Journal of Electronics and Information Science (2024)
Clausius Scientific Press, Canada

DOI: 10.23977/jeis.2024.090114
ISSN 2371-9524 Vol. 9 Num. 1

Indoor Visible Light Localization Algorithm Based on KNN and Bayesian Algorithm

Pang Hua*, **Cui Dongkai, Wang Xin**

Henan Polytechnic University, Jiaozuo, China
[*]Corresponding author: 957764408@qq.com

Keywords: Visible light communication; Wireless localization; Visible light; Indoor positioning; Fingerprint localization; WKNN; Fusion algorithm

Abstract: In this paper, based on KNN and Bayesian algorithm, the basic algorithm principle of KNN (WKNN) and Bayesian is expounded. Because the KNN and Bayesian algorithm develop the signal intensity matching strategy from the perspective of mean error and probability, so one of the single algorithm can not better cope with the complex and changeable positioning scenario. For this problem, this paper proposes a new signal intensity matching criterion based on the fusion of two algorithms. The main idea of the algorithm is to change the traditional weighting method in WKNN to the weighting method considering Bayesian estimation results. In order to verify the effectiveness of the fusion algorithm, the existing visible light indoor positioning algorithm based on fingerprint recognition is compared, and the fusion algorithm based on KNN and Bayesian algorithm is proposed. This improved algorithm not only reduces the complexity of Bayesian algorithm, but also significantly improves the positioning accuracy of WKNN algorithm.

1. Introduction

With the wide application of Internet of Things (IoT) technology and mobile Internet services, positioning services play an indispensable role in People's Daily work and life. In the outdoor environment, GPS and Beidou navigation and positioning system have been relatively mature, and their positioning accuracy can reach within a few meters[1], which can basically meet people's outdoor positioning needs. However, in indoor environments, due to the complex propagation conditions, the positioning accuracy is affected by multipath effect and signal weakness. Common indoor positioning technologies (WIFI, UWB, Bluetooth, radio frequency, etc.), due to the low positioning accuracy, easy to interference, instability and other problems, it is difficult to meet the positioning requirements [2]. Therefore, in the face of indoor complex environment and multiple factors, visible light positioning technology has attracted much attention with its advantages. Visible light positioning (VLP) is based on the development of visible light communication (VLC) technology. VLC technology has the characteristics of simple equipment, high transmission rate, green and high efficiency, while visible light positioning technology has the advantages of high accuracy, low power consumption, strong anti-interference, safety and reliability. The technology is widely used in complex indoor environments, such as large supermarkets, hospitals, underground parking lots, mines and other places. Therefore, the research on indoor localization techniques based on visible light is of great significance.

Compared with traditional communication technology, visible optical communication (VLC) technology with LED VLC is less affected by multi-path effect and electromagnetic wave; therefore, it is more suitable for the limited radio communication environment of hospitals, shopping malls and mines, etc, especially indoor positioning. VLP technology has the advantages of high precision, low cost, traceability and other advantages.

For VLC system[3], this paper proposes an indoor visible light positioning algorithm based on KNN and Bayesian algorithm, the algorithm used the KNN and Bayes algorithm to calculate the measurement position, signal strength of KNN and Bayesian algorithm of a single algorithm can not better cope with the complex and changeable positioning scenario. For this problem, this paper proposes a new signal intensity matching criterion based on the fusion of two algorithms.

2. System Model

Figure 1 is a typical indoor visible light communication model. The number and location of LED can be changed according to different room size requirements. In this paper, the classic indoor VLC four-lamp model is adopted. As a signal transmitter, the LED can play the functions of lighting, communication and positioning at the same time. Each LED light is assigned its corresponding ID information. The receiving end receives and analyzes the light signal through the photodetector to obtain the position information of each LED lamp and the received signal intensity value[4].

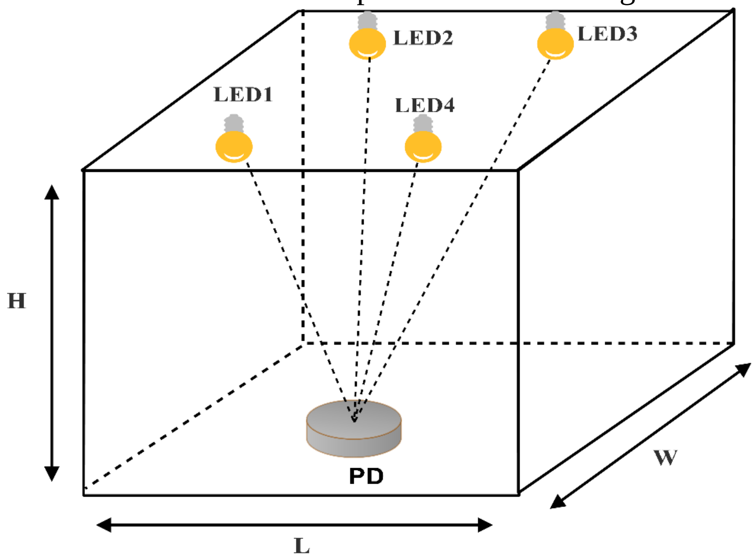

Figure 1: Visible light communication model

In the visible light communication system, the signal can be divided into direct signals and reflected signals, and in the whole VLC system the direct signal of all the largest proportion of the least, therefore, to facilitate calculation, this chapter only consider the distance of sight (LOS) link channel model, between the transmitting end and the receiving end is not affected by the barrier, the light signal from the transmitter can be directly transmitted between the receiver. Figure 1 shows the channel model diagram of the LOS link.

Introduce the Lambertian radiation model, the DC gain of the LOS channel can be expressed as:

$$H(0) = \frac{(m+1)A}{2\pi d^2} COS^m(\theta) T_s(\partial) g(\partial) COS(\partial) \delta(t - \frac{d}{c}) \tag{1}$$

3. Improved indoor visible light fingerprint positioning algorithm

Commonused indoor wireless positioning algorithm has arrival time (TOA), arrival time difference (TDOA), arrival Angle (AOA), and receive signal strength (RSSI) positioning algorithm, fingerprint positioning using multiple sensors or fixed nodes to collect fingerprint information of target equipment, and match the established fingerprint database, determine the location of the device.

Fingerprint positioning is divided into two stages: online stage and offline stage. In the offline stage, the positioning area is divided into multiple grid points. They obtain the received signal intensity information by receiving the signal from different transmitting terminals through the established grid points, which is represented by their respective fingerprints and stored in the fingerprint database.

In the online phase, position estimation can be achieved by the nearest reference point, obtained by measuring the received signal intensity indication (RSSI) value of the user position and comparing this RSSI value with a fingerprint in the database.

3.1. WKNN positioning algorithm

KNN algorithm is the K value nearest neighbor method, the main idea of the algorithm is: In a feature space, if a sample has K most similar to the sample in the space, the sample and these K samples belong to the same type. Simply put, KNN algorithm is also a classification algorithm, and different classifications are made according to the similarity of samples[5]. The specific steps of KNN algorithm are as follows:

(1). Prepare and preprocess the data.

(2). Calculate the distance between the sample to be measured and the other samples.

(3). The distance between the sample to be tested and each sample is sorted, and the K sample points with the minimum distance between the samples are selected.

(4). Classify and compare the selected K sample points, select the category with the most concentrated sample points, and summarize the sample points to this category.

In fingerprint positioning, KNN algorithm is a fingerprint identification positioning algorithm, which collects the fingerprint information of each grid point in the room, namely the RSSI value, and establishes a fingerprint information database. In the positioning stage, the measured RSSI is compared with the RSSI value in the fingerprint information database to calculate the Euclidean distance. The centroid of K grid points with the smallest distance is screened out as the position of the receiver. The formula for calculating the Euclidean distance is as follows:

$$D = \sqrt{\sum_{i=1}^{n}(rssi_i - \overline{rssi_i})^2} \tag{2}$$

Where, denotes the RSSI value of the first AP, and denotes the average of the RSSI collected by the first AP multiple times.

WKNN algorithm is the weighted K-nearest neighbor method, which is further optimized based on KNN algorithm, W is the weight, the weight reference, so that when calculating the positioning result, it is judged according to the contribution provided by K similar fingerprint data in the estimation of the positioning result. If the distance is larger, the location fingerprint point is farther away from the point to be located and measured, so the contribution of this point to the estimation of the positioning result is less. By using the weight to process the most similar position coordinates of K fingerprint data, the position coordinates obtained by using the average value directly are more convincing, the results are more realistic, and the positioning accuracy can be improved, so that the position coordinates are more accurate. Like KNN algorithm, WKNN algorithm needs to calculate the Euclidean distance between the sample to be measured and other samples, sort the distance between the sample to be measured and each sample, and screen out K sample points with the smallest distance between the samples. Finally, the final result is calculated by giving a certain weight to the fingerprint similarity of K sample points, and the calculation formula is as follows:

3.1.1. Bayesian algorithm

Bayes' theorem is about the conditional probability of random events A and B. In mathematical statistics, conditional probability can be expressed as follows:

$$P(AB) = P(A)P(B) \tag{3}$$

If event B is composed of N independent events, then:

$$P(A) = P(A \mid B_1)P(B_1) + P(A \mid B_2)P(B_2) + ... + P(A \mid B_n)P(B_n) \tag{4}$$

In fingerprint localization, the Bayesian algorithm can help to determine the probability distribution of the location of the receiver at different grid points based on the received RSSI signal data, so as to realize the localization function. The posterior probability is calculated as shown in Equation (4).

$$P(l_i \mid S) = \frac{P(S \mid l_i)P(l_i)}{P(S)} \tag{5}$$

Calculate the probability density function between the received RSSI value and the actual observed value at each grid point. Considering the independence of the signal strength between LED lights, the likelihood function of each LED can be multiplied to obtain the total likelihood function, as shown in equation (5).

$$(\, P(S \mid l_i) = P(S_1 \mid l_i)P(S_2 \mid l_i)P(S_3 \mid l_i) \cdots P(S_n \mid l_i) \tag{6}$$

In summary, when the conditional probability is maximum, the corresponding grid point is the location of the target.

3.1.2. Fusion algorithm

The above sections introduce the basic algorithm principles of KNN and Bayes, and the two algorithms formulate the signal strength matching strategy from the perspective of mean error and probability respectively. Therefore, a single use of one of the algorithms cannot deal with complex and variable positioning scenarios. To solve this problem, this subsection analyzes the positioning principle of KNN and Bayesian algorithm, and proposes a new signal strength matching criterion based on the fusion of the two algorithms. These include:

(1) Firstly, the KNN algorithm is used to calculate the distance between each collected reference point and the test point to be located, and then the previous reference point with similar information to the test point is found, and the corresponding signal strength difference $d_i\left(i = 1, 2, \cdots, K\right)$ is recorded .

(2) Based on the probability idea, the similarity $p_i\left(i = 1, 2, \cdots, K\right)$ between the K reference points selected in step 1 and the test points is calculated.

(3) Based on the WKNN idea in the previous subsection, the normalization in Step 1 is first performed, that is

$$\bar{d}_i = \frac{d_i}{\sum_{i=1}^{K} d_i} \tag{7}$$

Similarly, the normalization is performed in step two, that is

$$\bar{p}_i = \frac{p_i}{\sum_{i=1}^{K} p_i} \tag{8}$$

(4) By considering the two term weights in step three, the unknown result of the final estimation is calculated, which can be expressed as follows:

$$(\hat{x}, \hat{y}) = \frac{1}{K} \sum_{i=i}^{K} \frac{\bar{p}_i}{\bar{d}_i} (x_i, y_i) \tag{9}$$

The design of this algorithm is mainly aimed at the Bayesian location estimation algorithm. The greater the probability p_i, the more likely the corresponding i reference point is to be close to the real test point. For the traditional KNN algorithm, the smaller the signal strength error d_i, the more likely the corresponding first reference point is to be close to the real test point. It is further considered that the dimensions of p_i and d_i are not the same, so it needs to be normalized. Based on this, the final location estimation results given in this chapter are shown in equation (9).

Finally, in the field of visible light positioning, this paper changes the traditional WKNN weighting method to a weighting method that considers Bayesian estimation results. The effectiveness of the algorithm proposed in this chapter will be further explained by the experimental results.

4. Simulation and Analysis

In order to verify the effectiveness of the algorithm, this chapter uses MATLAB to simulate the four positioning algorithms of KNN, WKNN, Bayes and fusion algorithm proposed in this paper. The size of the simulation space is 5 meters long, 5 meters wide and 2 meters high. As shown in Figure 2, the test plane is divided into 25 grid points, of which 36 are reference points and 25 are test points. The signal strength of the four LED lights received by the test point is shown in Table 1. Equation (4-1) is the calculation formula of RSSI.

$$RSSI = 10log10 \frac{(m+1)A \times COS^m(\theta) \times COS(\theta)}{2\pi \| d \|} \tag{10}$$

Here, m is the Lambertian coefficient, A is the effective receiving area, d is the Euclidean norm from the LED to the point to be measured, and θ is the incidence Angle.

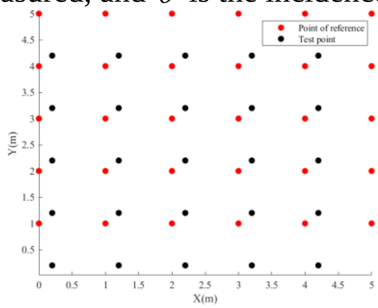

Figure 2: Experimental environment.

The main parameters of the system are shown in Table 1.

Table 1: VLC system parameters

parameters	Numerical value
Room size (length × width × height) / m	5×5×2
Grid size / (cm×cm)	100×100
Number of leds	4
LED position coordinates	A(0,5,2) B(5,0,2) C(5,5,2) D(0,0,2)
LED power / W	5
Field of view / (°)	60
Effective receptive area / cm2	1
Half power Angle / (°)	70
Optical filtering gain / W	1

In order to verify the effectiveness of the algorithm, the positioning errors of KNN algorithm, WKNN algorithm and fusion algorithm under different K values are compared. The error meter probability distribution of the localization algorithm is shown in the following figure.

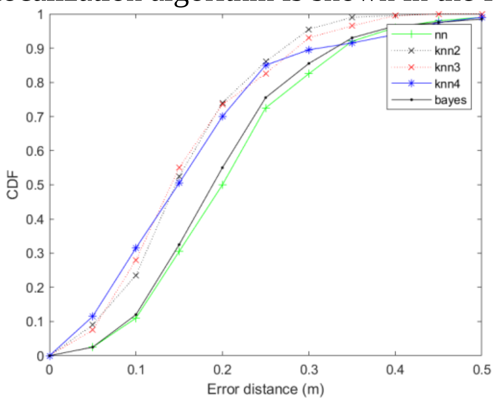

Figure 3: Cumulative probability distribution of KNN positioning error

Figure 3 shows the positioning error diagram of KNN. It can be obtained from Figure 4-5 that the KNN positioning algorithm is used to calculate one of the grid points, and the positioning error is also different with different values of K. When K=1, the CDF of positioning error reaches 70% within 0.52cm, and 90% within 0.75cm. When K=2, the CDF of the localization error reaches 70% within 0.41cm, and can reach 90% within 0.53cm. When K=3, the CDF of the localization error reaches 70% within 0.34cm, and it can reach 90% within 0.61cm. When K=4, the CDF of the localization error reaches 70% within 0.33cm and can reach 90% within 0.5cm. Under this model, the probability that the positioning error of Bayes positioning algorithm is 0.47m is 70%, and the probability that the error is 0.67m is 90%. Due to the limitations of the Bayes algorithm, the positioning error is slightly larger.

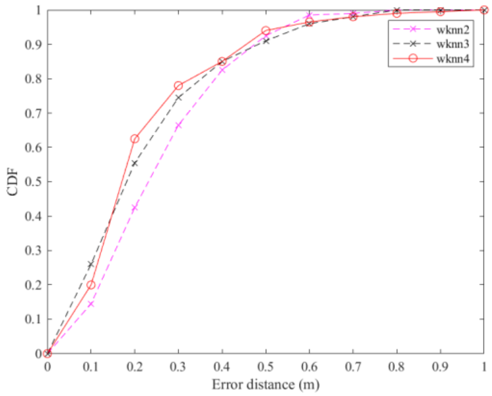

Figure 4: Cumulative probability distribution of WKNN positioning error

Figure 4 shows the positioning error diagram of WKNN under different K values. The WKNN positioning algorithm is used to calculate one of the grid points, and the positioning error is also different when the value of K is different. When K=2, the probability that the error of one grid point is 0.32cm is 70%, and the probability that the error reaches 0.47cm is 90%. When K=3, the probability that the localization error reaches 0.25cm is 70%, and the probability that the error reaches 0.44cm is 90%. When K=4, the probability that the positioning error reaches 0.22cm is 70%, and the probability that the error reaches 0.42cm is 90%. Figures 4-11 and 4-12 show the positioning effect when K=2 and K=3, respectively. In the process of increasing the value of K further, we may find a more appropriate value for K. However, after considering the practicability and effectiveness of the algorithm, K=3 is chosen as the best choice in this paper. It is worth emphasizing that the minimum localization error obtained for K=3 is only applicable to the specific scenario set in this study. In different scenarios, the value of K should be set according to the specific situation. The subsequent simulation works are all carried out based on K=3.

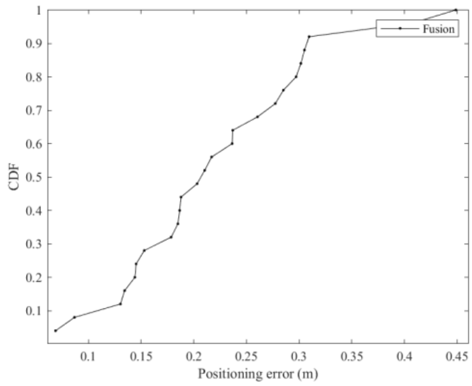

Figure 5: Cumulative probability distribution of positioning error of fusion algorithm

Figure 5 shows the cumulative probability distribution of error of the fusion localization algorithm based on KNN and Bayesian algorithm. Through experimental analysis, the localization error of the fusion algorithm can reach the minimum of 0.03m, and the probability of the localization error within 0.43m is 90%, and the average error is 0.25m.

In order to further verify the effectiveness of the fusion algorithm, the positioning error of the three positioning algorithms at each test point is analyzed and compared. Figure 6 shows the positioning accuracy map of different algorithms at each test point. It can be seen from the figure that the positioning error of the WKNN positioning algorithm is mainly distributed between 0.15m and 0.3m. The positioning error of the fusion algorithm is mainly distributed between 0.1m and 0.2m, indicating that compared with other algorithms, the positioning error of the fusion algorithm at different test points tends to be stable, the fluctuation of the error value is relatively small, and the fusion algorithm has higher positioning accuracy.

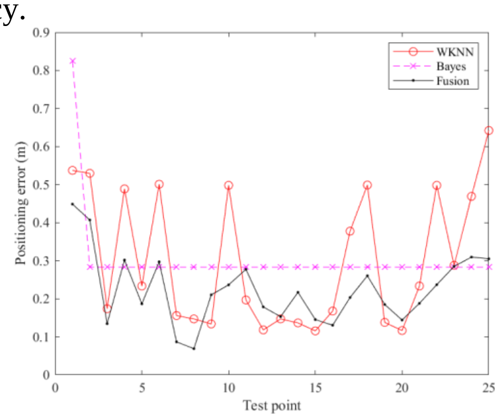

Figure 6: Test point positioning error

Through a series of comparative studies, it is finally inferred that the fused algorithm shows obvious practicality and feasibility in practical applications. The algorithm can put forward a new signal strength matching criterion and improve the positioning accuracy of the algorithm under the premise of ensuring the location accuracy.

5. Conclusions

The signal strength matching strategy of the two algorithms is improved, and a new signal strength matching strategy is proposed. Comprehensive simulation analysis shows that the improved fusion algorithm has higher positioning accuracy than KNN, WKNN and Bayesian positioning algorithm. The minimum positioning error of the fusion algorithm can reach 0.03m, and the average positioning error can reach 0.25m. Compared with KNN, WKNN and Bayesian positioning algorithms, the average positioning accuracy of the fusion algorithm is improved by 26%. This fusion algorithm is expected to be more effective in practice to complete the position positioning, while maintaining the positioning accuracy has significant practical application significance.

References

[1] Hong, Wei, et al. "Multibeam antenna technologies for 5G wireless communications." IEEE Transactions on Antennas and Propagation 65.12 (2017): 6231-6249.
[2] Yang, Se-Hoon, Eun-Mi Jung, and Sang-Kook Han. "Indoor location estimation based on LED visible light communication using multiple optical receivers." IEEE Communications Letters 17.9 (2013): 1834-1837.
[3] Xu, He, et al. "An RFID indoor positioning algorithm based on Bayesian probability and K-nearest neighbor." Sensors 17.8 (2017): 1806.
[4] Huan, Hai, et al. "Indoor location fingerprinting algorithm based on path loss parameter estimation and bayesian inference." IEEE Sensors Journal 23.3 (2022): 2507-2521.
[5] Deng, Zhi-An, et al. "Carrying position independent user heading estimation for indoor pedestrian navigation with smartphones." Sensors 16.5 (2016): 677.