In the transmitter section, Arduino Uno was used to control the operation of LEDs using an LED driver. In our approach, the time-division multiplexing technique was employed to help the receiver detect which LED signal is transmitting [31], as shown in Fig. 3. In the first slot, we turned off all LEDs within 2.2 ms to create the background light intensity. After this period, all four LEDs were turned on and off, respectively, in every cycle of 11 ms.

In addition to separating the transmission signals into four separate slots in each cycle for the purpose of easier identification by the PD, the fixed transmission frequency also helped to fix the light interference among LED groups.

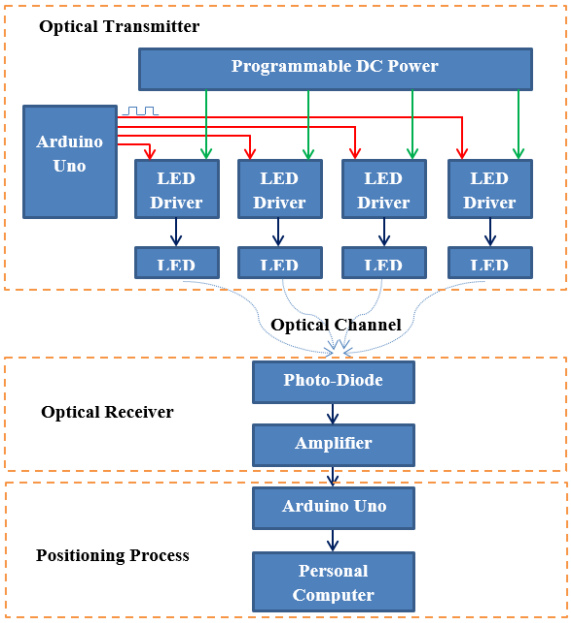

FIGURE 2. Flowchart of proposed visible light positioning system.

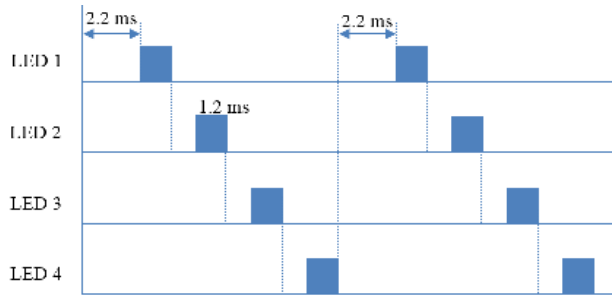

FIGURE 3. Time-division multiplexing scheme.

Furthermore, the selection of maximum RSS value based on MRR technique is a reliable solution to eliminate the light interference problem. These approaches maintained the reliable RSS values collected during the offline and online modes. Using PD, the receiver easily captures the RSS data of each LED and recognizes the order in which LEDs come on due to the transmitting order that was set in the transmitter. The amplifier unit is also a main part of the optical receiver. While PD is responsible for receiving RSS data, the amplifier unit acts as a bridge to amplify and convert the PD current into voltage signals before sending these signals to the available analog digital converter inside the second Arduino Uno. To determine the estimated location of the object, all the digital data are directly sent to a personal computer where our proposed WKONN is executed. Detailed information related to the optical transmitter, the optical receiver, and experimental space are described in Section III.

2) DATA COLLECTION AND PROCESSING IN OFFLINE AND ONLINE MODES

In the RSS-based fingerprint method, the positioning process is divided into two distinct modes: offline mode and online mode.

In the offline mode (Fig. 4), we first collect the RSS data at all fingerprint points. The final data set is defined as

$$RSS^{off}_{n(i,j)} = \begin{bmatrix} R^{off}_{11} & R^{off}_{12} & \dots & R^{off}_{1j} \\ R^{off}_{21} & R^{off}_{22} & \dots & R^{off}_{2j} \\ \dots & \dots & \dots & \dots \\ R^{off}_{i1} & R^{off}_{i2} & \dots & R^{off}_{ij} \end{bmatrix} \tag{8}$$

Then, these data are stored in the memory, and we use them to perform two important tasks: specific area recognition with MRR and K value optimization with MLP.

Similar to the offline mode, collecting RSS data for the real position of the PD is mandatory. Equation (9) shows the RSS data from 4 LEDs at a certain position

$$RSS^{on} = \begin{bmatrix} R^{on}_1 & R^{on}_2 & R^{on}_3 & R^{on}_4 \end{bmatrix} \tag{9}$$

The offline mode is used to collect, train, and analyze all the fingerprint data. The execution time of the online mode is faster, though the online mode undergoes more procedures as depicted in Fig. 5. This happens because this process utilizes the results from the previous process, and it only estimates the coordinate at a certain position under consideration. After conducting the same process as the offline mode, the final estimated position is determined using the WOKNN algorithm. Details of the MRR technique, K value optimization, and WOKNN algorithm are provided below.

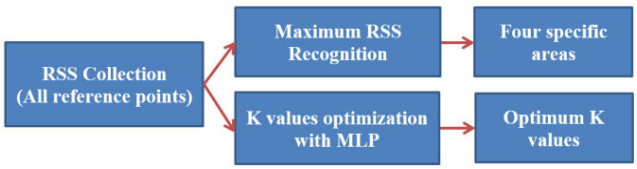

FIGURE 4. Offline mode flowchart.

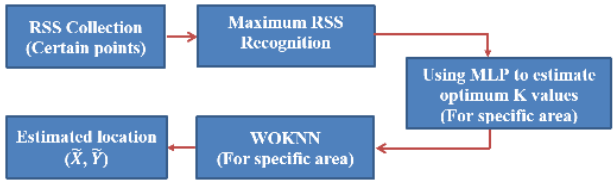

FIGURE 5. Online mode flowchart.

3) MAX RSS RECOGNITION TECHNIQUE

The purpose of MRR is to classify the entire experimental floor into four specific areas based on the maximum RSS data at each position. We set up a practical testbed with four LED groups (LEDs) separately installed on an aluminum frame at a height of 1.2 m. The distance between two adjacent LEDs was 0.6 m. The detailed locations of each LED are clearly shown in Table 1. At each reference point on the floor, the PD receives the RSS signals from all four LEDs. However, the actual RSSs at a certain position are not quite the same for two major reasons. First, although we used the same kinds of LED, the actual optical intensity distribution of every LED light is slightly different (Fig. 6) due to the power supply unit,

TABLE 1. Main parameters of experimental model.

Object	Parameter	Value
Transmitter	LED's brand name	Luxeon XR-3020
	Typical luminous efficacy	156 lm/W
	Typical drive current	200 mA
	Nominal CCT	5000 K
	Minimum CRI	80 CRI
	LED color	White
	No. of LED	24 x 4
	LED position (x, y, z) (m)	LED 1 (0.3, 0.3, 1.2)
		LED 2 (0.3, 0.9, 1.2)
		LED 3 (0.9, 0.9, 1.2)
		LED 4 (0.9, 0.3, 1.2)
Receiver	Brand name	Hamamatsu S1227-33BR
	Spectral range	340-1000 nm
	Spectral peak	720 nm
	Sensitivity	430 mA/W
	Dark current	5pA
	Capacitance	200 pF
	Resistance	20 GOhm
	Response time	500 ns
	Active area	5.76 mm^2
	Viewing angle	30 deg
	Amplifier	AD8041
Model	Model dimension	1.2 m x 1.2 m x 1.2 m

driver boards and minor technical faults of the LED itself. Therefore, there is always a maximum RSS value at a certain position, even if the considered point is in the center of the floor. Secondly, the distances from PD to all four LEDs are different according to the PD's current position (except for the center point of the floor). In addition to the two main reasons above, the influence of noise due to ambient light is also one of the possible causes of differing RSS values. In this paper, the MRR technique is first used to identify the maximum RSS data at all reference points; then, it can create four specific areas based on that RSS data as shown in Fig. 7.

From the classification result, the WOKNN algorithm only finds the nearest points for each real position within one of four classified areas instead of surveying the entire floor. This significantly reduces the execution time for WOKNN, which is also time-consuming due to K optimization with MLP.

To maintain the variation of light intensity, the power source for all LED drivers is kept as a constant value, and the transmission frequency is also fixed. Therefore, the overall variation of the whole system is unchanged. This leads to a stability in the intensity distribution of each LED light. Based on this procedure, the RSS data collected at any points on the floor of the test bed are still stable.

4) OPTIMUM NEIGHBORS CLASSIFICATION WITH MLP

In this part, we clarify the role of K value optimization with MLP - the second task in offline mode. As we discussed in Section I, many recent research articles focused on KNN and WKNN for VLPS applications; however, the constant K value is a serious limitation. In most cases, the K value is intentionally fixed after finding the best K value by manually running this value in a wide range. Obviously, there is not a unique K value for all cases. The fixed number of nearest neighbors, therefore, leads to errors in positioning results. To address this problem, we classified optimum K values by applying the MLP model.

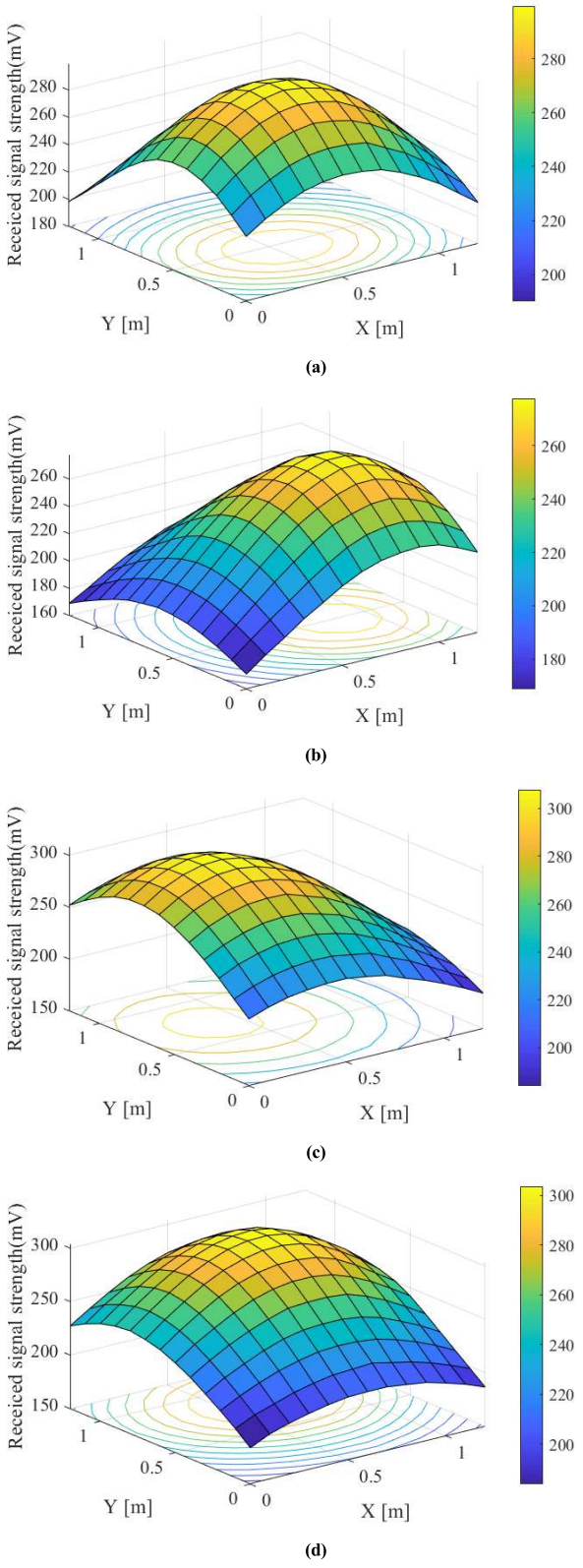

FIGURE 6. Practical optical power distribution of (a) LED 1, (b) LED 2, (c) LED 3, (d) LED 4.

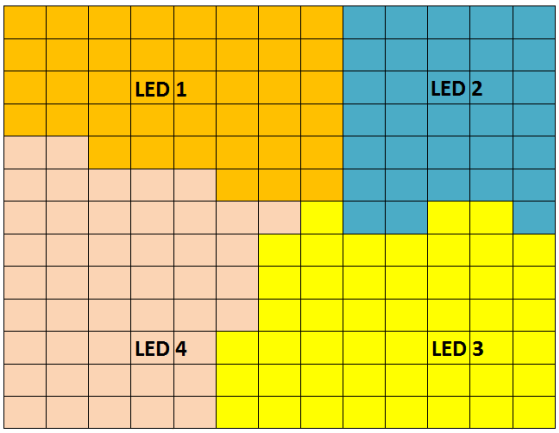

FIGURE 7. Optical power distribution of all LEDs on the entire floor.

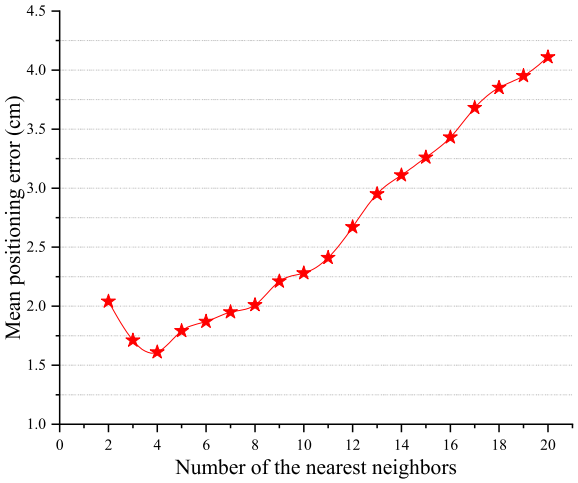

FIGURE 8. Mean positioning errors vs. number of the nearest neighbors.

To create a training dataset for MLP, we first determined the optimum K values at all reference points by running K from 2 to 7 with a conventional KNN algorithm in the offline mode. This does not affect the actual execution time during the online process. We limited the value of K for two main reasons. First, higher K values resulted in more neighbors and a longer system execution time. Secondly, as shown in Fig. 8 (where K = 4), a steady increase in K value resulted in a gradual reduction in the positioning accuracy, especially when K equals 8 or more. After obtaining the optimum K values at all survey points, we recognized that 2, 3, 4 were the best values in Fig. 9, where K = 2 accounts for 47.93%, K = 3 accounts for 27.22%, and K = 4 accounts for 24.85%. All these data and corresponding RSS values were used as training datasets when applying the MLP algorithm in the next step.

In the proposed model, MLP serves as both a training process in offline mode and a prediction process in online mode [32]. The specific purpose of MLP is to estimate the optimum K value for any given point based on the results of the training process conducted in offline mode. In Fig. 10, the proposed MLP model consists of 2 hidden layers, and 5 neurons for each hidden layer. The ReLU activation

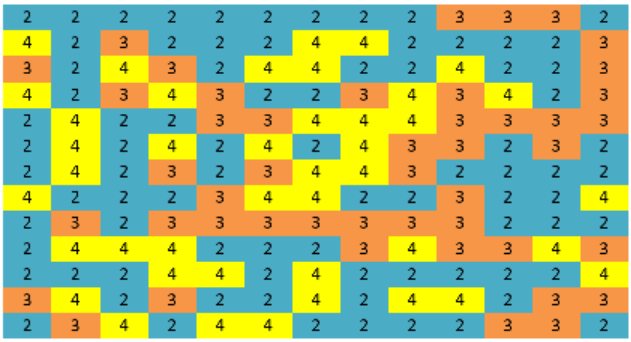

FIGURE 9. Optimum k values at all reference points.

function is used for hidden layers. The input layer receives RSS signals from 4 LEDs. The output layer is divided into 3 channels when K equals 2, 3, and 4. The result showed that MLP achieved an accuracy of approximately 94% during the process of classifying the optimum K values.

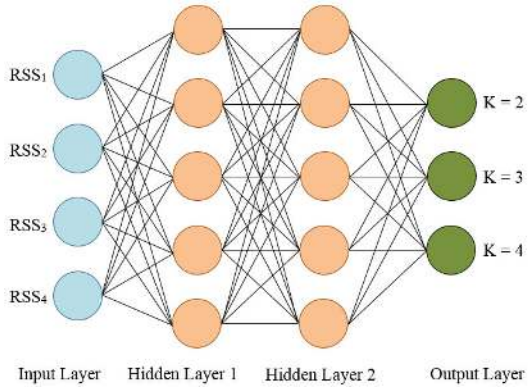

FIGURE 10. Optimum k values classification with MLP.

5) WEIGHTED OPTIMUM KNN

It is very important to note that the estimation of positions based on the average calculation of Euclidean distances is one limitation of a traditional KNN algorithm besides the slow running time. In this paper, the Euclidean distance is described as

$$d_{i,j} = \sqrt{\sum_{n=1}^{4} \left(RSS_n^{on} - RSS_{n(i,j)}^{off} \right)} \tag{10}$$

where n is the number of LED lights, RSS_n^{on} and $RSS_{n(i,j)}^{off}$ are the RSS data in the online mode and offline mode, respectively

After computing all the Euclidean distances thanks to the RSS received from the LED system, we determine the number of nearest neighbors based on the optimum K values estimated from the MLP in the previous step. The corresponding coordinates are easily inferred from these nearest points, and then we can utilize this information to determine the

estimated position of the object as follows:

$$x_{KNN} = \frac{\sum_{n=1}^{K} x_n}{K}; \quad y_{KNN} = \frac{\sum_{n=1}^{K} y_n}{K} \tag{11}$$

where x_n, y_n are the coordinates corresponding to nearest neighbors, K is the number of nearest points.

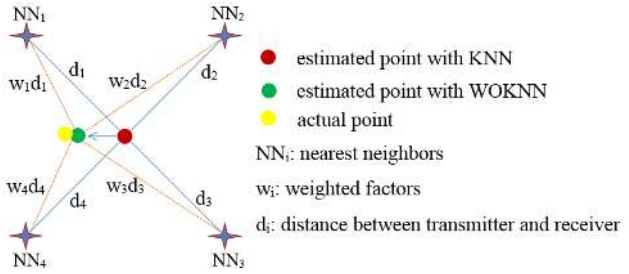

FIGURE 11. Error reduction with WOKNN.

According to (11), a very high accuracy can be set if the actual point is either at the middle of the line (in the case of two closest neighbors) or in the center of the figure formed by the nearest points (in the case of more than two closest neighbors). In fact, the point under consideration could be anywhere on the floor, and the average Euclidean distances can create unexpected errors, as shown in Fig. 11. Therefore, we apply weighted factors for every Euclidean distance, in which case these weights are in inverse proportion to the distances. By employing this method, we can minimize the positioning inaccuracy caused by averaging the Euclidean distances. The related formula of weighted factors corresponding to each Euclidean distance d_{i_min} is presented as follows:

$$w_i = \frac{\frac{1}{d_{i\_\min}}}{\sum_{n=1}^{K} \frac{1}{d_{i\_\min}}} \tag{12}$$

When we applied these weights to (11), the current position of the PD was estimated as

$$x_{WOKNN} = \frac{\sum_{n=1}^{K} x_n w_n}{\sum_{n=1}^{K} w_n}; \quad y_{WOKNN} = \frac{\sum_{n=1}^{K} y_n w_n}{\sum_{n=1}^{K} w_n} \tag{13}$$

6) FRIEDMAN TEST AND POST-HOC ANALYSIS

To statistically evaluate the performance of our proposed algorithm and other existing methods, we apply Friedman test, which is used to rank the algorithms' performance. The test statistic suggested by Friedman (14) and F distribution by Iman and Davenport (15) are as follows [33], [34]:

$$T_F = \frac{12n}{m(m+1)} \left[\sum_{i=1} R_i^2 - \frac{m\,(m+1)^2}{4} \right] \tag{14}$$

$$F_F = \frac{T_F\,(n-1)}{n(m-1) - T_F} \tag{15}$$

where n is the number of datasets, m is the number of compared algorithms, R_i^2 be the average ranks of the i-th of m algorithms.

According to the null hypothesis, all the compared algorithms have the same rank. Therefore, the null hypothesis is rejected when the ranks of all algorithms are not equivalent. In this case, we continue to apply post-hoc analysis to find which algorithms differ from the others. The algorithms a and b are considered significantly different when [34]:

$$|R_a - R_b| \geq CD; \quad CD = q\sqrt{\frac{m(m+1)}{6n}} \tag{16}$$

where q is the critical value, which can be found in the studentized range statistic table.

III. EXPERIMENTAL RESULT AND DISCUSSION

In this section, we continue to provide an overview of the main structure and specifications of the experimental system, and then some experimental results are discussed in more detail.

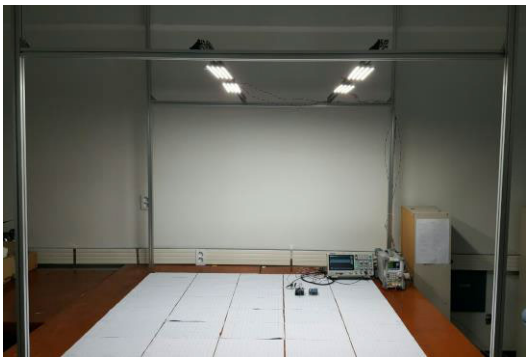

FIGURE 12. Experimental positioning system.

A. EXPERIMENTAL SETUP

To demonstrate the efficiency of the proposed MRR technique and WOKNN algorithm in real environment, a practical LED light-based positioning system was established. As illustrated in Fig. 12, this system consists of a 1.2 m × 1.2 m × 1.2 m aluminum frame, four LED lights suspended from the top of the frame, and one optical receiver, which can freely move around the floor (Fig. 13). The main specifications of the transmitter and receiver are detailed in Table 1.

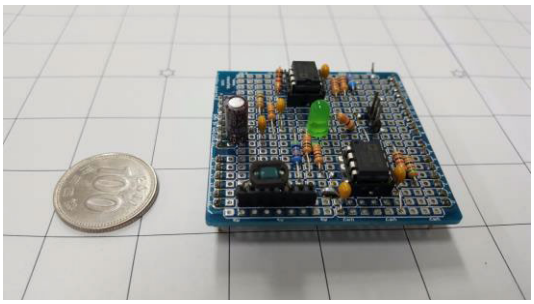

FIGURE 13. Receiver circuit board with photodiode and amplifier.

B. RESULT AND DISCUSSION

In our experiment, RSS data at 51 different points that intentionally form the letters VLP (an abbreviation for Visible Light Positioning) were collected in the real mode to evaluate the positioning quality and the ability to shorten execution time by using our proposed algorithms.

1) POSITIONING ERRORS EVALUTION

Before drawing a comparison of positioning accuracy between our proposed algorithm and other existing algorithms, we evaluated the performance of the MLP model, which was used to estimate the optimum K value for the suggested algorithm, using 10-fold cross validation. In this procedure, we randomly divided the whole dataset into 10 subsets, and each subset is given the opportunity to be a testing set one time on each iteration. This process helps reduce the risk of overfitting problem and estimate the skill of the MLP model. For every iteration, we computed the evaluation score based on the percentage of prediction accuracy as illustrated in Fig. 14. Obviously, the prediction accuracy is very impressive on each iteration and the mean accuracy of approximately 94% was, therefore, obtained in total.

FIGURE 14. 10-fold cross validation.

In this experiment, RSS data at 51 different points that intentionally form the letters VLP were collected in the real mode. After gathering enough testing datasets, all the above data were first distributed to individual areas by the MRR technique. The detailed distribution is depicted in Fig. 15, in which the blue star signs are actual positions, and the red plus signs are estimated positions. Next, the estimation process of MLP for estimating optimal K values takes place. Then, we apply the WOKNN algorithm to all 51 points under consideration to determine the estimated track. From the results in Fig. 15, the best performance is shown in the last case with WOKNN (Fig. 15d), where the estimated results are almost identical to the actual data. In contrast, the worse positioning accuracy distributions are shown in Figs. 15a, b, and c when the distances between the estimated points and the actual points are relatively large (especially in the first case with KNN). This means that the positioning quality of a traditional KNN-based algorithm is the worst compared to WKNN, OKNN, and WOKNN.

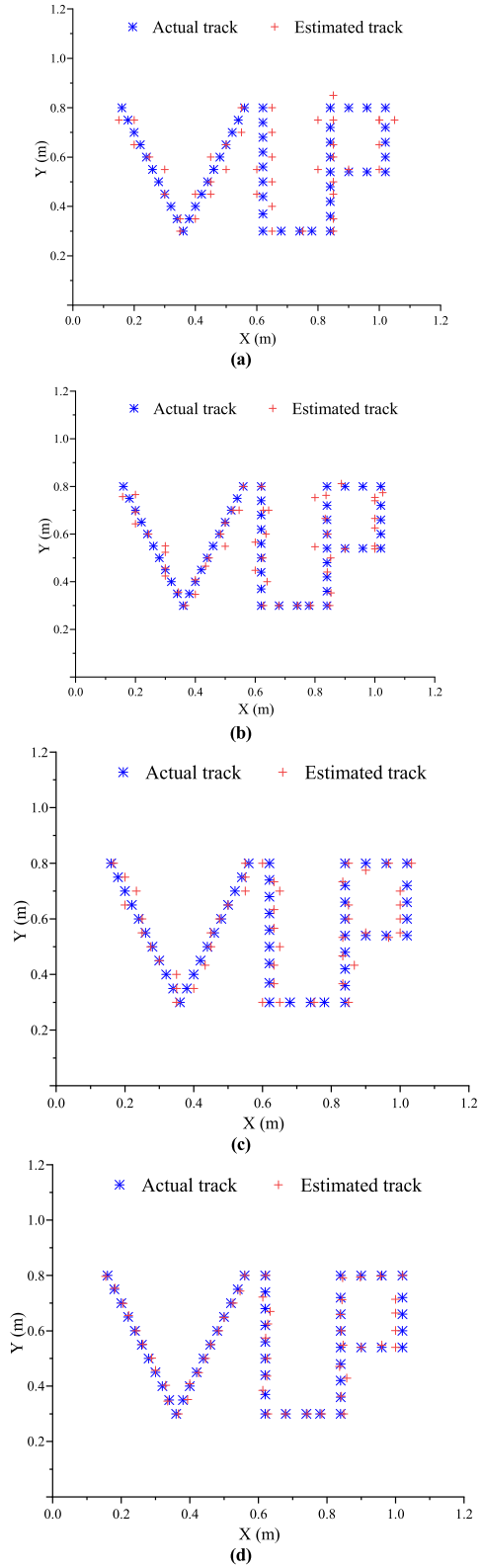

FIGURE 15. Position estimation by (a) KNN, (b) WKNN, (c) OKNN, and (d) WOKNN.

The results in Fig. 15 show that a combination of optimum K value and weighted Euclidean distances provides the highest positioning accuracy. This is understandable because

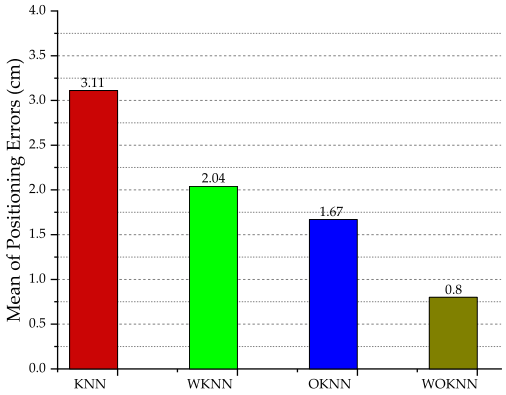

FIGURE 16. Mean positioning error with various version of KNN.

the optimum K value is the best value at the considering location, and the weighted factors help adjust the estimated position close to the actual position. To clarify the above conclusion, the mean positioning errors with various versions of KNN are shown in Fig. 16. In this figure, the degree of improvement of positioning errors was gradually enhanced thanks to our proposed techniques. Obviously, in comparison to other approaches, the traditional KNN algorithm shows the worst result with an error of 3.11 cm. However, when we apply WKNN and OKNN, we can easily observe a significant change in the blue column and green column, which show errors of 2.04 cm and 1.67 cm, respectively. Finally, WOKNN reduced the error to less than 1 cm. In our experiment, an error of 0.8 cm was been achieved by the WOKNN algorithm. In addition to mean errors, the histograms and standard deviations of positioning errors are also illustrated in Fig. 17, where we presented a detailed comparison between WKNN with a consistent K value and WOKNN with optimum K values. We conducted this comparison because WOKNN algorithm is an obvious connection between WKNN and optimum K values. In the first three subfigures, we applied a fixed value of K for each one, while optimum K values was set into subfigure d, where the considered point was always received the best value of K. Depending on the location of the object being considered on the floor, every point in Fig. 17d has a different K value, and it is also called the optimum K value. In four mentioned cases, the second case (Fig. 17b) and third case (Fig. 17c) show better performance than the first case (Fig. 17a) in both the distribution of error, mean positioning error (47.5% and 21.1%), and standard deviation (46.5% and 40.3%). The best result was obtained using optimum K values, Fig. 17d, with the best error distribution, the lowest mean positioning inaccuracy (0.8 cm) and the lowest standard deviation (0.59 cm). This means that the combination of WKNN and optimum K values significantly improve positioning accuracy.

For further evaluation of the performance of our proposed algorithm in term of the positioning accuracy, the Friedman test and post-hoc analysis was investigated in detail (Table 2). In the first step, we computed the average ranks of each considering algorithm (i.e., R_{KNN}, R_{OKNN}, R_{WKNN}, and R_{WOKNN})

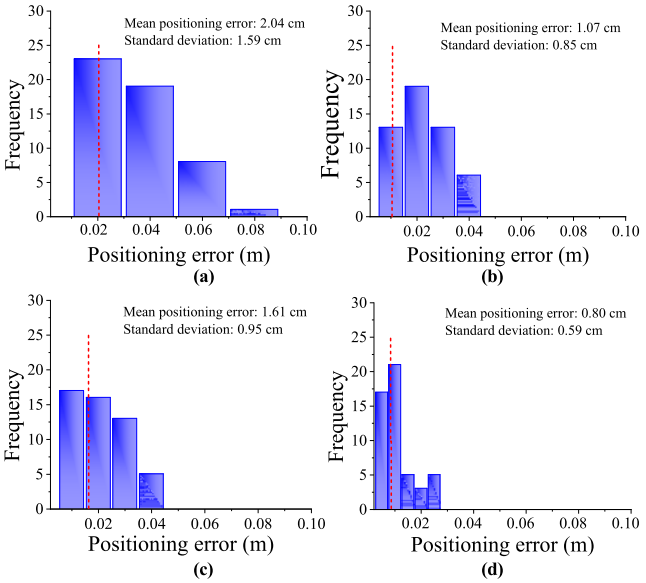

FIGURE 17. Histogram of positioning error with WKNN by (a) K = 2, (b) K = 3, (c) K = 4, and (d) Optimum K.

TABLE 2. Parameters of Friedman test and post-hoc analysis.

R_{KNN}	R_{OKNN}	R_{WKNN}	R_{WOKNN}	R_{mean}	T_F	D_F	CD
3.559	2.461	2.618	1.363	2.5	1.519	0.502	0.612

TABLE 3. A comparison of post-hoc rank with CD = 0.612.

KNN vs OKNN	KNN vs WKNN	KNN vs WOKNN	OKNN vs WOKNN	WKNN vs WOKNN
1.098	0.941	2.196	1.098	1.225

and the mean rank R_{mean} of algorithms. Then, we computed the values of T_F (14), F_F (15), and the critical value of F distribution for significant level of 0.05. With 4 algorithms and 51 data points, the critical value of F distribution is 2.79, which is greater than the mean rank R_{mean}. Therefore, the null hypothesis was rejected. Finally, we applied post-hoc analysis method to compare the performance of our proposed algorithm with the existing algorithms thanks to the difference in ranks between them. The results in Table 3 showed that the traditional KNN is the worst algorithm compared with the others because

$$|R_{KNN} - R_{OKNN}| = 1.089 > 0.612$$
$$|R_{KNN} - R_{WKNN}| = 0.941 > 0.612$$
$$|R_{KNN} - R_{WOKNN}| = 2.196 > 0.612$$

Using the same procedure, we also recognized that our proposed algorithms WOKNN achieved the highest rank. This means that this solution performed significantly better than the remaining

$$|R_{KNN} - R_{WOKNN}| = 2.196 > 0.612$$
$$|R_{OKNN} - R_{WOKNN}| = 1.098 > 0.612$$
$$|R_{WKNN} - R_{WOKNN}| = 1.225 > 0.612$$

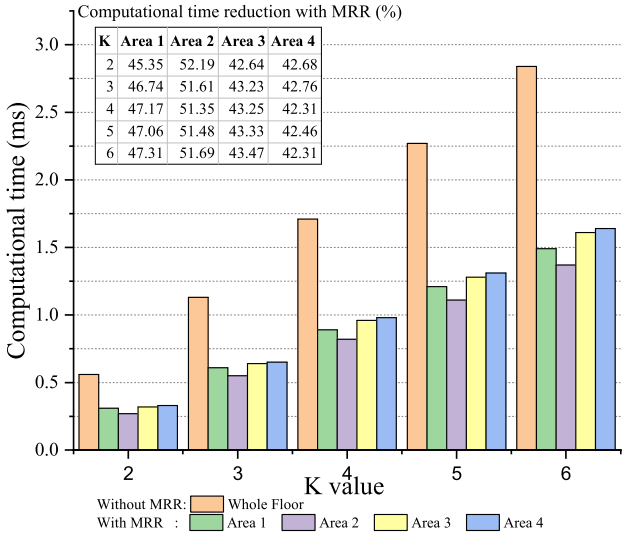

FIGURE 18. Comparison of computational time with and without MRR.

2) COMPUTATIONAL TIME

In addition to a noticeable enhancement in the positioning accuracy thanks to the WOKNN algorithm, the improvement in the computational time to meet real-time requirements is also addressed. We significantly reduced the execution time of the proposed algorithm by developing the MRR technique as discussed in Section 2. The results in Fig. 18 show that the total running time of each area decreased from approximately 42% to 52% reduction (compared with the entire floor before applying MRR) for all values of K (K runs from 2 to 6). As far as we known, the MRR method helped to classify the floor into four distinct areas, and each different area has a different number of reference points. Thus, the computational time for each area was not the same. However, the improvement in each area is consistent and significant. Obviously, the MRR technique paved the way for the WOKNN algorithm to become more feasible for indoor positioning applications that require high accuracy and real-time operation.

IV. CONCLUSION

In this paper, we developed a high precision indoor visible light positioning system by using the WOKNN algorithm, which resulted in a very low mean positioning error (0.8 cm). To improve the complexity cost, which is considered to be one of the most serious limitations of the KNN-based algorithm, the MRR technique was applied in the second step after collecting the RSS data and before employing WOKNN. This method helped to reduce the running time by 42-52% compared to the total execution time before applying MRR. These positive results confirmed the advantages of KNN, and also opened up the possibility of enhancing positioning accuracy by improving the inherent limitations of traditional positioning algorithms. The motivation for this approach was to obtain a model which could apply for mini mobile robots with a very high positioning accuracy in a specific space. However, our proposed solution could be a potential solution for larger experiment environment because of its simplicity and low positioning error. Finally, future research could also employ the proposed solution for more practical space where the negative impact of multipath reflection and noises due to furniture, surfaces of rooms, and humans are fully considered. In addition, localization in the 3-D environment could also be a key research direction in the near future.

REFERENCES

[1] E. D. Kaplan and C. J. Hegarty, *Understanding GPS: Principles and Applications*. Norwood, MA, USA: Artech House, 2006.
[2] 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 Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 1963–1988, 3rd Quart., 2018, doi: 10.1109/COMST.2018.2806558.
[3] T.-H. Do and M. Yoo, "An in-depth survey of visible light communication based positioning systems," *Sensors*, vol. 16, no. 5, p. 678, May 2016, doi: 10.3390/s16050678.
[4] H. Steendam, "A 3-D positioning algorithm for AOA-based VLP with an aperture-based receiver," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 1, pp. 23–33, Jan. 2018, doi: 10.1109/JSAC.2017.2774478.
[5] T. Akiyama, M. Sugimoto, and H. Hashizume, "Time-of-arrival-based smartphone localization using visible light communication," in *Proc. Int. Conf. Indoor Positioning Indoor Navigat. (IPIN)*, Sep. 2017, pp. 1–7, doi: 10.1109/IPIN.2017.8115904.
[6] C. Amini, A. Taherpour, T. Khattab, and S. Gazor, "Theoretical accuracy analysis of indoor visible light communication positioning system based on time-of-arrival," in *Proc. IEEE Can. Conf. Electr. Comput. Eng. (CCECE)*, Vancouver, BC, Canada, May 2016, pp. 1–5, doi: 10.1109/CCECE.2016.7726718.
[7] A. Naeem, N. U. Hassan, M. A. Pasha, C. Yuen, and A. Sikora, "Performance analysis of TDOA-based indoor positioning systems using visible LED lights," in *Proc. IEEE 4th Int. Symp. Wireless Syst. Int. Conf. Intell. Data Acquisition Adv. Comput. Syst. (IDAACS-SWS)*, Lviv, Ukraine, Sep. 2018, pp. 103–107, doi: 10.1109/IDAACS-SWS.2018.8525567.
[8] A. Naz, H. M. Asif, T. Umer, and B.-S. Kim, "PDOA based indoor positioning using visible light communication," *IEEE Access*, vol. 6, pp. 7557–7564, 2018, doi: 10.1109/ACCESS.2018.2796623.
[9] S. Buyukcorak and G. K. Kurt, "A Bayesian perspective on RSS based localization for visible light communication with heterogeneous networks extension," *IEEE Access*, vol. 5, pp. 17487–17500, 2017, doi: 10.1109/ACCESS.2017.2746141.
[10] X. Yu, J. Wang, and H. Lu, "Single LED-based indoor positioning system using multiple photodetectors," *IEEE Photon. J.*, vol. 10, no. 6, pp. 1–8, Dec. 2018, doi: 10.1109/JPHOT.2018.2848947.
[11] S. Zhang, P. Du, C. Chen, W.-D. Zhong, and A. Alphones, "Robust 3D indoor VLP system based on ANN using hybrid RSS/PDOA," *IEEE Access*, vol. 7, pp. 47769–47780, 2019, doi: 10.1109/ACCESS.2019.2909761.
[12] J.-W. Lee, S.-J. Kim, and S.-K. Han, "3D visible light indoor positioning by Bokeh based optical intensity measurement in smartphone camera," *IEEE Access*, vol. 7, pp. 91399–91406, 2019, doi: 10.1109/ACCESS.2019.2927356.
[13] S. Sadowski and P. Spachos, "RSSI-based indoor localization with the Internet of Things," *IEEE Access*, vol. 6, pp. 30149–30161, 2018, doi: 10.1109/ACCESS.2018.2843325.
[14] H. Liu, H. Darabi, P. Banerjee, and J. Liu, "Survey of wireless indoor positioning techniques and systems," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 37, no. 6, pp. 1067–1080, Nov. 2007, doi: 10.1109/TSMCC.2007.905750.
[15] K. Wang, X. Yu, Q. Xiong, Q. Zhu, W. Lu, Y. Huang, and L. Zhao, "Learning to improve WLAN indoor positioning accuracy based on DBSCAN-KRF algorithm from RSS fingerprint data," *IEEE Access*, vol. 7, pp. 72308–72315, 2019, doi: 10.1109/ACCESS.2019.2919329.
[16] A. Khalajmehrabadi, N. Gatsis, and D. Akopian, "Modern WLAN fingerprinting indoor positioning methods and deployment challenges," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 3, pp. 1974–2002, 3rd Quart., 2017, doi: 10.1109/COMST.2017.2671454.
[17] R. Faragher and R. Harle, "Location fingerprinting with Bluetooth low energy beacons," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 11, pp. 2418–2428, Nov. 2015, doi: 10.1109/JSAC.2015.2430281.

[18] J.-S. Jeon, Y. Kong, Y. Nam, and K. Yim, "An indoor positioning system using Bluetooth RSSI with an accelerometer and a barometer on a smartphone," in *Proc. 10th Int. Conf. Broadband Wireless Comput., Commun. Appl. (BWCCA)*, Kraków, Poland, Nov. 2015, pp. 528–531, doi: 10.1109/BWCCA.2015.142.

[19] Y. Alvarez and F. Las Heras, "ZigBee-based sensor network for indoor location and tracking applications," *IEEE Latin Amer. Trans.*, vol. 14, no. 7, pp. 3208–3214, Jul. 2016, doi: 10.1109/TLA.2016.7587622.

[20] V. Bianchi, P. Ciampolini, and I. De Munari, "RSSI-based indoor localization and identification for ZigBee wireless sensor networks in smart homes," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 2, pp. 566–575, Feb. 2019, doi: 10.1109/TIM.2018.2851675.

[21] F. Alam, N. Faulkner, M. Legg, and S. Demidenko, "Indoor visible light positioning using spring-relaxation technique in real-world setting," *IEEE Access*, vol. 7, pp. 91347–91359, 2019, doi: 10.1109/ACCESS.2019.2927922.

[22] X. Wang and J. Shen, "Machine learning and its applications in visible light communication based indoor positioning," in *Proc. Int. Conf. High Perform. Big Data Intell. Syst. (HPBD IS)*, Shenzhen, China, May 2019, pp. 274–277, doi: 10.1109/HPBDIS.2019.8735490.

[23] H. Tran and C. Ha, "Improved visible light-based indoor positioning system using machine learning classification and regression," *Appl. Sci.*, vol. 9, no. 6, p. 1048, Mar. 2019, doi: 10.3390/app9061048.

[24] P. Du, S. Zhang, C. Chen, H. Yang, W.-D. Zhong, R. Zhang, A. Alphones, and Y. Yang, "Experimental demonstration of 3D visible light positioning using received signal strength with low-complexity trilateration assisted by deep learning technique," *IEEE Access*, vol. 7, pp. 93986–93997, 2019.

[25] M. Xu, W. Xia, Z. Jia, Y. Zhu, and L. Shen, "A VLC-based 3-D indoor positioning system using fingerprinting and K-Nearest neighbor," in *Proc. IEEE 85th Veh. Technol. Conf. (VTC Spring)*, Sydney, NSW, Australia, Jun. 2017, pp. 1–5, doi: 10.1109/VTCSpring.2017.8108345.

[26] F. Alam, M. T. Chew, T. Wenge, and G. S. Gupta, "An accurate visible light positioning system using regenerated fingerprint database based on calibrated propagation model," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 8, pp. 2714–2723, Aug. 2019, doi: 10.1109/TIM.2018.2870263.

[27] T. Wenge, M. T. Chew, F. Alam, and G. S. Gupta, "Implementation of a visible light based indoor localization system," in *Proc. IEEE Sensors Appl. Symp. (SAS)*, Seoul, South Korea, Mar. 2018, pp. 1–6, doi: 10.1109/SAS.2018.8336711.

[28] X. Guo, S. Shao, N. Ansari, and A. Khreishah, "Indoor localization using visible light via fusion of multiple classifiers," *IEEE Photon. J.*, vol. 9, no. 6, pp. 1–16, Dec. 2017, doi: 10.1109/JPHOT.2017.2767576.

[29] T. Komine and M. Nakagawa, "Fundamental analysis for visible-light communication system using LED lights," *IEEE Trans. Consum. Electron.*, vol. 50, no. 1, pp. 100–107, Feb. 2004, doi: 10.1109/TCE.2004.1277847.

[30] J. M. Kahn and J. R. Barry, "Wireless infrared communications," *Proc. IEEE*, vol. 85, no. 2, pp. 265–298, Feb. 1997, doi: 10.1109/5.554222.

[31] M. Yasir, S.-W. Ho, and B. N. Vellambi, "Indoor positioning system using visible light and accelerometer," *J. Lightw. Technol.*, vol. 32, no. 19, pp. 3306–3316, Oct. 1, 2014, doi: 10.1109/JLT.2014.2344772.

[32] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, J. Vanderplas, A. Passos, and D. Cournapeau, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Oct. 2011.

[33] D. G. Pereira, A. Afonso, and F. M. Medeiros, "Overview of Friedman's test and post-hoc analysis," *Commun. Statist.-Simul. Comput.*, vol. 44, no. 10, pp. 2636–2653, Nov. 2015, doi: 10.1080/03610918.2014.931971.

[34] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *J. Mach. Learn. Res.*, vol. 7, pp. 1–30, Jan. 2006.

HUY QUANG TRAN received the B.E. degree in automation engineering from the Danang University of Science and Technology, Vietnam, in 2007, and the M.Eng. degree in automation engineering from the University of Transport and Communication, Vietnam, in 2012. He is currently pursuing the Ph.D. degree with the Department of Aerospace Engineering, University of Ulsan, South Korea. His current research interests include applied artificial intelligence, indoor positioning and navigation systems, visible light communication, computer vision, and the Internet-of-Things applications.

CHEOLKEUN HA (Member, IEEE) received the B.S. and M.S. degrees in aeronautical engineering from Seoul National University, in 1984 and 1986, respectively, and the Ph.D. degree from the University of Washington (UW), in 1993. In 1993, he worked as a Postdoctoral Researcher with UW. Since 1993, he has been a Professor with the University of Ulsan, South Korea. His current research interests include indoor positioning using visible light, monocular vision for industrial applications, motion control of mechanical dynamic systems, and autonomous flight control of aerial vehicles.