Received June 11, 2019, accepted June 21, 2019, date of publication June 26, 2019, date of current version July 17, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2924992

A Novel High Precision and Low Consumption Indoor Positioning Algorithm for Internet of Things

JIN REN[1,2], YUNAN WANG[1], CHANGLIU NIU[1], AND WEI SONG[3]

[1]School of Information Science and Technology, North China University of Technology, Beijing 100144, China
[2]Beijing Key Laboratory on Integration and Analysis of Large-Scale Stream Data, Beijing 100144, China
[3]School of Information Engineering, Minzu University of China, Beijing 100081, China

Corresponding author: Jin Ren (rj@ncut.edu.cn)

This work was supported in part by 2017 Annual Outstanding Young Teacher Training Program Project of North China University of Technology under Grant XN019009, in part by Scientific Research Project of Beijing Educational Committee under Grant KM201710009004, in part by 2018 Science and Technology Activities Project for College Students of North China University of Technology under Grant 110051360007, in part by Research Project on Teaching Reform and Curriculum Construction of North China University of Technology under Grant 18XN009-011, in part by 2019 Beijing University Student Scientific Research and Entrepreneurship Action Plan Project under Grant 218051360019XN004, in part by 2019 Education and Teaching Reform General Project of North China University of Technology under Grant 108051360019XN141/021, and in part by 2019 Fundamental Research Funds for Beijing Universities under Grant 110052971921/004.

ABSTRACT Internet of Things (IoT) is digitizing the world, and indoor positioning is one of the important applications of them. Indoor positioning refers to the realization of positioning in the indoor environment. The recent research on indoor positioning focuses on Wi-Fi-based methods since GPS cannot achieve the desired effect. A core algorithm in those methods is the K nearest neighbor (KNN) search. In this paper, we proposed an improved indoor positioning algorithm named IpKNN with better accuracy and efficiency. The IpKNN mainly includes two parts. The first part is to use the proposed clustering algorithm to classify the data set, which can improve the computational efficiency. The second part is to improve the positioning accuracy by using the proposed KNN algorithm. The proposed algorithm can achieve high precision and low consumption, and the experiment results also proved it.

INDEX TERMS Internet of Things (IoT), indoor positioning, clustering, KNN algorithm, IpKNN.

I. INTRODUCTION

The Internet of Things (IoT) [1]–[4] is the next technological revolution that affects all areas of application. It is expected to be large and widespread, with more than 50 billion smart items and objects connected by 2020, and many more. It will affect many applications, from agriculture to smart cities, Industry, energy and transportation. The exact number of IoT microfilms in the long term is not realistic and predictable. However, we can be sure that IoT will be huge and growing.

The core and foundation of IoT technology is still Internet technology. It is a network technology that extends and expands on the basis of Internet technology. Its client extends and extends to any item and item for information exchange and communication.

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

Internet of Things is the transmission and control of information between things and things, or between people and things [5], [6]. There are three key technologies in IoT applications. The first is sensor technology, which is also a key technology in computer applications. Everyone knows that most computers have processed digital signals so far. Since the computer has been required, the sensor needs to convert the analog signal into a digital signal computer. Secondly, RFID tag is also a sensor technology. RFID technology is a comprehensive technology that integrates radio frequency technology and embedded technology. RFID has broad application prospects in automatic identification and item logistics management. Finally, the embedded system technology is a complex technology that integrates computer hardware and software, sensor technology, integrated circuit technology and electronic application technology. After decades of evolution, smart terminal products featuring embedded systems

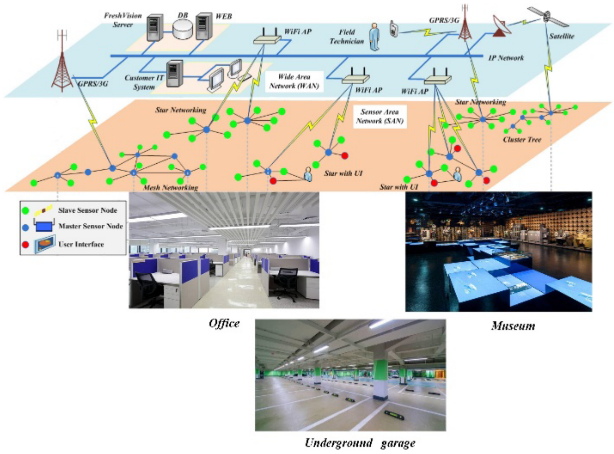

FIGURE 1. Internet of Things and positioning.

can be seen everywhere; from small MP3s to aerospace satellite systems. Embedded systems are changing people's lives and driving industrial production and the development of the defense industry. If the Internet of Things is a simple metaphor for the human body, the sensor is equivalent to the human eye, nose, skin and other senses, the network is the nervous system used to transmit information, the embedded system is the human brain, and the information is classified after receiving the information to deal with. This example is a very vivid description of the location and role of sensors and embedded systems in the Internet of Things.

Positioning is one of the common applications of the Internet of Things. GPS-based outdoor positioning is the most common positioning method, which has good signal, high positioning accuracy and wide application range. Indoor positioning technology based on wireless networks has gradually become a research hotspot. The relationship between the Internet of Things and location is shown in Figure 1. We will focus on the discussion of indoor positioning technology.

By only using GPS [7]–[9], it is difficult to solve some locating problems on indoor environments. This is mainly because the GPS signal power is very low, and the antenna is highly affected by the obstruction between the indoor environment the sky. Sometimes it can be positioned by the window, because it is possible to receive satellite signals for locations with a largely visible sky. In addition, because most of the houses are now constructed with reinforced concrete, the GPS signal cannot enter the room due to the obstruction and reflection of the walls. In the present days, popular indoor positioning technologies consist of ultrasonic positioning, radio frequency positioning and ultra-wide band positioning. Although these methods have good positioning effects, they have many problems such as high cost, small positioning range and low flexibility.

With the development of technology, Wi-Fi [10]–[12] has covered most homes and public places. In order to provide Wi-Fi locating service, access points (APs) will be deployed in indoor areas. Each wireless AP has a globally unique MAC address, and generally the wireless APs will not change for

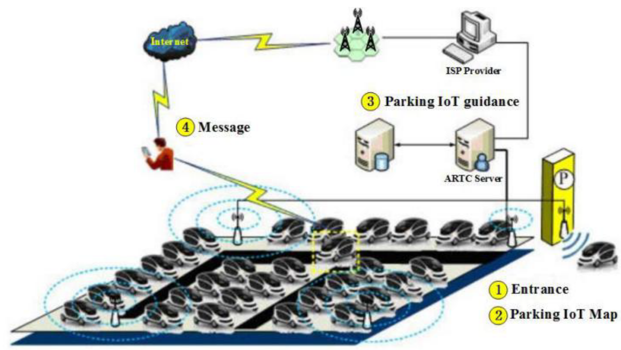

FIGURE 2. Indoor positioning diagram.

a period of time. The Wi-Fi terminal can constantly scan and collect surrounding AP signals and obtain their signal strengths. Therefore, people have developed an indoor positioning method by processing the signal strength of Wi-Fi, which is a very cost-effective method. The traditional method of locating is based on time and angle include AOA (Angle of Arrival) [13], [14], TDOA (Time Difference of Arrival) [15], [16], which cannot be adopted for Wi-Fi. In the scenario of indoor Wi-Fi locating, the locating method based on fingerprint is widely studied and adopted.

Indoor positioning using Wi-Fi requires placing a certain number of wireless access points indoors, so that the entire indoor environment is covered by Wi-Fi signals in Figure 2. Then use the mobile device to get the RSSI and mac address from the AP. The RSSI signals from different wireless access points collected by the mobile terminal at one location are collated and stored to obtain a set of RSSI vectors, and the RSSI values in the vector correspond to each wireless access point. Collecting at different locations yields an array RSSI vector. These vectors are stored in the database with the corresponding indoor locations. Suppose we have five wireless access points, the data stored in the database is as follows:

$$S = [(p_{i,1}, p_{i,2}), (r_{i,1}, r_{i,2}, r_{i,3}, r_{i,4}, r_{i,5})] \tag{1}$$

where $(p_{i,1}, p_{i,2})$ is the coordinate of the i-th position, and $(r_{i,1}, r_{i,2}, r_{i,3}, r_{i,4}, r_{i,5})$ is the RSSI vector collected by the i-th position. After the database is built, the RSSI signal of a certain location is collected again by the mobile device, and uploaded to the positioning server, and the vector is compared and calculated with the vector in the database, and finally the calculated position is returned by the positioning server.

Wi-Fi based fingerprint locating is usually divided into two stages: offline collection and online positioning. In the offline stage, a Wi-Fi signal detecting instrument should be used to collect RSSI [17], [18] from each AP in its positioning range, as well as to sort out and generate fingerprint database. In order to make the collected data stable and reliable, the data is typically filtered. In the online stage, the data measured in real time is compared with the fingerprint database. After which the final location is obtained through positioning algorithms.

Through the research, we found that many scholars have improved the KNN algorithm. Mohsen proposed a Weighted Differential Coordinate Probabilistic-KNN (WDCP-KNN) method based on probabilistic weighting of generalized Reference Points and differential coordinates [19]. Zihan Liu proposed to use the correlation between main neighbor and (K-1) auxiliary neighbors, and it combining it with the variance weighting method [20]. Lei Yen proposed the differential coordinate based WKNN using Wi-Fi technology to further improve the accuracy [21]. Long Cheng proposed an improved weighted KNN algorithm to calculate the final positioning coordinates of the measurement point [22].

Clustering algorithms are also often used in positioning algorithms to increase efficiency. ChoRong Park adopted Improved K-means clustering for an efficient and accurate classification in KNN algorithm [23]. Hao jiewang compare positioning performance of three clustering algorithms which are spatial clustering, K-means clustering and Affinity Propagation Clustering [24].

Above research can improve positioning accuracy and reduce consumption, but only unilateral scheme is put forward. Our research will suggest a new plan include improving accuracy and reducing consumption.

II. TRADITIONAL ALGORITHM

A. KNN ALGORITHM

The KNN algorithm is the most commonly used algorithm in Wi-Fi positioning. It calculates the Euclidean distance between the online RSSI vector and the RSSI vector in the fingerprint database [25], [26]. Then it selects the K fingerprint positions close to the distance calculated, and averages the coordinates of those to obtain the estimated coordinates as the result. The algorithm is shown in equation (2) and (3).

$$(d_j) = \left(\sum\nolimits_{i=1}^{n} |rss_i - \overline{rss}_i|^2\right)^{\frac{1}{2}} \tag{2}$$

$$(\bar{x}, \bar{y}) = \frac{1}{\mathrm{K}} \sum\nolimits_{i=1}^{\mathrm{K}} (x_i, y_i) \tag{3}$$

where n denotes the number of access points detected on the targeted location, rss_i represents the signal strengths of the test points which received by the APs, d_j represents the Euclidean distance between the test point and the j-th reference point, K is the number of the previous reference points, (x_i, y_i) represents the coordinates of the i-th reference point in the previous K reference points. $(\bar{x}, \bar{y})$ represents the resulting estimated coordinates of the targeted location.

The KNN algorithm uses its regression model in the localization domain, which assigns the average of the neighbor properties near the sample to the sample. This makes the KNN algorithm have the following disadvantages when used for positioning: each calculation requires traversing all the fingerprint points in the database, and the calculation efficiency is low [27]; and the calculated neighbor points may be far away from the sample. This point is used as a calculation would lead the final positioning coordinates sometimes deviates too far from the actual position. We decided to improve the indoor positioning process in two ways. The first is to add a clustering algorithm before KNN algorithm [28]; the second is to combine the KNN algorithm with other algorithms, the main purpose is to solve the KNN algorithm through the neighbors the point coordinates average get the instability of the positioning result and optimize the positioning result.

The usual KNN algorithm requires large calculation amount and long calculation time, which leads to inefficiency [27]. The accuracy of the algorithm fluctuates due to the instability of RSSI. Therefore, using the clustering algorithm is an effective way to reduce the number of fingerprint points required during calculation [28].

B. FUZZY C-MEANS CLUSTERING

Fuzzy C-mean algorith (FC) mainly consists of two parts, one is the fuzzy theory and the other is the C-means algorithm [29], [30].

The input of FCM is a data set of pending clustering, and each element has p features. The output is a matrix with c lines and n columns named U, where c is the amount of clusters, n is the number of elements of the data set; we can use this matrix to represent the result of the classification [31]. A column in the matrix represents the degree to which this element belongs to each class, and which the value is the largest represents which class this element belongs to equation (4) is FCM's value function.

$$J_f = \sum\nolimits_{j=1}^{c} \sum\nolimits_{i=1}^{t} \left[u_j(x_i)\right]^b \left\|x_i - m_j\right\|^2 \tag{4}$$

where $u_j\left(x_j\right)$ is the membership function of the i-th sample corresponding to class j. $\left\|x_i - m_j\right\|^2$ is the Euclidean distance from the x-th data and the j-th clustering center. t is the amount of the dataset.

The FCM clustering algorithm is as follows: First, set the clustering number c and the iterative convergence condition to initialize the clustering centers [32]. Second, calculate the membership matrix using the current cluster center based on equation (5), and calculate the new center of clustering using the current membership function based on equation (6). Then, after achieving the convergence condition, obtain the final cluster centers and membership degree matrix U.

$$u_j(x_i) = \frac{\left\|x_i - m_j\right\|^{-2/(b-1)}}{\sum_{s=1}^{t} \|x_i - m_s\|^{-2/(b-1)}} \tag{5}$$

$$m_j = \frac{\sum_{i=1}^{n} \left[u_j(x_i)\right]^b x_i}{\sum_{i=1}^{n} \left[u_j(x_i)\right]^b} \tag{6}$$

where $u_j(x_i)$ is the Membership matrix, m_j is the clustering center, $\sum_{i=1}^{n} \left[u_j(x_i)\right]^b = 1$, b is the weighted coefficient, usually equals to 2.

The advantage of the clustering method is that before the clustering algorithm is used as the fingerprint indoor positioning, clustering the fingerprint point database is an effective method to improve the calculation efficiency and reduce the calculation amount. After the clustering is completed, the fingerprint database is classified. When the real-time positioning is performed, the RSSI vector of the anchor point measured by the mobile terminal and the RSSI vector of the cluster center point are calculated by Euclidean distance, and the class with the smallest distance is selected as the new fingerprint database. The new fingerprint database has similar data characteristics and less data volume, which effectively solves the problem that the KNN fingerprint localization algorithm has large computational complexity and many invalid calculations.

III. PROPOSED ALGORITHM

A. PROPOSED CLUSTERING ALGORITHM

We use the characteristics of FCM algorithm membership matrix to propose an improved clustering algorithm clustering algorithm, which is an improvement of the original algorithm FCM clustering algorithm. Its innovation is that each class has a common part after clustering, which is at the boundary of the cluster. It can improve the accuracy of positioning. Based on the FCM clustering algorithm, the improved clustering algorithm adjusts the clustered membership probability matrix by the control variables. When the fingerprint points have similar membership degrees to different classes, it is judged that this point is a common point and is shared by different classes. The advantage of the improved clustering algorithm is that it can improve the positioning accuracy of the cluster boundary area. The main method of KNN algorithm is to obtain the positioning result according to the average of the coordinates of the nearest neighboring K fingerprint points near the positioning point. The clustering algorithm will make the points which is near the class and class boundaries unable to obtain accurate neighbors. When the point near the class A (a red area in Figure 3) is positioned, the nearest neighbor is in the class B (a green area in Figure 3) and cannot be acquired, so the positioning accuracy is lowered.

Figure 3 shows the results of 100 random points clustered by FCM clustering algorithm in a planar area. However, when positioning in the cluster boundary area, because of the influence of the classification, some reference points closer to the Euclidean distance was assigned to another. In this case, we can't use these neighbor points to locate. This leads to a reduction in positioning accuracy.

Therefore, we propose to take the point in the middle of the plane as the common part of classification. When the degree membership of fingerprint is similar, then add it to these classes at the same time. The new clustering results are shown in the Figure 4. New classifications are blue area and brown area. These two kinds of public parts have effectively solved the problem of large distance error in the calculation of fingerprint points after classification. The area depends on equation (7).

$$|p_1 - p_2| < \alpha \tag{7}$$

where p is the membership probability of different pairs of fingerprint points. α is the threshold that determines the size of the public area. The sum of membership probabilities

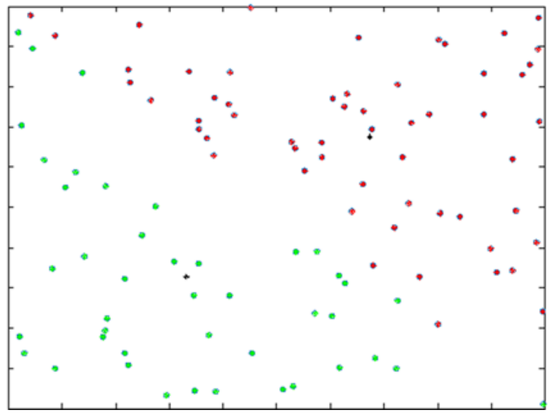

FIGURE 3. Original FCM algorithm classification diagram.

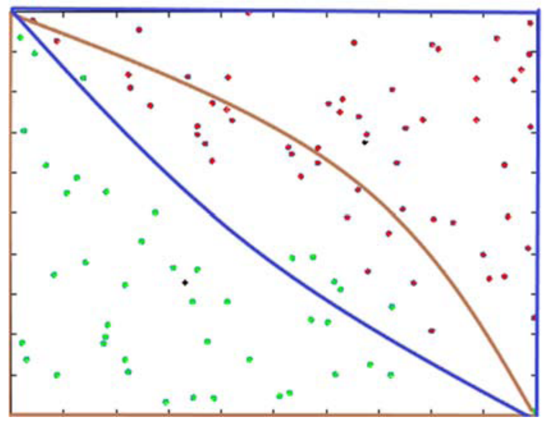

FIGURE 4. Improved clustering algorithm classification diagram.

is 1. In this way, the number of reference points can be effectively reduced and the efficiency can be improved when the accuracy is guaranteed.

Although Improved clustering algorithm (Ipc) requires more fingerprint points than the traditional algorithm, the positioning accuracy is better than the traditional clustering algorithm.

B. PROPOSED KNN ALGORITHM

The general KNN algorithm has low positioning accuracy due to the instability of indoor environment. With KNN algorithm, a target tag must choose K nearest reference neighbor tag to calculate the coordinate position. In the improved KNN algorithm, we first get the position of a target with K nearest reference neighbor tags. After identifying the one target, this solution will estimate this K nearest neighbors' new position step by step [33]. The improved KNN Algorithm's principle is to do KNN algorithm for K reference points with known positions, get their positioning coordinates, and then calculate the error between the original position and the positioning position. Sum the deviation value and take the average. The algorithm as shown in equation (8):

$$(\Delta x, \Delta y) = \left(\frac{1}{\mathrm{K}}\sum\nolimits_{i=i}^{\mathrm{K}}\left(x_i - x_i'\right), \quad \frac{1}{\mathrm{K}}\sum\nolimits_{i=1}^{\mathrm{K}}\left(y_i - y_i'\right)\right) \tag{8}$$

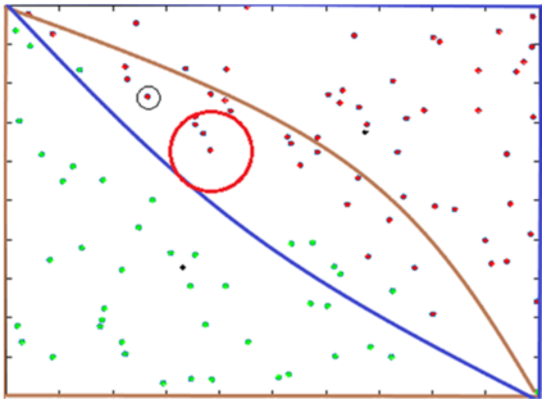

FIGURE 5. Reference point screening diagram.

where (x_i, y_i) is the actual coordinate of the reference point, (x_i', y_i') is the calculated position coordinate. $(\Delta x, \Delta y)$ is the deviation value of the location in this environment.

In the calculation of reference point coordinate difference, in order to avoid some point distance value too large to affect the accuracy of the final deviation value, we propose a new method to reduce the bias value. We adopt the method of threshold screening. If the distance value at some points is greater than the threshold, get rid of this reference point and select the next neighbor point.

An example is shown in Figure 5. We select the nearest neighbor fingerprint point of location point as the center of the circle, make a circle with d_0 as the radius like the red circle, the threshold of d_0 depends on the size of the room, then determine whether the first K neighbor points are in the circle. If a neighbor point is too far from the center of the circle to be in the circle, such as the point in a black circle in Figure 5, on behalf of this neighbor fingerprint point is discrete, which is not meet the requirements. We will eliminate this neighbor and select the K + 1 neighbor in order. Repeat this step until all neighbors meet the requirements. The area depends on equation (9).

$$\sqrt{(x_i - x_0)^2 + (y_i - y_0)^2} < d_0 \tag{9}$$

where (x_i, y_i) are the coordinates of the fingerprint points, (x_0, y_0) is the coordinate of the center in the circle. The final result is shown in equation (10).

$$(x, y) = \left(x', y'\right) + (\Delta x, \Delta y) \tag{10}$$

where (x', y') is the positioning coordinate obtained by the original KNN.

The focus of the screening method is the value of d_0. When d_0 is too large, the detection points may be covered by the circle and cannot be screened. When d_0 is too small, there are too few neighbors in the range of the hour circle, and the normal neighbors are excluded. After the screening, the KNN algorithm will increase the error. The number of detection points and the screening range also need to be adapted. Therefore, factors such as indoor size and fingerprint point space density should be considered when selecting the value of d_0. Generally, before positioning, we need to use the control variable or cross-validation method to get the best screening radius. By controlling the values of other indoor parameters, we select some known points as test points, and the remaining points as the reference points for positioning test. Then change the value of d_0, observe the change of the positioning result, and select the value of d_0 when the average error is the smallest as the optimal screening radius in this indoor scene.

TABLE 1. Algorithm process overview.

Proposed KNN algorithm
Input: (x_0, y_0)=the coordinates of the nearest neighbor fingerprint point of location point; d_0 depends on the size of the room.
Output: (x, y)= the final positioning result.
1. Make a circle with (x_0, y_0) as the center of the circle, d_0 as the radius.
2. Screen the first K neighbor points through the following formula, if a point does not satisfy the formula, it is not on the circle, eliminate the point and select the K + 1 neighbor point in order. $\sqrt{(x_i - x_0)^2 + (y_i - y_0)^2} < d_0$
3. Repeat step2 until there are K neighbor points meet the requirements.
4. Use the proposed KNN algorithm to get the location error of neighbor points. $(\Delta x, \Delta y) = \left(\frac{1}{K}\sum_{i=i}^{K}(x_i - x_i'), \frac{1}{K}\sum_{i=1}^{K}(y_i - y_i')\right)$
5. Get the final positioning result (x, y) through the following equation. $(x, y) = (x', y') + (\Delta x, \Delta y)$

After optimization, the discrete fingerprint points will be rejected. The procedure is as follows Table 1.

Follow Table 1, combine this screening method into the improved algorithm which can further improve positioning accuracy. Since all neighbors near the measurement point are to be calculated, as long as one of the neighbors does not belong to the vicinity of the point to be measured, the final deviation vector will be affected. Our screening method is combined with the improved KNN algorithm, so that all the calculated neighbors are within the range of the point to be measured, and the neighbors with no actual distance are affected by the calculation results.

In the above we introduced the proposed KNN algorithm and the proposed clustering algorithm. We combined proposed clustering algorithm and proposed KNN algorithm in order to get better positioning results. We named it IpKNN. The positioning flow chart is shown in Figure 6. We first use proposed clustering algorithm to improve computational efficiency, reduce the consumption, and then use proposed KNN algorithm to improve positioning accuracy.

As shown in the flow chart, our main improvements to Wi-Fi positioning are divided into two categories. One method is to add a clustering algorithm before KNN positioning algorithm and improve it based on the traditional FCM clustering algorithm. This method is mainly for the computational

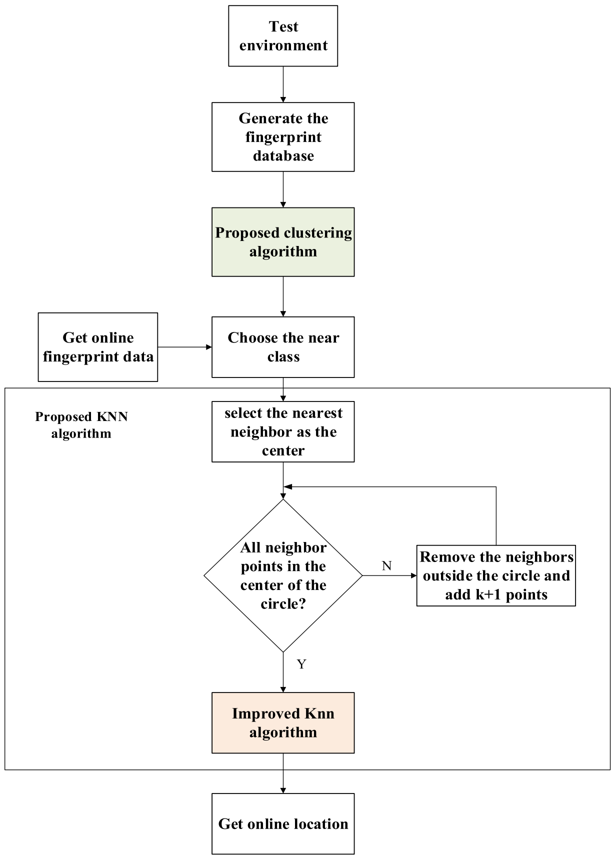

FIGURE 6. Positioning flow chart with IpKNN.

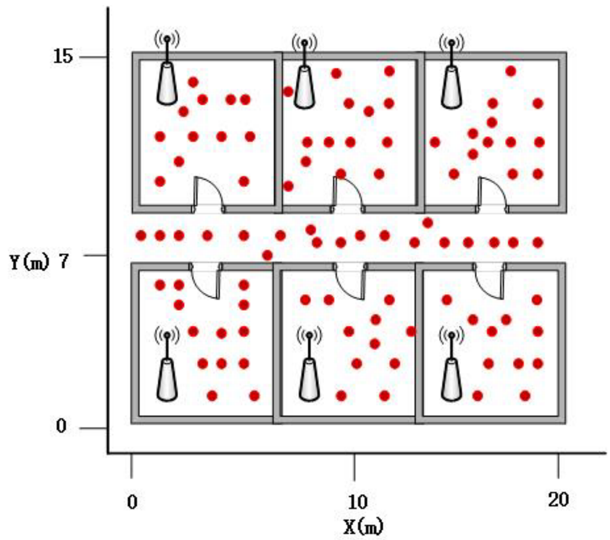

FIGURE 7. Layout of the indoor positioning system.

efficiency problem of KNN algorithm. Another method is to perform the screening operation after KNN algorithm obtains the neighboring point. The main purpose of the method is to improve the positioning accuracy.

In summary, we have improved the method of indoor positioning of Wi-Fi fingerprints on the platform based on Internet of Things. Firstly, the shortcomings of traditional KNN algorithm in computational efficiency and positioning accuracy are analyzed. Then decided to improve from these two aspects, using improved clustering algorithm to improve computational efficiency; using screening methods to improve positioning accuracy. We will analyze the positioning performance of IpKNN algorithm through experiments.

IV. EXPERIMENTAL CLASSIFICATION RESULTS AND ANALYSIS

We obtain position fingerprint data through the office as Figure 7. The test area is 20 meters long and 15 meters wide. There are 6 Aps in this area. Divide the area into 300 square cells and receive the test RSSI in the lower right corner of each cell, so we obtained 266 ($19 * 14$) distributed reference points (red dot) in this space. Each fingerprint point contains a vector of 6 RSSI values. We selected 50 fingerprint point as test point, the remaining fingerprint points as reference points. The method of selecting test points is uniform sampling, and one fingerprint point is selected as a test point at a certain distance. After the selection is completed, the test points are reasonably distributed indoors. We use the traditional positioning algorithm and the improved positioning algorithm to locate the collected data and compare the positioning effects. We make the performance indicator with the average value of the positioning error obtained by multiple tests.

We need to determine the optimal number of clusters for the experiment before performing the positioning test. There are many ways to select the optimal number of clusters. We select Sum of Squared Error (SSE) as the indicator to select. SSE represents the sum of the squared errors of each sample data for each level or group and its group mean. When the number of clusters is increasing, the data in each class will be reduced accordingly, and the distance will be closer, and SSE will also decrease. We pay attention to the change of the amplitude of SSE curve. When the magnitude of SSE reduction is not obvious, it is considered that continuing to increase the number of clusters cannot continue to increase the clustering effect. The curve of SSE is shown in Figure 8.

From the figure, we can see that SSE has a significant decrease when the number of clusters is from 1 to 2. From 2 to 5, the magnitude of SSE reduction tends to be flat, so we choose the optimal cluster number to be 2.

Then we control the size of α in the improvedclustering algorithm to determine the number of fingerprint points in the public area. The relationship of them is shown in the Table 2.

As shown in the above table, too few public points will lead to the improved clustering algorithm not obvious, too many public points will affect the efficiency of calculation. Therefore, we seek a trade-off between the calculation efficiency and positioning accuracy, then choose $\alpha = 0.15$. In this way, the function of the improved clustering algorithm clustering algorithm can be utilized to improve the positioning accuracy

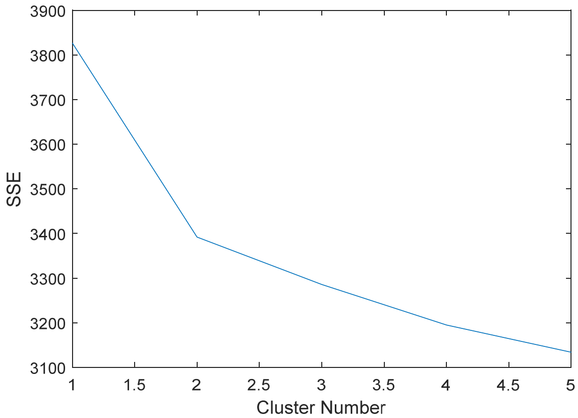

FIGURE 8. SSE vs number of clusters.

TABLE 2. The relationship between α and public points.

α	Public points
0.01	2
0.05	9
0.1	15
0.15	24
0.2	30
0.25	42

and reduce the number of fingerprint points that need to be used.

We select K = 5 in the KNN algorithm and $\alpha = 0.15$ as the threshold after control variable tests. In the screening process, considering the size of the room, the size of each cell is 1m×1m. We obtained the best screening radius of 4m by cross-validation. In this case, neighbor points greater than 4 meters away will be removed. If this number is too large, the effect of screening is not significant, and if it is too small, there is not enough data in the circle. After many tests, we find that it is the most appropriate to select $d_0 = 4$ according to the size of this indoor space.

For the accuracy of the results, we take the error mean of multiple tests as rmse. We record the positioning average error of 50 test points after the calculation and reselect 50 test points to calculate again. Repeat this step 20 times, sum each distance error and take the mean to get the final positioning indicator. The comparison algorithm we choose is the traditional KNN algorithm and the WDCP-KNN algorithm of reference 16 and the improved WKNN algorithm of reference 19.

We first perform a separate performance test on the proposed KNN algorithm. The rmse of different KNN algorithms under Gaussian noise of different intensity is shown in Figure 9. From the figure we find that IpKNN algorithm is better than the other positioning algorithms in positioning performance.

After using the screening method and cluster method, the IpKNN algorithm has lower positioning accuracy than other Improved KNN algorithm. When using the Improved KNN algorithm for positioning, when Gaussian intensity

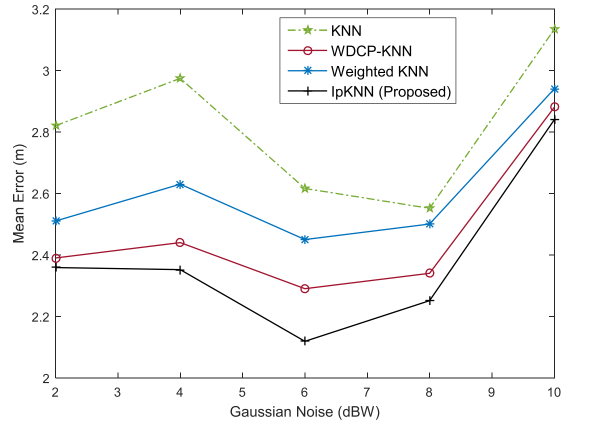

FIGURE 9. Rmse of four positioning algorithms under different Gaussian noise.

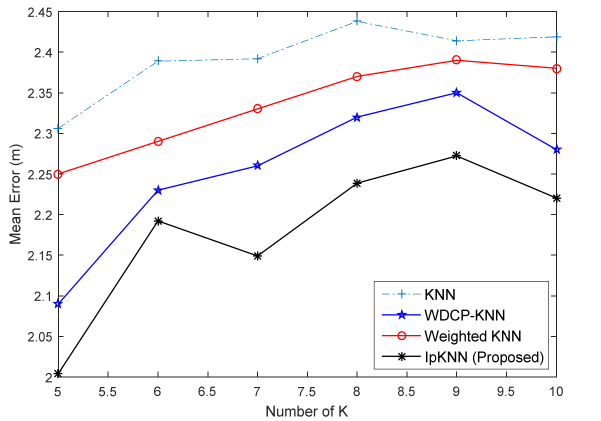

FIGURE 10. Comparisons of positioning effects of different positioning algorithms under different K values.

is 2, 4, 6, 8, 10, the average error is 2.36m, 2.35m, 2.12m, 2.24m, and 2.84m.

Then we change the number of K and observe the variation of the average positioning error of the four algorithms under different conditions. The result is shown in the Figure 10. From the figure, we can see that K from 5 to 10, the average positioning error of the four algorithms is generally rising. Whatever the value of K is, the average error of IpKNN algorithm is the lowest.

Then we observe the cumulative probability distribution curve of the four algorithms for the positioning error of 50 test points with K being 5, and the results are shown in Figure 11 and Table 3.

It can be seen from the figure and the table that the average positioning error of IpKNN algorithm is the smallest, 64% of the 50 reference points have a positioning error within 2 meters, and 98% of the 50 reference points have a positioning error within 4 meters. The average positioning accuracy of IpKNN algorithm is 0.09 meters higher than that of WDCP-KNN algorithm, and 0.25 meters higher than Weighted KNN algorithm. In terms of computational efficiency, the number of fingerprint points used in IpKNN algorithm is less than other algorithms. The IpKNN algorithm

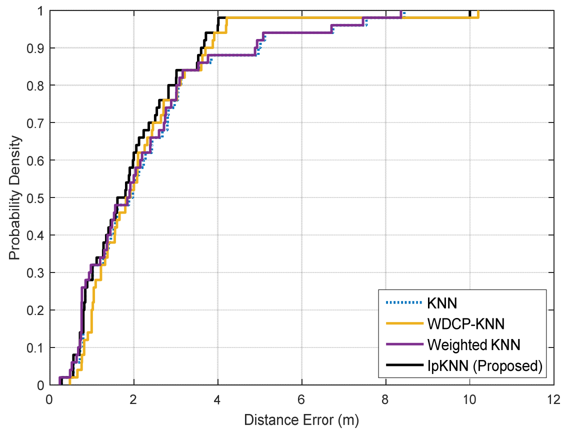

FIGURE 11. Four algorithms cumulative probability distribution curve.

TABLE 3. Comparison of four algorithms.

Algorithm	Using points	RMSE (m)	Distance error within 2m	Distance error within 4m
KNN	216	2.31	50%	88%
WDCP-KNN	216	2.09	52%	94%
Weighted KNN	216	2.25	54%	88%
IpKNN (proposed)	152	2.00	64%	98%

requires 152 fingerprint points, and other algorithms require 216 fingerprint points.

The WDCP-KNN algorithm does not take into account the accuracy of the nearest neighbors calculated by KNN; the Weighted KNN algorithm improves the positioning by the weight method, and the positioning accuracy is not significantly improved. IpKNN considered the selection of neighboring points in the cluster boundary region in clustering. In the positioning calculation, the neighboring points with far distance are screened, and the positioning accuracy is greatly improved by two-step optimization.

In order to verify the integrity of the IpKNN algorithm, we chose another area II for positioning. The laboratory is 16 meters long and 12 meters wide. We collected 165 ($15*11$) RSSI vectors as reference point data in this classroom. These reference points are evenly distributed in the classroom. We selected 40 reference points as test points for location algorithm verification. These test points can cover the entire area.

First, we analyze the positioning of the test fingerprint points under different Gaussian noises. We simulate a Gaussian noise signal from 2 to 10. The comparative positioning algorithm is still the four algorithms of the previous experiment. The experimental results are shown in Figure 12.

From the figure we can clearly see that the average positioning error of IpKNN algorithm is the lowest under the influence of Gaussian noise in each segment. When using the Improved KNN algorithm for positioning, when Gaussian intensity is 2, 4, 6, 8, 10, the average error is 2.4m, 2.14m, 2.34m, 2.43m, and 2.51m.

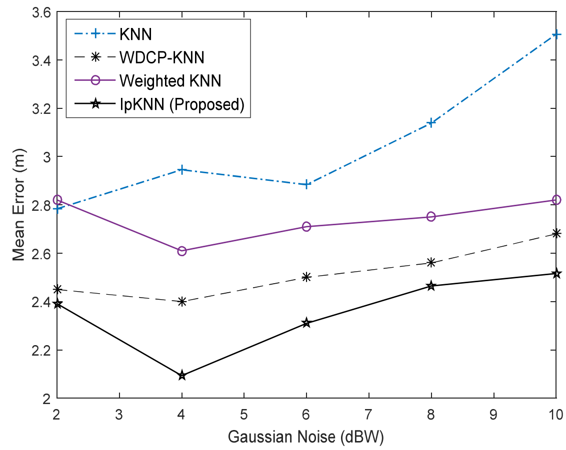

FIGURE 12. Rmse of four positioning algorithms under different Gaussian noise in new area.

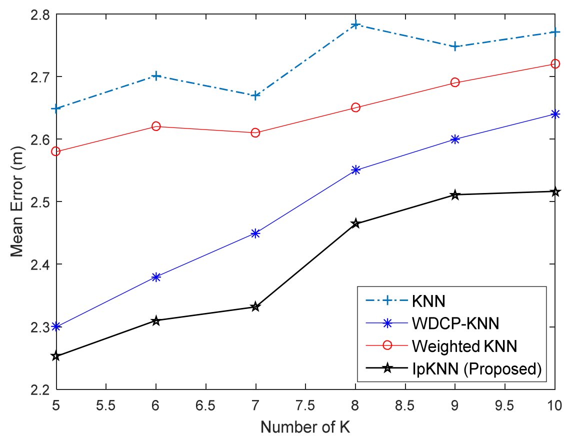

FIGURE 13. Comparisons of positioning effects of different positioning algorithms under different K values in new area.

Then we change the number of K and observe the variation of the average positioning error of the four algorithms under different conditions. The result is shown in the Figure 13.

It can be seen from the figure that the positioning errors of the four algorithms are from high to low is KNN algorithm, Weighted KNN algorithm, WDCP-KNN algorithm and IPKNN algorithm. When $K = 5$, the positioning error of the four algorithms is the lowest, and then increases the value of K. When $K = 5$, the positioning error of KNN algorithm is 2.66m; the positioning error of Weighted KNN algorithm is 2.58m; the positioning error of WDCP-KNN algorithm is 2.3m; the positioning error of IpKNN algorithm is 2.25m.

We selected two different experimental environments, and selected test points in these two environments for positioning experiments and algorithm comparison. The IpKNN algorithm first uses the improved clustering algorithm. Compared with the traditional clustering algorithm, the concept of public area is introduced, and the data of the boundary between one class and another class is regarded as the common data, which effectively solves the problem of uneven distribution of the boundary fingerprint points after clustering. The second advantage of IpKNN is the improvement of KNN algorithm. The screening function does not make a change in the nature of KNN's algorithm, but rather a reasonable screening of the K neighbors calculated by KNN algorithm. By setting the screening radius, dividing the area, all neighboring points in the area are unchanged, and neighboring points outside the area are screened out, and the remaining neighbors are used to continue the calculation after the screening ends. The experimental results show that the K-screening algorithm can improve the positioning accuracy by removing the Euclidean distance of the signal strength vector after KNN calculation, but the actual distance is too close.

V. CONCLUSION

This experiment has achieved a raise in both accuracy and efficiency through the improvement of IpKNN. The IpKNN algorithm is mainly improved in two aspects. One is to introduce a public fingerprint point on the basis of traditional clustering, which can improve the positioning accuracy when the boundary point is located; the second is to add a screening module when calculating KNN. Neighboring point which does not meet the requirements will be removed, which can effectively improve positioning accuracy.

From the experimental results, positioning accuracy of IpKNN algorithm is better than other algorithms with lower Consumption and has a 0.31 meter improvement in positioning error over the traditional KNN algorithm.

REFERENCES

[1] M. M. Dhanvijay and S. C. Patil, "Internet of Things: A survey of enabling technologies in healthcare and its applications," *Comput. Netw.*, vol. 153, pp. 113–131, Apr. 2019.

[2] J. Ruan, Y. Wang, F. T. S. Chan, X. Hu, M. Zhao, F. Zhu, B. Shi, Y. Shi, and F. Lin, "A life cycle framework of green IoT-based agriculture and its finance, operation, and management issues," *IEEE Commun. Mag.*, vol. 57, no. 3, pp. 90–96, Mar. 2019.

[3] W. Wei, X. Xia, W. Marcin, X. Fan, and R. Damaševičius, "Multi-sink distributed power control algorithm for cyber-physical-systems in coal mine tunnels," *Comput. Netw.*, to be published. doi: 10.1016/j.comnet.2019.04.017.

[4] W. Wei, M. Woźniak, R. Damaševičius, X. Fan, and Y. Li, "Algorithm research of known-plaintext attack on double random phase mask based on WSNs," *J. Internet Technol.*, vol. 20, no. 1, pp. 39–48, 2019.

[5] J. Ruan, X. Hu, X. Huo, Y. Shi, F. T. S. Chan, X. Wang, G. Manogaran, G. Mastorakis, C. X. Mavromoustakis, and X. Zhao, "An IoT-based E-business model of intelligent vegetable greenhouses and its key operations management issues," *Neural Comput. Appl.*, pp. 1–16, Mar. 2019. doi: 10.1007/s00521-019-04123-x.

[6] W. Wei, J. Su, H. Song, H. Wang, and X. Fan, "CDMA-based anti-collision algorithm for EPC global C1 Gen2 systems," *Telecommun. Syst.*, vol. 67, no. 1, pp. 63–71, 2018.

[7] I. Akbar and A. F. Misman, "Research on semantics used in GPS based mobile phone applications for blind pedestrian navigation in an outdoor environment," in *Proc. Int. Conf. Inf. Commun. Technol. Muslim World (ICT4M)*, Jul. 2018, pp. 196–201.

[8] E. Choi and S. Chang, "A consumer tracking estimator for vehicles in GPS-free environments," *IEEE Trans. Consum. Electron.*, vol. 63, no. 4, pp. 450–458, Nov. 2017.

[9] F. Yucel and E. Bulut, "Clustered crowd GPS for privacy valuing active localization," *IEEE Access*, vol. 6, pp. 23213–23221, 2018.

[10] Q. Ma, L. Gao, Y.-F. Liu, and J. Huang, "Incentivizing Wi-Fi network crowdsourcing: A contract theoretic approach," *IEEE/ACM Trans. Netw.*, vol. 26, no. 3, pp. 1035–1048, Jun. 2018.

[11] M.-C. Tsai, F. Y.-S. Lin, and Y.-F. Wen, "Lagrangian-relaxation-based self-repairing mechanism for Wi-Fi networks," *IEEE Access*, vol. 7, pp. 15868–15883, 2019.

[12] Y.-J. Lee and W.-W. Liao, "Ultimate performance of Wi-Fi access points with multiple interfaces: An application of software defined network," in *Proc. 20th Int. Conf. Adv. Commun. Technol. (ICACT)*, Feb. 2018, pp. 590–594.

[13] C.-B. Ko and J.-H. Lee, "Performance of ESPRIT and root-music for angle-of-arrival (AOA) estimation," in *Proc. IEEE World Symp. Commun. Eng. (WSCE)*, Dec. 2019, pp. 49–53.

[14] L. Zhang, T. Du, and C.-X. Wang, "Detection of an unknown radio transmitter using joint RSSD and AoA information based on factor graph," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC)*, Oct. 2017, pp. 1–5.

[15] S. K. Das and M. F. Hossain, "TDOA based localization architecture for M2M communications over cellular networks," in *Proc. 10th Int. Conf. Elect. Comput. Eng. (ICECE)*, Dec. 2018, pp. 333–336.

[16] Z. Su, G. Shao, and H. Liu, "Semidefinite programming for NLOS error mitigation in TDOA localization," *IEEE Commun. Lett.*, vol. 22, no. 7, pp. 1430–1433, Jul. 2018.

[17] M. Ivanić and I. Mezei, "Distance estimation based on RSSI improvements of orientation aware nodes," in *Proc. Zooming Innov. Consum. Technol. Conf. (ZINC)*, May 2018, pp. 140–143.

[18] A. Booranawong, N. Jindapetch, and H. Saito, "A system for detection and tracking of human movements using RSSI signals," *IEEE Sensors J.*, vol. 18, no. 6, pp. 2531–2544, Mar. 2018.

[19] M. B. Afousi and M. R. Zoghi, "Wi-Fi RSS indoor positioning system using online layer clustering and weighted DCP-KNN," in *Proc. Iranian Conf. Elect. Eng. (ICEE)*, May 2018, pp. 710–715.

[20] Z. Liu, X. Luo, and T. He, "Indoor positioning system based on the improved W-KNN algorithm," in *Proc. IEEE 2nd Adv. Inf. Technol., Electron. Automat. Control Conf. (IAEAC)*, Mar. 2017, pp. 1355–1359.

[21] L. Yen, C.-H. Yan, S. Renu, A. Belay, H.-P. Lin, and Y.-S. Ye, "A modified WKNN indoor Wi-Fi localization method with differential coordinates," in *Proc. Int. Conf. Appl. Syst. Innov. (ICASI)*, May 2017, pp. 1822–1824.

[22] L. Cheng, Y. Li, M. Zhang, and C. Wang, "A fingerprint localization method based on weighted KNN algorithm," in *Proc. 18th Int. Conf. Commun. Technol. (ICCT)*, Oct. 2018, pp. 1271–1275.

[23] C. Park and S. H. Rhee, "Indoor positioning using Wi-Fi fingerprint with signal clustering," in *Proc. Int. Conf. Inf. Commun. Technol. Converg. (ICTC)*, Oct. 2017, pp. 820–822.

[24] H. Wang, X. Zhang, Y. Gu, L. Zhang, and J. Li, "Indoor Wi-Fi RSS-fingerprint location algorithm based on sample points clustering and AP reduction," in *Proc. 6th Int. Conf. Intell. Control Inf. Process. (ICICIP)*, Nov. 2015, pp. 264–267.

[25] A. Y. Wang and L. Wang, "Research on indoor localization algorithm based on WIFI signal fingerprinting and INS," in *Proc. Int. Conf. Intell. Transp., Big Data Smart City (ICITBS)*, Jan. 2018, pp. 206–209.

[26] J. Song, J. Zhao, F. Dong, J. Zhao, Z. Qian, and Q. Zhang, "A novel regression modeling method for PMSLM structural design optimization using a distance-weighted KNN algorithm," *IEEE Trans. Ind. Appl.*, vol. 54, no. 5, pp. 4198–4206, Sep./Oct. 2018.

[27] I. Bisio, F. Lavagetto, M. Marchese, and A. Sciarrone, "Smart probabilistic fingerprinting for WiFi-based indoor positioning with mobile devices," *Pervasive Mobile Comput.*, vol. 31, pp. 107–123, Sep. 2016.

[28] Z. Wu, C. Rengin, X. Shuyan, W. Xiaosi, and F. Yuli, "Research and improvement of WiFi positioning based on k nearest neighbor method," *Comput. Eng.*, vol. 43, no. 3, pp. 289–293, 2017.

[29] L. Mengual, O. Marbán, and S. Eibe, "Clustering-based location in wireless networks," *Expert Syst. Appl.*, vol. 37, no. 9, pp. 6165–6175, 2010.

[30] M.-S. Yang, S.-J. Chang-Chien, and Y. Nataliani, "A fully-unsupervised possibilistic c-means clustering algorithm," *IEEE Access*, vol. 6, pp. 78308–78320, 2018.

[31] G.-Y. Heo, J.-S. Seo, and I.-G. Lee, "Problems in fuzzy C-means and its possible solutions," *J. Korea Soc. Comput. Inf.*, vol. 16, no. 1, pp. 39–46, 2011.

[32] Y. Sun, Y. Xu, L. Ma, and Z. Deng, "KNN-FCM hybrid algorithm for indoor location in WLAN," in *Proc. 2nd Int. Conf. Power Electron. Intell. Transp. Syst. (PEITS)*, Dec. 2009, pp. 251–254.

[33] M. E. Rida, F. Liu, Y. Jadi, A. A. A. Algawhari, and A. Askourih, "Indoor location position based on Bluetooth signal strength," in *Proc. 2nd Int. Conf. Inf. Sci. Control Eng.*, Apr. 2015, pp. 769–773.

JIN REN received the Ph.D. degree in information and communications engineering from the School of Electronic and Information Engineering, Dankook University, South Korea, in 2011. He is currently a Lecturer with the School of Information Science and Technology, North China University of Technology. His current research interests include intelligent information processing, wireless communications, and the Internet of Things (IoT) technology.

YUNAN WANG was born in Beijing, China. He is currently pursuing the master's degree with the North China University of Technology with a focus on advanced indoor positioning solutions. He is also a member of the Cyber-Physical Systems Laboratory, North China University of Technology, where he joined Dr. J. Ren's Research Team. His research interests include the low-cost Internet of Things (IoT) systems and indoor positioning algorithm.

CHANGLIU NIU was born in 1967. He received the Ph.D. degree in field and wave electromagnetics from the Beijing University of Posts and Telecommunications, in 2006. He is currently an Associate Professor with the School of Information Science and Technology, North China University of Technology. The main courses he taught include field and wave electromagnetics, the principle of modern communication, and theoretical basis of communication network. His research interests include high-speed optical communication, communication network switching, and wireless sensor networks.

WEI SONG received the Ph.D. degree in traffic information engineering and control from Beijing Jiaotong University, in 2010. He was an Assistant Professor with the Research Institute of Information Technology, Tsinghua University, from 2011 to 2012. He was a Visiting Scholar with the New Jersey Institute of Technology, from 2017 to 2018. He is currently an Associate Professor with the School of Information Engineering and the Vice Dean of the School of Information Engineering, Minzu University of China. He is also the Director of the Media Computing Laboratory. His research interests include multimedia networks, image processing, and natural language processing.