IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received September 10, 2020, accepted September 16, 2020, date of publication October 21, 2020, date of current version November 11, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3031089

KF-KNN: Low-Cost and High-Accurate FM-Based Indoor Localization Model via Fingerprint Technology

CANYANG DU[1], BAO PENG[2], ZHAOBO ZHANG[1], WEICHENG XUE[3], AND MINGXIANG GUAN[2]

[1]Guangdong Yue Hai Pearl River Delta Water Supply Company Ltd., Guangzhou 611130, China
[2]Shenzhen Institute of Information Technology, Shenzhen 518172, China
[3]Jiangsu Broadcasting Corporation, Nanjing 210009, China

Corresponding author: Bao Peng (pengb@sziit.edu.cn)

This work was supported in part by the Shenzhen Science and Technology Innovation Committee under Grant JCYJ20190809145407809, in part by the Longgang District Science and Technology Plan under Grant KJ2020D027, in part by the Guangdong Province Higher Vocational Colleges and Schools Pearl River Scholar Funded Scheme (2016), and in part by the Zhongshan Innovative Research Team Program under Grant 180809162197886.

ABSTRACT Localization and tracking of personnel and equipment are technical issues that urgently need to be solved for Indoor positioning. To improve the accuracy and environmental adaptability of personnel and equipment localization algorithms in the construction and operation of smart water platform, this paper proposes a fingerprint localization algorithm (KF-KNN) based on FM signals. Firstly, use FM data collection device to obtain RSSI fingerprint information within the coverage area, and train them to build a fingerprint database; secondly, KNN technology is used to complete the rough localization calculation based on the RSSI data received by the module to be located, the RSSI fingerprint database and environmental noise parameters; finally, the Kalman filter model is used to predict and optimize the rough position information, so as to have better environmental adaptability and effectively improve the accuracy of localization. The analysis results show that: compared with the original KNN and WKNN fingerprint localization models, the KF-KNN algorithm has better performance in localization, and its average localization error can be as low as 1.9 meters.

INDEX TERMS Fingerprint, localization, KNN, Kalman filter model.

I. INTRODUCTION

With the development of science and technology in recent years, Fingerprint Location Technology is gradually mature and popular in indoor location and other fields Its unique advantages such as high localization accuracy, strong anti-interference ability, no need to deploy any hardware facilities, no need for any knowledge about beacon position, lower cost, less influence by the surrounding environment and non-line-of-sight and no need to consider signal propagation paths and propagation models have aroused great attention in pipe system Safety Management, Indoor Personnel Localization and Tracking and other application fields, especially the fingerprint localization based on received signal strength has become the preferred localization technology in this application field [1]–[5]. There are still four problems to be solved based on the researches of fingerprint localization used in pipe system [6]–[10]:

(1) Lack of features for generating high-precision fingerprints;

(2) Lack of fingerprint suitable for band-shaped sparse distribution of wireless access points;

(3) The location description of single fingerprint is inaccurate;

(4) RSSI fingerprint is vulnerable to environmental interference and its localization accuracy is not high, which cannot meet the requirements of specific applications for high-precision localization.

Therefore, it will have important theoretical value and practical significance to apply the high-precision fingerprint positioning method to the research of pipeline system [3], [11]–[13].

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

In order to solve this problem, some projects use various sensors, robots and other auxiliary devices, which provide the location system with fingerprints updated from predefined points in the region of interest. Chen *et al.* designed a Wi-Fi-based localization system, which uses RFID-based sensors to provide a reference location for the system when users pass by. Assuming that the user's walking speed is constant, the system can regularly update the calibration data even during RFID reading [14]. Another method is to use a set of fixed signal sniffers or a mobile robot that can automatically collect Wi-Fi signal strength measurements at different locations. Despite the above limitations, fingerprint identification provides the best accuracy in complex environments (such as indoor) and works well when receiving the reflected, diffracted and scattered signals from LOS or NLOS [15]–[19]. In addition, compared with all other methods, fingerprint identification does not require any knowledge about beacon location, which makes it the only option for localization systems that utilize external beacons, such as cellular network nodes or broadcast FM stations. Based on the above reasons, this paper proposes that fingerprint identification technology is mainly adopted to achieve FM location calculation and service.

Comparison the indoor fingerprint location method based on Wi-Fi, geomagnetic and RFID network technology, because the FM fingerprint signal has better time domain robustness, it has higher practical value in the field of indoor location service [20]. For this reason, researchers have devoted more attention to this field in the last 20 years, such as in 2005, chen *et al.* use FM broadcast signal to realize coarse localizationused in six areas of Seattle, USA, location error up to 8 km, difficult to use in indoor location services [21]; V. Moghtadaiee *et al.* presents a new fingerprint location algorithm based on FM broadcast signal, this method compares the localization performance of NN, KNN and WKNN algorithms, the experimental results show that: ideally, WKNN algorithm can achieve an optimal positioning accuracy of 2.96 m at $k = 6$ [22]; S.H.Fang and Piyush Kumar at al. used different methods to study fusion localization technology of FM and GPS, experimental results show that: using 924 power spectra distributed over 200 miles, the positioning error is less than 5 miles and in a specific campus area, the ideal location error is less than 35 meters [1], [23]; Andrei Popleteev systematic study on FM indoor positioning [24], data acquisition and analysis over a long period of one year on several experimental platforms, he studied the effects of different classifiers, training methods and fingerprint widths on localization performance, experimental results show that: the long-term stability of the ambient FM localization approach even several months after training. Although these studies have promoted the development of FM indoor fingerprint location technology to a great extent, the technology is still not mature enough at present. There are still some problems to be solved in practical application scenarios such as smart water.

Facing the specific application scene of smart water, aiming at the common problems existing in indoor fingerprint localization, based on the unique advantages of FM signal, such as low frequency, low susceptibility to human and multi-path fading, strong penetration ability, small changes in signal, low cost, no need to build infrastructure, etc [25]. A precise positioning algorithm based on frequency modulation signal KF-KNN is proposed in this paper. Firstly, to acquire FM broadcast signals or FM signal data of FM local transmitters by FM data acquisition unit, minimize the location area of personnel to be located by KNN technology, and continuously optimize adjacent K points by the weighted centroid method, so as to overcome the randomness shortage of K value selection in the previous FM localization algorithm based on KNN, and improve the localization accuracy of the algorithm. Then, Kalman filter model is introduced for optimizing the strong discreteness after KNN localization, in which the prediction and correction performance of Kalman filter is brought into full play. The personnel localization and tracking method which combines the Kalman filter with RSSI fingerprint localization model can effectively reduce the influence of noise on the performance of localization model, and improve the localization accuracy and comprehensive localization performance [26]–[30].

This paper outlined our contributions as follows. In order to meet the needs of personnel and equipment localization in the construction and application of smart water, a new fingerprint localization algorithm based on FM model is proposed. After extensive simulation analysis, a practical fingerprint localization model is proposed for only using publicly available FM transmission information and floor plans of buildings to predict RSS distribution so as to achieve high-precision localization.

II. KF-KNN ALGORITHM DESCRIPTION

According to the requirement of improving high-quality location service for intelligent terminals in different indoor environments, this section proposes a new indoor localization algorithm, namely weighted centroid KNN-FM precise location algorithm (KF-KNN), based on FM broadcast signals. The algorithm can achieve the accurate localization of the intelligent terminal under the condition that the intelligent terminal can move freely indoors without any network conditions and without deploying any reference equipment or anchor nodes in advance.

Training and localization are two localization processes of KF-KNN algorithm. After selecting the environment for the experiment, first set up four data collection points (i.e. reference points) in each office, and collect RSSI, DNS, CSC and other data of each channel in FM broadcasting through special testing equipment. Assuming $O_{r,t-1}$ indicates the FM signal measurement data set of the reference point R at time t-1, $p(l_{r,t}|l_{r,(t-1)})$ indicates the probability value that the fingerprint at time t of the intelligent terminal and the fingerprint at time t-1 satisfy the relation model at the reference point R,

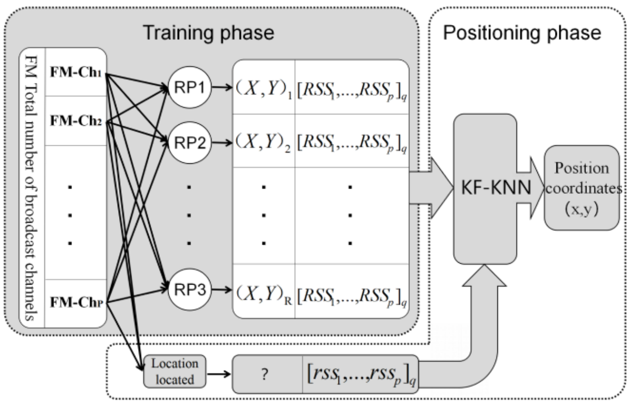

FIGURE 1. **Two Fingerprint Identification Stages of KF-KNN Algorithm.**

the reference points with the correlation degree lower than the threshold value with respect to the position K can be filtered according to the matching degree of the FM signal data set acquired by the intelligent terminal at point k at time T and the fingerprint data at time t of each reference point $p(l_{r,t}|o_{k,t}), r = 1, 2, \ldots, R$. The KNN method is used to calculate the reference points (i.e. sample points) larger than the threshold value, and the probability value is used as the weight to update the position value of the reference points at time t and the centroid of the geo.

A. TRAINING PHASE

The first step is to establish the RSSI fingerprint database of R reference points (RP) in the test environment (only RSSI data features are used in this chapter) as a training sample in the localization phase.

The database contains Q fingerprint data of all P FM channels measured by each RP in a specified time period, i.e.

$$\{RSSI_{rq} = [RSSI_{r1}, RSSI_{r2}, \ldots, RSSI_{rp}, \quad r = 1, 2, \ldots, R \quad q = 1, 2, \ldots, Q\} \tag{1}$$

$$\overline{RSSI} = \frac{1}{Q}\sum_{q=1}^{Q} RSSI_{rq} \tag{2}$$

According to equation (1), the average value of all measured values of each FM channel is calculated and recorded as the reference data of the RP in the database. The process of the fingerprint recognition training phase is described in Figure.1 [16].

B. ROUGH LOCALIZATION PHASE

In the localization stage, compare $\overline{RSSI}$ to the Q RSSI measurement data (3) obtained with the position to be positioned through a specific calculation model, adopt different matching algorithms to find the best matching point, and calculate the coordinate value of the position to be positioned.

$$\{RSSI_q = [RSSI_1, RSSI_2, \ldots, RSSI_p], \quad q = 1, 2, .., Q\} \tag{3}$$

In the stage of fingerprint localization, deterministic method and probabilistic method are mainly used to estimate the location. In this paper, three different algorithms are compared and analyzed. The first is nearest neighbor algorithm (NN), in which the position closest to RP to the unknown point is regarded as the estimated position. The nearest RP is determined by the shortest distance from the unknown point. This distance calculation is based on the Manhattan distance and Euclidean distance between the observed fingerprint and the fingerprint recorded in the database.

The second method is K nearest neighbor (KNN). The basic principle of the algorithm is that the fingerprints in the fingerprint database and the fingerprints collected online $\bar{r}$ referring to the $r_i, i = 1, 2, \ldots, L$ mean values of all FM broadcast signal samples. In the execution process, the Euclidean distance of $\bar{r}$ and each RP_i corresponding fingerprint r_i stored in the fingerprint database is solved, and after ascending order, select K RP_i corresponding positions l_i with the smallest Euclidean distance for weighting, and the weighted position information is the position estimation result of the terminal. The mathematical model of the algorithm is as follows:

$$r = \underset{r_i, i=1,2,\ldots,K}{\arg\_\min} \sum_{i=1}^{K} \|r_i - \bar{r}\|^2 \tag{4}$$

$r = [r_1, r_2, \ldots, r_k]^T$ represents the set of fingerprints, and $r = [r_1, r_2, \ldots, r_k]^T$ at the k reference points with the minimum Euclidian distance.The set of reference points for this point is $l = [l_1, l_2, \ldots, l_k]^T$, then the location estimation result of KNN algorithm is:

$$\bar{l} = \frac{1}{k}\sum_{i=1}^{K} l_i \tag{5}$$

The position estimation for the weighted KNN algorithm can be optimized as follows:

$$\bar{l} = \frac{1}{\sum_{i=1}^{K} w_i}\sum_{i=1}^{K} w_i l_i \tag{6}$$

In $w_i = \frac{1}{\varsigma + \|r_i - \bar{r}\|}$, ς is a constant approximately zero. From equation (5), the weighted KNN algorithm is a special case when the weighted factors in the KNN algorithm are DE-averaged.

Due to the short distance between sampling points of FM broadcast data, the similarity between adjacent reference points is high, which makes the general KNN algorithm often misjudge adjacent points and has a great impact on the localization accuracy, resulting in erroneous position estimation. In order to improve the reliability and robustness of localization in complex indoor environment, this chapter proposes to integrate the geometric layout of indoor environment layout into the fingerprint database establishment, and to optimize the K parameter selection and usage of localization algorithm by using the graph structure formed by RP to improve the traditional indoor localization model.

Assuming that the real coordinates of an intelligent terminal A to be located in a 2D experimental environment are (x, y). Set the RP_i coordinate as (x_i, y_i), then calculate $RSSI_{rq}$ of each reference point and $RSSI_i, i = 1, 2, \ldots, p$ received from point A according to formula (5), and select K reference points with the highest geometric spatial correlation. In this chapter, Geometric space correlation by distance d_n, the specific formula is as follows:

$$d_n = \left\|\overline{rssi} - \overline{RSSI_r}\right\|^2, \quad r = 1, 2, \ldots, R \tag{7}$$

Consider a two-dimensional plane surrounded by the selected K reference points, and the coordinates (x_{Ob}, y_{Ob}) of the O_b are:

$$x_{ob} = \frac{1}{K}\sum_{k=1}^{K} x_k, y_{ob} = \frac{1}{K}\sum_{k=1}^{K} y_k \tag{8}$$

Then the distance between the O_b and A is:

$$\begin{aligned} d_{Ob} &= \sqrt{(x_{Ob}-x)^2 + (y_{Ob}-y)^2} \\ &= \sqrt{(\frac{1}{k}\sum_{k=1}^{K} x_k - x)^2 + (\frac{1}{k}\sum_{k=1}^{K} y_k - y)^2} \\ &= = \frac{1}{K}\sqrt{[\sum_{k=1}^{K}(x_k - x)]^2 + [\sum_{k=1}^{K}(y_k - y)]^2} \\ &= = \frac{1}{K}[\sum_{k=1}^{K}(x_k - x)^2 + \sum_{k=1}^{K}(y_k - y)^2 \\ &\quad + 2\sum_{i=1}^{K-1}\sum_{j=i+1}^{K}(x_i - x)(x_j - x) \\ &\quad + 2\sum_{i=1}^{K-1}\sum_{j=i+1}^{K}(y_i - y)(y_j - y)]^{1/2} \end{aligned} \tag{9}$$

where,

$$\begin{aligned} &2(x_i - x)(x_j - x) \\ &\quad = 2x_ix_j - 2x_ix - 2x_jx + 2x^2 \\ &\quad = (x_i - x)^2 + (x_j - x)^2 - (x_i - x_j)^2 \end{aligned} \tag{10}$$

$$\begin{aligned} &2(y_i - y)(y_j - y) \\ &\quad = 2y_iy_j - 2y_iy - 2y_jy + 2y^2 \\ &\quad = (y_i - y)^2 + (y_j - y)^2 - (y_i - y_j)^2 \end{aligned} \tag{11}$$

$$\begin{aligned} d_{Ob} = \frac{1}{K}[K\sum_{k=1}^{K}(d_n)^2 - \sum_{i=1}^{K-1}\sum_{j=i+1}^{K}(x_i - x_j)^2 \\ - \sum_{i=1}^{K-1}\sum_{j=i+1}^{K}(y_i - y_j)^2]^{1/2} \end{aligned} \tag{12}$$

Therefore, given the coordinates of K reference points (x_k, y_k) and the distance d_n between them and the node A to be located, the coordinates (x_{Ob}, y_{Ob}) of the centroid O_b of the space enclosed by the selected reference points and the distance d_{Ob} between the centroid O_b and A can be obtained by equations (7) and (1-12). By further simplifying equation (12), the following results can be obtained:

$$\begin{aligned} d_{Ob}^2 = \frac{1}{K}\sum_{k=1}^{K}(d_n)^2 - \frac{1}{K^2}[\sum_{i=1}^{K-1}\sum_{j=i+1}^{K}(x_i - x_j)^2 \\ + \sum_{i=1}^{K-1}\sum_{j=i+1}^{K}(y_i - y_j^2)] = F_1 - F_2 \end{aligned}$$

$$F_1 = \frac{1}{K}\sum_{k=1}^{K}(d_n)^2, \quad F_2 = \frac{1}{K^2}\sum_{i=1}^{K-1}\sum_{j=i+1}^{K}(d_{ij})^2 \tag{13}$$

d_{ij} indicates the distance between reference points RP_i and RP_j. Assuming that the distance between the selected K reference points and the intelligent terminal A to be located satisfies the following relation:

$$0 < d_1 \leq d_2 \leq d_3 \cdots \leq d_{K-1} \leq d_K \tag{14}$$

Substitute into (13)

$$\begin{cases} F_1 = \frac{1}{K}\sum_{k=1}^{K}(d_n)^2 \leq d_K^2 \\ F_2 = \frac{1}{K^2}\sum_{i=1}^{K-1}\sum_{j=i+1}^{K}(d_{ij})^2 > 0 \end{cases} \tag{15}$$

The following results can be obtained by calculating the comprehensive equations (13)-(15):

$$d_{Ob} < d_K \tag{16}$$

That is, among the known K reference points, there is at least one reference point whose distance d_N to the intelligent terminal A to be located must be greater than the obtained distance d_{Ob} between the current centroid O_b and the intelligent terminal A to be located. Therefore, it is considered to replace the reference point farthest from the intelligent terminal A to be located with the current centroid O_b. The plane enclosed by the new K reference points must be smaller than the plane enclosed by the original K reference points, which can further reduce the range of the plane on which the intelligent terminal A to be located is located and improve the localization accuracy of nodes through multiple iterations.

C. KALMAN FILTERING PHASE

In fingerprint localization, $RSSI(k)$ indicates RSSI vector at k time, $RSSI(k-1)$ indicates RSSI vector (old fingerprint vector), $Z(k)$ indicates RSSI vector actually collected by k time sensor (new fingerprint vector). Assuming that environmental factors are not considered, The RSSI vector and the vector at time K-1 are consistent at time K, that is, the system state remains consistent at any time. At time K, the RSSI vector is consistent with the measured value, and can also be represented by A = H = I; Since there are no control variables in the system, $B = 0$, $U(k) = 0$. Therefore, Kalman filtering can be correspondingly converted into the formulas shown in (17) and (18) in fingerprint location.

$$RSSI(k) = RSSI(k-1) + W(k), W(k) \sim (0, Q) \tag{17}$$

$$Z(k) = RSSI(k) + V(k), V(k) \sim (0, R) \tag{18}$$

Therefore, after the fusion model of the old and new data in fingerprint localization is established, the fused fingerprint data $RSSI(k|k)$ is obtained according to equations (17) to (18).

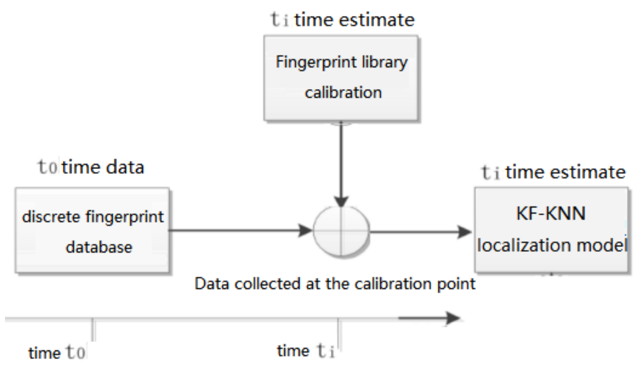

FIGURE 2. Kalman filtering process in fingerprint localization.

This paper provides a dynamic map construction method to solve the above problems. The RF signal map before construction is called static map or offline map, and the signal strength will not change with time. The construction of dynamic radio frequency map is based on a new environment. Specifically, a static radio frequency map is established at time 0t. At any recent time it, the RSSI of all reference points can be estimated by combining the RSSI collected by a small number of calibration points with the static map at time 0t. fig.2 shows the model of the dynamic map. The previous analysis indicates that the key to building a dynamic radio frequency map is how to build the RSSI function mapping relationship between several calibration points and each reference point at 0t, and then collect the RSSI of calibration points at it time to estimate the RSSI of each reference point. An important assumption here is that the functional mapping relationship between the calibration point and the reference point will not change over time.

Define an area to build a map (calculate the map), m indicates the number of channels, M indicates the number of calibration points, and n indicates how many reference points there are. At O_t time, the mapping relationship between the reference point and the RSSI value of the calibration point can be expressed by Equation 4-12.

$$R_{nk}(t_0) = f_{nk}(S_{1k}(t_0), S_{2k}(t_0), \ldots, S_{Mk}(t_0)) \tag{19}$$

S_{MK} is the RSSI value of the k(1<k<m) the channel collected at the m-th calibration point. R_{nk} is defined as the RSSI value of the m-th channel collected by the n-th reference point. Assuming that the fingerprint database is constructed at O_t time, the function f_{nk} can be solved. The functional relation at t_i time can still be expressed as:

$$R_{nk}(t_i) = f_{nk}(S_{1k}(t_i), S_{2k}(t_i), \ldots, S_{MK}(t_i)) \tag{20}$$

During the localization calculation, a position estimation (observation position) can be obtained by some localization technology (such as position fingerprint method), according to past experience and law, to determine the current position (predicted position)(moving target is usually moving with uniform speed), can be based on the position and speed of the previous moment The observation result and the prediction result are weighted average as the localization result. The weight value depends on the uncertainty degree of the observation position and the prediction position, and is weighted optimal according to the Kalman method. The specific operation steps of the algorithm proposed in this chapter are as follows:

1) Introducing external input to predict the current state from the state at the previous moment.

$$\tilde{x}_k = A\tilde{x}_{k-1} + Bu_{k-1} \tag{21}$$

2) Adding new uncertainties in the prediction process, plus the uncertainties that existed before.

$$\tilde{P}_k = AP_{k-1}A^T + Q \tag{22}$$

3) Calculate the kalman gain (weight) from the uncertainty Pk− of the prediction result and the uncertainty RR of the observation result.

$$K_k = \tilde{P}_k H^T (H\tilde{P}_K H^T + R)^{-1} \tag{23}$$

4) Performing weighted average on the prediction result and the observation result to obtain the state estimation at the current time.

$$\hat{x}_k = \tilde{x}_k + K_k(z_k - H\tilde{x}_k) \tag{24}$$

5) Update Pk to represent the uncertainty of this state estimation.

$$P_k = \tilde{P}_k - K_k H\tilde{P}_k \tag{25}$$

It should be noted that in the localization, the state x_k is a vector, and besides coordinates, it can also contain velocity, such as x_k = (coordinate x, coordinate y, velocity x, velocity y). The state is a vector rather than just a scalar. The matrix multiplication in the above formulas actually calculates multiple States at the same time, indicating that the variance of uncertainty, i.e. the covariance matrix.

III. EXPERIMENTAL SETUP AND RESULTS

A. SETTINGS OF SIMULATION ENVIRONMENT

The experimental test site is located on the fourth floor of the office building. Figure 3 shows the layout of the test environment. The testing area is 88m x 88m, including 7 rooms (typical indoor offices) and corridors, and the sign (□) refers to test point (TP) or reference point (RP). There are 9 RPs and 3 TPs in this experiment. RP is a strong nearby FM broadcast signal searched by FM test equipment. TP is defined as the point that user most likely needs localization. Each RP senses 17 FM channels (P = 17), covering the entire FM bandwidth from 88MHz to 108MHz. At each RP, the user would first face north and record the RSS of the sensed FM channel. Then user would change the direction to south and again record the RSS value. The reason for above is that the antenna is wide and suitable for narrow corridors in the east-west direction. A total of 120 measurements are made at each point (Q = 120). Since our FM-based localization is a two-dimensional method, the height of the FM antenna remains constant in all measurements (height = 75cm). The data was

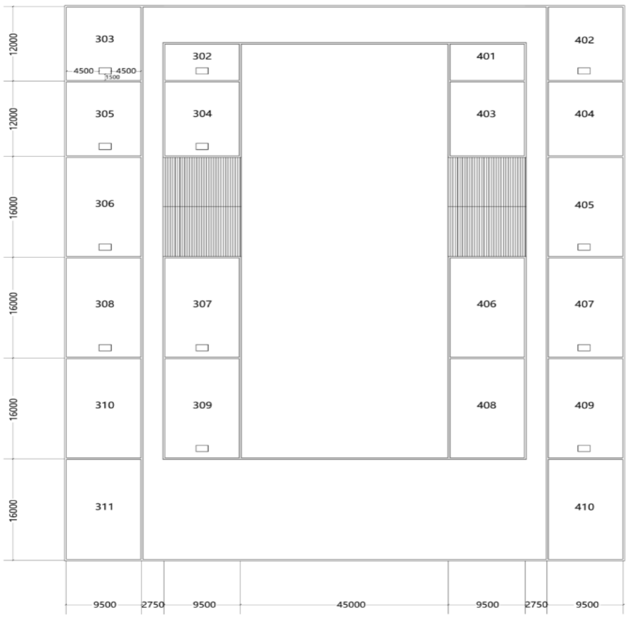

FIGURE 3. Arrangement of wireless FM data acquisition points.

collected three times a day and measured 120 times each time within a two-week period, thus effectively avoiding the presence of people.

B. ALGORITHM PERFORMANCE ANALYSIS

1) LOCALIZATION ACCURACY

Figure.4 shows the statistics of the FM signal. From the results, the following conclusions can be drawn: the positioning characteristics of FM signal is good. Considering the complexity of signal propagation in indoor environment, this chapter would first analyze the FM statistical characteristics of different locations at the same time and the characteristics of the same location at different times before analyzing the localization accuracy. In this way, we can have a better understanding of which factors affect the performance and accuracy of the positioning algorithm.

According to the characteristics of the indoor environment, the relationship between RSSI and location should be one-to-one, which is also the key to realizing indoor localization. Theoretically, the RSSI values received at different data collection locations will show differences. At the same time, the signal strength received from different radio stations at the same location will also be different. Figure.4 intuitively embody this one-to-one matching relationship. Figure 4 shows a variety of corresponding mean RSSI distribution characteristics measured at different physical locations. The distribution of mean RSSI from multiple FM radio stations at different data collection locations can be seen from the figure; at the same time, as this chapter has selected multiple radio stations for comparison, the mean RSSI of different radio stations measured at the same location also differs as shown in Figure.4.

Figure.4 shows the probability distribution histogram of RSSI received from one radio station at the same data collection location. Data in the figure was collected every half an hour using MS9801 small solar field strength meter for 7 days, with a total of 30,000. The test took place on the third and fourth floors of the office building in November 2019. It can be seen from the figure that the richer the signal sample data set, the better and closer is the sample's RSSI distribution to the true one. Theoretically, the larger the number of samples, the more it can truly reflect the relationship between RSSI and location, and the higher the localization accuracy will be. However, it will take a lot of time and labor to collect signal samples, especially for the fingerprint localization method. In addition, the larger the sample size, the more complicated the calculation and the more difficult the maintenance of fingerprint database would be.

The FM fingerprint was generated by the MS9801 small solar field strength meter after collecting FM radio signals. The RSSI value was collected at 12 data collection points. Only 37 data collection points were set in 12 rooms of 12×9.5 m because of the original furnishings of office, etc. Each FM fingerprint includes 13 channels, and each channel collects 10 RSSI samples at a time. In order to adopt the weighted centroid KNN algorithm to realize indoor FM localization, this chapter uses a K-nearest neighbor classifier ($k = 30$). Randomly select a path consisting of 100 points as the true trajectory (shown by the green line in the figure), and take the location estimated by the algorithm in this chapter as the predicted trajectory (shown by the red line in the figure). Comparing the relationship between the predicted trajectory

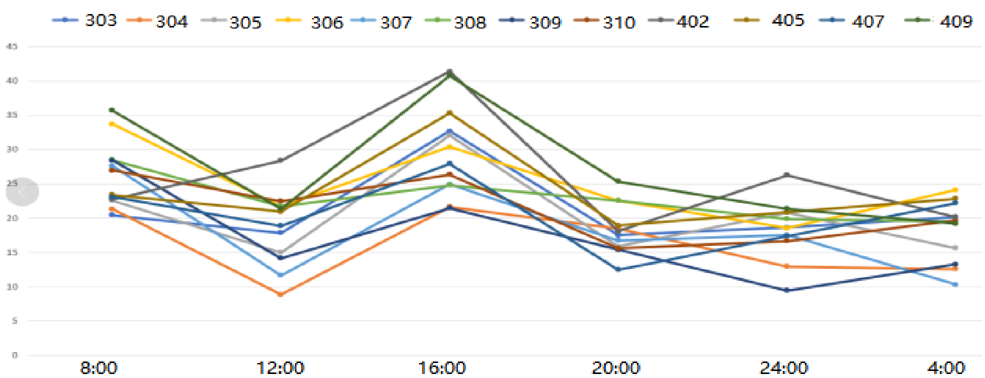

FIGURE 4. Typical distribution of RSS of an FM channel measured at a location.

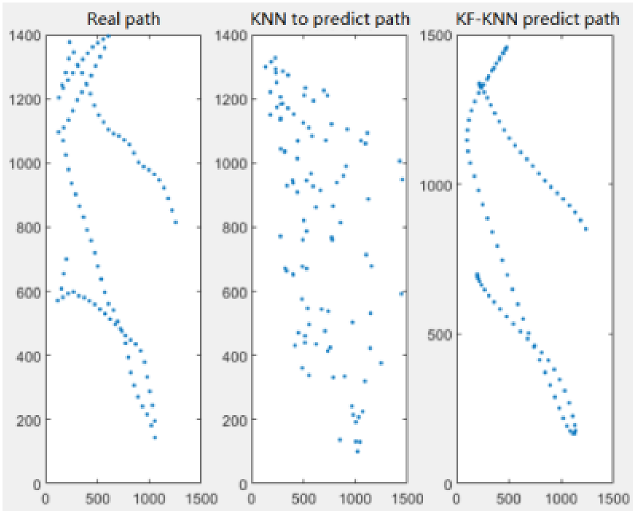

FIGURE 5. Influence of ranging error on localization error.

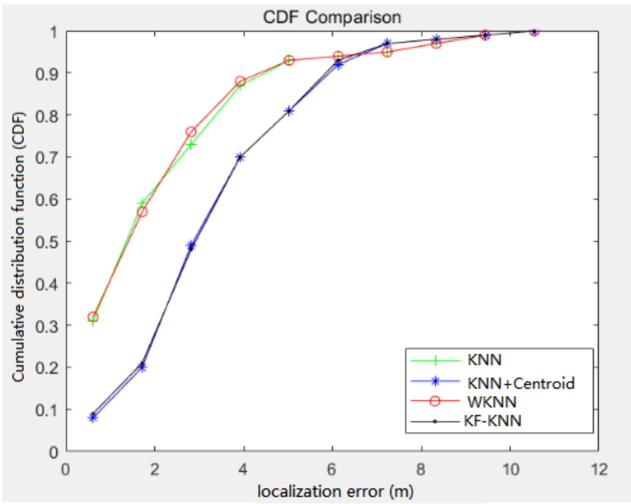

FIGURE 6. Performance of different localization algorithms.

and the true trajectory, the result shows that the algorithm can better serve indoor localization.

In the performance evaluation of dynamic localization technology, firstly, Figure 5 shows the improvement of corresponding performance after comparing KF-KNN algorithm with KNN algorithm. Curve in Figure.5a represents the true trajectory of the intelligent terminal for testing, while Figure. 5b represents the estimated trajectory using the KNN algorithm, and Figure. 5c represents the estimated trajectory using the KF-KNN algorithm. It is known from Figure.5 that the trajectory estimated by the KF-KNN algorithm is closer to the actual route of the terminal's movement, that is, the localization performance of the KF-KNN algorithm is significantly better than the KNN algorithm.

As shown in Figure.6, all ranging errors are combined into a cumulative distribution function (CDF), to analyze the performance of four algorithms, KNN, KNN + Kalman filtering, WKNN and KF-KNN. It can be seen from Figure.8 that when the localization error is low, the covariance of the Kalman filter is also small, so when the localization error is less than 5 meters, the algorithms without Kalman filtering have better localization accuracy, and when the localization error is greater than 6 meters, on the contrary, algorithm using Kalman filtering method has better performance.

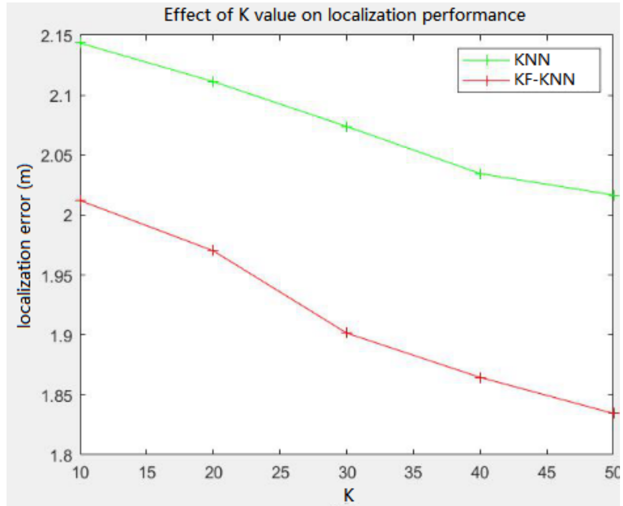

FIGURE 7. Influence of different K values on localization error.

2) THE EFFECT OF K VALUE ON LOCALIZATION ERROR

Figure 7 shows the influence of changing K value on positioning accuracy. K value being 1 is equivalent to the NN method. As the value of K increases, the algorithm's localization error continues to decrease and reached the best average error when K = 40, and then fall into a relatively stable convergence stage. Under the experimental conditions in this chapter, the overall localization performance is the best when K = 40, and the best estimation of location can be obtained with an average error on distance of 1.85 meters. On the other hand, using more nearest neighbor reference points is equivalent to that points far from the target would also have influence on classification of unknown objects. The advantage of this case is that it is robust, yet the shortcomings are also obvious that it will cause under-fitting, that is, the unclassified objects are not really labeled and classified; while using less neighbor points may miss some useful information, which requires that the unclassified objects shall be very close to its neighbors. This might raise a problem if the neighbor is a noise point, then the classification of the unclassified object will produce errors, then the KF-KNN algorithm will be overfitting.

3) INFLUENCE OF PREDICTED NOISE COVARIANCE MATRIX Q ON LOCALIZATION ERROR

The size of the predicted noise covariance matrix Q depends on the degree of trust in the prediction process. It can be seen from Figure.8, with the continuous increase of Q value, the algorithm will converge rapidly. The abscissa in the figure shows different Q value while the ordinate represents the localization error of the algorithm. As can be seen from Figure.10, initialization and stabilization comprise the entire positioning process. In the initialization stage, as the confidence in the prediction process increases, the localization

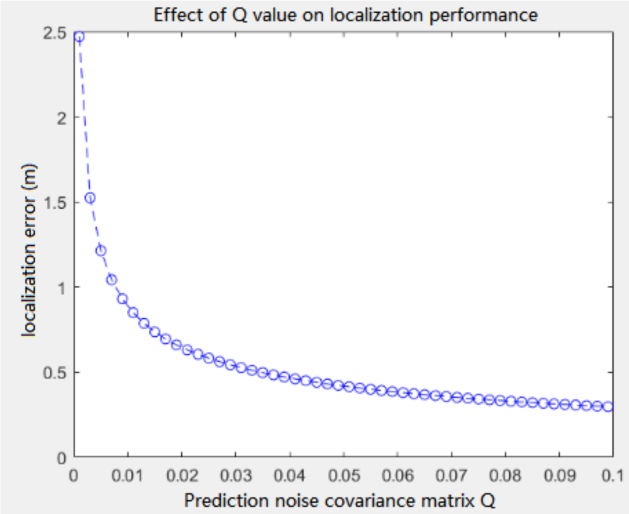

FIGURE 8. Influence of different Q values on localization error.

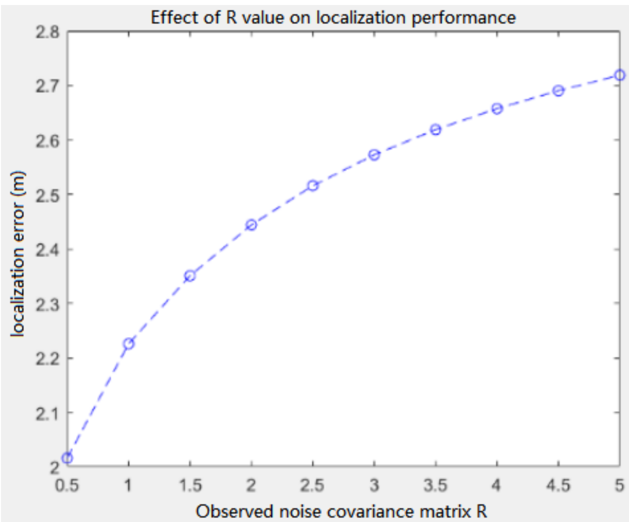

FIGURE 9. Influence of different R values on localization error.

error of the KF-KNN algorithm is improved very quickly. In the stable stage, when the degree of trust in the prediction process reaches a certain value, the impact on the localization error reaches balance to some degree, and the localization error tends to stabilize at this moment.

4) INFLUENCE OF OBSERVED NOISE COVARIANCE R ON LOCALIZATION ERROR

The size of the observed noise covariance matrix R depends on the degree of trust in the observation process. It can be seen from Figure.9 that as the R value increases, the localization error of the algorithm gradually increases as well. The abscissa in the figure shows different R value while the ordinate represents the localization error of the algorithm. It can be seen from Figure.9 that the R value shows an approximately linear relation to the algorithm's localization error. The R value is an important adjustable index when optimizing and adjusting the performance of the KF-KNN algorithm.

IV. CONCLUSION

This paper presents an FM - indoor location algorithm based on weighted centroid KNN technology. It overcomes the shortcoming of randomness for K value selection in the previous KNN-based FM localization algorithm and improves accuracy as the KNN technology can narrow the location scope of the target intelligent terminal, and weighted centroid method is combined to continuously optimize the neighboring K points. At the same time, this algorithm can meet the needs of various application scenarios by using different motion models. Simulation results show that, compared with the traditional KNN model, the localization accuracy of the KF-KNN model is significantly improved. Furthermore, the comprehensive performance of KNN, WKNN and KF-KNN models is compared systematically.It can be seen that the comprehensive performance of KF-KNN model is much higher than that of the other two models. In addition, KF-KNN model has low algorithm difficulty, which allows it to be well applied to the precise localization of personnel and equipment in construction and operation of pipe system theoretically. The above is only a preliminary study, and there are still many problems to be further discussed and improved in the practical life and project application of this design:

1) This FM localization algorithm has not been compared with other indoor localization methods such as WI-FI, Bluetooth, RFID, UWB, GSM, etc., as well as the in-depth performance analysis of these algorithm and their combinations for the application scenario of pipe system construction and operation in this project;
2) Establishment of multi-dimensional wide fingerprint localization model. For example, fingerprint localization using WI-FI + FM, RSSI + CSI, video analysis + fingerprint localization and other combinations still need in-depth study;
3) The fingerprint database needs to be constantly updated in the fingerprint localization model, resulting in a large amount of work in collecting dynamic data. In order to make a breakthrough in fingerprint localization, it is necessary to continuously improve and optimize the self-adaptive capacity of the fingerprint database, and to establish a dynamic mapping model of the fingerprint database with better performance, which will directly determine whether the fingerprint localization technology can be widely applied on a large scale.

REFERENCES

[1] S.-H. Fang, J.-C. Chen, H.-R. Huang, and T.-N. Lin, "Is FM a RF-based positioning solution in a metropolitan-scale environment? A probabilistic approach with radio measurements analysis," *IEEE Trans. Broadcast.*, vol. 55, no. 3, pp. 577–588, Sep. 2009.
[2] H. Hashemi. "The indoor radio propagation channel," *Proc. IEEE*, vol. 81, no. 7, pp. 943–968, Jul. 1993.
[3] R. Harle, "A survey of indoor inertial positioning systems for pedestrians," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 3, pp. 1281–1293, 3rd Quart., 2013.

[4] Y. Chen, D. Lymberopoulos, J. Liu, and B. Priyantha, "FM-based indoor localization," in *Proc. 10th Int. Conf. Mobile Syst., Appl., Services (MobiSys)*, 2012, pp. 1–5.

[5] X. Liu, M. Jia, X. Zhang, and W. Lu, "A novel multichannel Internet of Things based on dynamic spectrum sharing in 5G communication," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 5962–5970, Aug. 2019.

[6] Z. Na, J. Lv, F. Jiang, M. Xiong, and N. Zhao, "Joint subcarrier and subsymbol allocation-based simultaneous wireless information and power transfer for multiuser GFDM in IoT," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 5999–6006, Aug. 2019.

[7] X. Guo, N. Ansari, F. Hu, Y. Shao, N. R. Elikplim, and L. Li, "A survey on fusion-based indoor positioning," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 1, pp. 566–594, 1st Quart., 2020.

[8] A. Popleteev, "Indoor positioning using FM radio signals," Ph.D. dissertation, Univ. Trento, Trento, Italy, 2011, pp. 8–125.

[9] A. Popleteev, "AmbiLoc: A year-long dataset of FM, TV and GSM fingerprints for ambient indoor localization," in *Proc. IPIN*, 2017, pp. 1–4.

[10] D. Lymberopoulos, J. Liu, X. Yang, R. R. Choudhury, V. Handziski, and S. Sen, "A realistic evaluation and comparison of indoor location technologies: Experiences and lessons learned," in *Proc. 14th Int. Conf. Inf. Process. Sensor Netw. (IPSN)*, 2015, pp. 178–189.

[11] X. Liu and X. Y. Zhang, "NOMA-based resource allocation for cluster-based cognitive industrial Internet of Things," *IEEE Trans. Ind. Informat.*, vol. 16, no. 8, pp. 5379–5388, Aug. 2020.

[12] C. Chen, Y. Chen, Y. Han, H. Q. Lai, and K. J. R. Liu, "Achieving centimeter-accuracy indoor localization on WiFi platforms: A frequency hopping approach," *IEEE Internet Things J.*, vol. 4, no. 1, pp. 111–121, Feb. 2017.

[13] Z. Na, J. Wang, C. Liu, M. Guan, and Z. Gao, "Join trajectory optimization and communication design for UAV-enabled OFDM networks," *Ad Hoc Netw.*, vol. 98, Mar. 2020, Art. no. 102031.

[14] X. Ye, X. Yin, X. Cai, A. P. Yuste, and H. Xu, "Neural-network-assisted UE localization using radio-channel fingerprints in LTE networks," *IEEE Access*, vol. 5, pp. 12071–12087, 2017.

[15] L. Mo and C. Li, "Passive UHF-RFID localization based on the similarity measurement of virtual reference tags," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 8, pp. 2926–2933, Aug. 2019.

[16] X. Tian, S. Zhu, S. Xiong, B. Jiang, Y. Yang, and X. Wang, "Performance analysis of Wi-Fi indoor localization with channel state information," *IEEE Trans. Mobile Comput.*, vol. 18, no. 8, pp. 1870–1884, Aug. 2019.

[17] J. Luo, Z. Zhang, C. Wang, C. Liu, and D. Xiao, "Indoor multifloor localization method based on WiFi fingerprints and LDA," *IEEE Trans. Ind. Informat.*, vol. 15, no. 9, pp. 5225–5234, Sep. 2019.

[18] S.-C. Yeh, W.-H. Hsu, W.-Y. Lin, and Y.-F. Wu, "Study on an indoor positioning system using Earth's magnetic field," *IEEE Trans. Instrum. Meas.*, vol. 69 no. 3, pp. 865–872, Mar. 2020.

[19] S. Shao, A. Khreishah, and I. Khalil, "Enabling real-time indoor tracking of IoT devices through visible light retroreflection," *IEEE Trans. Mobile Comput.*, vol. 19, no. 4, pp. 836–851, Apr. 2020.

[20] A. Matic, A. Papliatseyeu, V. Osmani, and O. Mayora-Ibarra, "Tuning to your position: FM radio based indoor localization with spontaneous recalibration," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. (PerCom)*, Mannheim, Germany, Mar. 2010, pp. 153–161.

[21] A. Youssef, J. Krumm, E. Miller, G. Cermak, and E. Horvitz, "Computing location from ambient FM radio signals [commercial radio station signals]," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Las Vegas, NV, USA, Mar. 2005, pp. 824–829.

[22] V. Moghtadaiee, A. G. Dempster, and S. Lim, "Indoor localization using FM radio signals: A fingerprinting approach," in *Proc. Int. Conf. Indoor Positioning Indoor Navigat.*, Guimaraes, Portugal, Sep. 2011, pp. 1–7.

[23] P. Kumar, T. Mukherjee, E. Pasiliao, and L. Xu, "Cheap approximate localization using FM radio," in *Proc. 23rd SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst. (GIS)*, Seattle, WA, USA, 2015, pp. 1–4.

[24] A. Popleteev, "Improving ambient FM indoor localization using multipath-induced amplitude modulation effect: A year-long experiment," *Pervasive Mobile Comput.*, vol. 58, Aug. 2019, Art. no. 101022.

[25] M. S. Corrington, "Frequency-modulation distortion caused by multipath transmission," *Proc. IRE*, vol. 33, no. 12, pp. 878–891, Dec. 1945.

[26] J. Koster, "Multipath transmission effects in FM reception and their simulation in the laboratory," *Philips Tech. Rev.*, vol. 22, no. 12, pp. 393–402, 1960.

[27] W. Cheng, X. Zhang, and H. Zhang, "Statistical-QoS driven energy-efficiency optimization over green 5G mobile wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 12, pp. 3092–3107, Dec. 2016.

[28] Z. Li, Q. Q. Pei, and Y. Liu, "Spoofing attacks and countermeasures in FM indoor localization system," *World Wide Web*, vol. 22, pp. 2407–2436, 2019.

[29] L. Cong, J. Tian, and H. Qin, "A practical floor localization algorithm based on multifeature motion mode recognition utilizing FM radio signals and inertial sensors," *IEEE Sensors J.*, vol. 20, no. 15, pp. 8806–8819, Aug. 2020.

[30] T. Mukherjee, P. Kumar, D. Pati, E. Blasch, E. Pasiliao, and L. Xu, "LoSI: Large scale location inference through FM signal integration and estimation," *Big Data Mining Anal.*, vol. 2, no. 4, pp. 319–348, Dec. 2019.

CANYANG DU received the M.Eng. degree in electrical engineer from the Huazhong University of Science and Technology.

He is a Senior Engineer (Vice Senior), a registered Safety Engineer, and the President of Guangdong Yue Hai Pearl River Delta Water Supply Company Ltd. His research interests are in the areas of smart water affairs, smart water conservancy, IT project management, the Internet of Everything (IoE), the Internet of Things (IoT), and indoor location technique.

BAO PENG received the Ph.D. degree in information and communication engineering from the Institute of Electronic Information Engineering, Harbin Institute of Technology, Harbin, China, in 2009.

His research interests are in the areas of distributed systems and information management.

ZHAOBO ZHANG received the master's degree in management science and engineering from Hohai University, Nanjing.

He is a Senior Engineer of computer technology and software, a System Analyst, and an Information System Project Manager. His research interests are in the areas of smart water affairs, smart water conservancy, IT project management, the Internet of Things (IoT), and indoor location technique.

WEICHENG XUE is currently pursuing the Ph.D. degree with the Computer Science and Engineering Department, University of the Chinese Academy of Sciences.

He is also a Senior Engineer with Jiangsu Broadcasting Corporation, China. His research interests include indoor localization and machine learning.

MINGXIANG GUAN was born in Gong'an, Hubei, China, in 1979. He received the bachelor's degree in electronic engineering from HUST, China, in 2002, and the master's and Ph.D. degrees in communication engineering from HIT, China, in 2004 and 2008, respectively.

His main interests are in the areas of wireless communication, HAPS communication, and network communication research.