AP Optimization for Wi-Fi Indoor Positioning-Based on RSS Feature Fuzzy Mapping and Clustering

XIAOLONG YANG, (Member, IEEE), ZHU LIU, (Member, IEEE), WEI NIE, (Graduate Student Member, IEEE), WEI HE, and QIAOLIN PU

School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

Corresponding author: Zhu Liu (lz1532098657@foxmail.com)

This work was supported in part by the Science and Technology Research Project of Chongqing Education Commission under Grant KJQN201800625, in part by the Chongqing Natural Science Foundation Project under Grant cstc2019jcyj-msxmX0635, in part by the Chongqing Natural Science Foundation Project under Grant cstc2019jcyj-msxmX0742, and in part by the National Natural Science Foundation of China under Grant 61301126 and Grant 61471177.

ABSTRACT In indoor environments, Access Points (APs) are widely deployed in various locations of buildings, and thereby the AP optimization-based Wi-Fi indoor positioning technology is of great significance for achieving the satisfactory indoor Location-based Services (LBSs). However, the current Wi-Fi indoor positioning methods rarely pay attention to the diversity of Received Signal Strength (RSS) features for AP optimization, which may result in the low positioning accuracy and high positioning overhead. In order to deal with such issues, this article proposes a new concept of multi-dimensional RSS feature fuzzy mapping and clustering for AP optimization in Wi-Fi indoor positioning. Besides, the extensive experiments conducted in an actual indoor environment show that compared with the existing positioning methods, the proposed method can not only achieve higher positioning accuracy by using the optimized APs but also reduce the positioning overhead in the online phase.

INDEX TERMS Wi-Fi indoor positioning, AP optimization, multi-dimensional RSS feature, fuzzy mapping and clustering

I. INTRODUCTION

In recent years, with people’s longing for a convenient and technologically equipped smart life and the development of communication technology, Location-based Services (LBSs) [1] have been given a different meaning from the conventional indoor positioning, such as the indoor personnel collaborative navigation guidance, intelligent warehouse location management, and merchant advertising. In such a massively interconnected smart-device environment [2], [3], how to ensure the reliability and convenience of LBSs with high data rate and spectral efficiency has become the main focus of many relevant researchers at home and abroad. Besides, in the actual indoor environment, due to the obstruction of the furniture, walls, and many other obstacles including the human body shadowing, indoor wireless signals are severely affected by the signal strength attenuation and multi-path fading during the propagation, which limits the development of indoor positioning technologies. At present, there are a large number of common indoor positioning technologies such as Ultra-wide Band (UWB) [4]–[6], Infrared Ray (IR) [7], Radio Frequency Identification (RFID) [8]–[10], Bluetooth [11], Ultrasonic Wave (UW) [12], ZigBee [13], Visible Light (VL) [14], and Wi-Fi [15] indoor positioning technologies. Most of them involve the high deployment cost due to the additional hardware requirement, which makes them difficult to be widely applied. In comparison, the Wi-Fi indoor positioning technology can be selected as a good candidate in indoor scenarios due to its advantages of easy deployment, low cost, and extensive coverage range [16].

In the era of ubiquitous computing [17], technologies such as the terminal communication, data collection, and information processing have achieved sufficient development, and the Wi-Fi indoor positioning technology in indoor environment has been highly valued and deeply studied as a
hot topic. Without incurring any additional hardware, Wi-Fi indoor positioning methods generally fall into two categories, namely propagation model-based method and location fingerprinting-based method. The former is based on the signal propagation model, which utilizes the relationship between RSS features and the corresponding propagation distances. In this method, the target position is generally estimated by using the geometric positioning algorithm, such as the Approximate Point-in-triangulation Test (APIT) [18], with the known locations of Wi-Fi Access Points (APs). However, since the Non-line-of-sight (NLOS) scenario often exists during the signal propagation in the indoor environment [19], how to construct an appropriate signal propagation model in the target environment to precisely describe the diversity of Received Signal Strength (RSS) features is challenging [20]. Comparatively speaking, independent of the signal propagation model construction, the latter involves two phases, namely offline phase and online phase. In the offline phase, we collect the RSS readings at several pre-calibrated points in the target environment to construct the fingerprint database, and these pre-calibrated points are called as Reference Points (RPs). Then, in the online phase, the proximity matching algorithm, such as the K-nearest Neighbor (KNN) [21], is used to estimate the target position by matching the newly-collected RSS features against the pre-constructed location fingerprint database.

The structure of this article is organized as follows. An overview of related works on the existing Wi-Fi indoor positioning technologies is given in Section 2. Then, the detailed steps of the proposed method in offline and online phases are illustrated in Section 3 and 4 respectively. In Section 5, the effectiveness of the proposed method in terms of the positioning accuracy and overhead is verified through the extensive experiments conducted in an actual indoor environment. Finally, Section 6 summarizes this article and provides the future direction.

II. RELATED WORK

In recent years, many researchers at home and abroad have shown great interest in developing the Wi-Fi indoor positioning technology to provide a more satisfactory LBS. For example, the authors in [22] use the T-test method to optimize the RSS sample size for Wi-Fi positioning, where the Operating Characteristics (OC) function is applied to depict the relationship between the positioning accuracy and RSS sample size. The authors in [23] use the nonlinear Partial Least Square (PLS) method to extract RSS features to eliminate the correlation between different variables by reducing variable dimensions. The two methods above can reduce the overhead of location fingerprint database construction, but they do not take into account the difference of the location resolution of RSS features from different APs and the existence of redundant APs, resulting in the decrease of positioning accuracy. In response to this compelling problem, considering the diversity of RSS features and the randomness of the noise in the actual indoor environment, the authors in [24] propose a new collaborative positioning method, namely the Maxlifd. In this method, the joint maximum likelihood estimation method is used to integrate the prior information of location fingerprints and the corresponding mutual distance between them, based on which the Semi-definite Program (SDP) algorithm is applied to estimate the target position. This method depends on a single RSS feature for positioning, and the corresponding positioning overhead in the online phase is relatively high, thus limiting its practical applications.

Different from the approaches above, a novel AP optimization method for Wi-Fi indoor positioning based on multi-dimensional RSS feature fuzzy mapping and clustering is proposed in this article, as shown in Fig. 1. Specifically, in the offline phase, we perform the offline RSS data collection and pre-processing, and then calculate the Maximum Information Coefficient (MIC) between different APs as the corresponding AP correlation, based on which the fuzzy clustering algorithm is used to select the non-redundant APs for positioning. In addition, we also utilize the AP information gain ratio to calculate the RSS feature fuzzy weight, which plays an essential role in calculating the online AP fuzzy membership. Then, in the online phase, by selecting APs with the high location resolution (or called the high online AP fuzzy membership) for positioning, the target position is estimated based on the Bayesian positioning approach.

III. OFFLINE PHASE

A. AP REDUNDANCY REDUCTION

In the offline phase, the first task is to select the non-redundant APs for positioning by eliminating the APs with the high substitutability. To this end, we do the offline RSS data collection and pre-processing, calculate the MIC between different APs as the corresponding AP correlation, and then perform fuzzy clustering to select the non-redundant APs for positioning. The specific process of achieving the AP redundancy reduction is given below.

1) RSS DATA PRE-PROCESSING

By supposing that there are $n$ APs and $m$ RPs in the environment, we set $\text{rss}_{ij}^s(i = 1, \cdots, n; j = 1, \cdots, m; s = 1, \cdots, \chi)$ as the $s$-th offline RSS feature, such as the RSS mean, RSS variance, RSS maximum, RSS minimum, the difference between the RSS maximum and minimum, RSS median, RSS...
with the highest probability of occurrence, and the probability of exceeding the RSS mean [25], from the \( i \)-th AP at the \( j \)-th RP. Specifically, RSS mean is the average value of each group of RSS data; RSS mean is the variance of each group of RSS data; RSS maximum is the maximum value of each group of RSS data; RSS maximum is the minimum value of each group of RSS data; The difference between the RSS maximum and minimum is the difference between the maximum and minimum values of each set of RSS data; We arrange the RSS data in order of size, and the RSS in the middle of this RSS data is the RSS median of this set of data; RSS with the highest probability of occurrence is the RSS with the highest occurrence frequency in each set of RSS data and the probability of exceeding the RSS mean is the proportion of each group of RSS data that is higher than the RSS mean in all RSS data. Based on them, we can construct the offline RSS feature matrix as

\[
Z^{\text{off}} = \begin{bmatrix}
\mathcal{Z}_1^{\text{off}}, & \cdots, & \mathcal{Z}_c^{\text{off}}
\end{bmatrix}^T = \begin{bmatrix}
\mathcal{Z}^{\text{off}}, & \cdots, & \mathcal{Z}^{\text{off}}, & \cdots, & \mathcal{Z}^{\text{off}}
\end{bmatrix}
\]

(1)

where the notation “\( T \)” represents the transpose operation, \( \mathcal{Z}_s^{\text{off}} = (\mathcal{Z}_{s,1}^{\text{off}}, \cdots, \mathcal{Z}_{s,n}^{\text{off}}) \) \( (s = 1, \cdots, \gamma) \), and \( \mathcal{Z}_{s,i}^{\text{off}} \) \( (i = 1, \cdots, n) \) represents the \( s \)-th offline RSS feature from the \( i \)-th AP, which is given by \( \mathcal{Z}_{s,i}^{\text{off}} = \mathcal{Z}_{s,i} - \mathcal{Z}_{s} \mathbf{1} \).

Considering the time-varying characteristics of the Wi-Fi indoor environment caused by the serious multi-path effect and unstable environment noise, we perform the standardization of \( \mathcal{Z}_{s,i}^{\text{off}} \), as \( \tilde{\mathcal{Z}}_s = \frac{\mathcal{Z}_s^{\text{off}} - \mathcal{Z}_s}{\mathcal{Z}_s} \), where \( \mathcal{Z}_s = \frac{1}{n} \sum_{i=1}^{n} \mathcal{Z}_{s,i}^{\text{off}} \) and \( \mathcal{Z}_{s,i}^{\text{off}} \) are respectively the mean and standard deviation of the \( s \)-th offline RSS feature. We continue to take the difference of scales of different RSS features into consideration by performing the normalization of \( \mathcal{Z}_{s,i}^{\text{off}} \), as \( \eta_s = \mathcal{Z}_{s,i}^{\text{off}} - \min_{\mathcal{Z}_{s,i}^{\text{off}}} \mathcal{Z}_{s,i}^{\text{off}} \), and then take the logarithm of \( \mathcal{Z}_{s,i}^{\text{off}} \) (i.e., \( \mathcal{Z}_{s,i}^{\text{off}} = \log \mathcal{Z}_{s,i}^{\text{off}} \)) to reduce the order of magnitude of \( \mathcal{Z}_{s,i}^{\text{off}} \). After pre-processing RSS data, we construct the standardized offline RSS feature matrix as

\[
\tilde{Z} = \begin{bmatrix}
\mathcal{Z}_{11}^{\text{off}}, & \cdots, & \mathcal{Z}_{1n}^{\text{off}} \\
\mathcal{Z}_{21}^{\text{off}}, & \cdots, & \mathcal{Z}_{2n}^{\text{off}} \\
\mathcal{Z}_{m1}^{\text{off}}, & \cdots, & \mathcal{Z}_{mn}^{\text{off}} \\
\mathcal{Z}_{1}^{\text{off}}, & \cdots, & \mathcal{Z}_{n}^{\text{off}} \\
\mathcal{Z}_{2}^{\text{off}}, & \cdots, & \mathcal{Z}_{n}^{\text{off}} \\
\mathcal{Z}_{m}^{\text{off}}, & \cdots, & \mathcal{Z}_{n}^{\text{off}}
\end{bmatrix}
\]

(2)

where \( \mathcal{Z}_{s,i}^{\text{off}} \) represents the \( s \)-th standardized offline RSS feature from the \( s \)-th AP.

2) AP CORRELATION CALCULATION

For any two APs (e.g., the \( p \)-th and \( q \)-th AP), we set \( \text{RSS}_p = \{ z_{1p}^{\text{off}}, \cdots, z_{\gamma p}^{\text{off}} \} \) and \( \text{RSS}_q = \{ z_{1q}^{\text{off}}, \cdots, z_{\gamma q}^{\text{off}} \} \), as their standardized offline RSS features, and then construct the set of scatters corresponding to \( \text{RSS}_p \) and \( \text{RSS}_q \) as \( D_{pq} = \{ (z_{sp}^{\text{off}}, z_{sq}^{\text{off}}), s = 1, \cdots, \gamma \} \). In this case, for any given division \( G_{r \times c} \) having the scale \( r \times c \), the scatter plot \( \Psi \) which is composed of scatters in \( D_{pq} \) can be divided into \( r \times c \) regions, where \( r \) and \( c \) stand for the number of divided rows and columns respectively. Based on this, we calculate the mutual information for the \( p \)-th and \( q \)-th AP under the division \( G_{r \times c} \) as

\[
I(p,q) = H(p) + H(q) - H(p,q)
\]

(3)

where

\[
H(p) = - \sum_{z_{sp}^{\text{off}} \in \text{RSS}_p} p(z_{sp}^{\text{off}}) \log_2 p(z_{sp}^{\text{off}})
\]

(4)

\[
H(q) = - \sum_{z_{sq}^{\text{off}} \in \text{RSS}_q} p(z_{sq}^{\text{off}}) \log_2 p(z_{sq}^{\text{off}})
\]

(5)

\[
H(p,q) = - \sum_{z_{sp}^{\text{off}} \in \text{RSS}_p} \sum_{z_{sq}^{\text{off}} \in \text{RSS}_q} p(z_{sp}^{\text{off}}, z_{sq}^{\text{off}}) \log_2 p(z_{sp}^{\text{off}}, z_{sq}^{\text{off}})
\]

(6)

where \( \gamma = \max \{ p(z_{sp}^{\text{off}}), p(z_{sq}^{\text{off}}) \} \) \( p(z_{sp}^{\text{off}}) \) and \( p(z_{sq}^{\text{off}}) \) stand for the marginal probability density of \( z_{sp}^{\text{off}} \) and \( z_{sq}^{\text{off}} \) respectively, besides, \( p(z_{sp}^{\text{off}}, z_{sq}^{\text{off}}) \) represents the joint probability density of \( z_{sp}^{\text{off}} \) and \( z_{sq}^{\text{off}} \). The calculation of \( p(z_{sp}^{\text{off}}, z_{sq}^{\text{off}}) \) and \( p(z_{sp}^{\text{off}}) \) is shown as follows.

i) To calculate \( p(z_{sp}^{\text{off}}, z_{sq}^{\text{off}}) \), we use the concept of the two-dimensional histogram [26], which gives

\[
p(z_{sp}^{\text{off}}, z_{sq}^{\text{off}}) = \frac{\text{card} \left( \Omega_f \right)}{x \cdot \text{area} \left( \Omega_f \right)} \cdot \frac{\left( z_{sp}^{\text{off}}, z_{sq}^{\text{off}} \right) \in \Omega_f}{\left( z_{sp}^{\text{off}}, z_{sq}^{\text{off}} \right) \notin \Psi}
\]

(7)

where \( \Omega_f (f = 1, \cdots, r \times c) \) represents the \( f \)-th region under the division \( G_{r \times c} \) and \( \bigcup_{f=1}^{r \times c} \Omega_f = \Psi, \text{card} \left( \Omega_f \right) \) represents the number of scatters in \( \Omega_f \), and \( \text{area} \left( \Omega_f \right) \) represents the area of \( \Omega_f \).

ii) To calculate \( p(z_{sp}^{\text{off}}) \), we define \( z_{sp}^{\min} = \max \{ z_{1p}^{\text{off}}, \cdots, z_{\gamma p}^{\text{off}} \} \) and \( z_{sp}^{\max} = \min \{ z_{1p}^{\text{off}}, \cdots, z_{\gamma p}^{\text{off}} \} \), and divide the interval \( \Delta p = \left[ \frac{z_{sp}^{\min} + \xi_1 \tau, z_{sp}^{\max} + \xi_1 \tau} \right] \) \( (\xi = 1, \cdots, \tau) \) as the \( \eta \)-th sub-interval, and use the concept of the one-dimensional histogram [27] to obtain

\[
p(z_{sp}^{\text{off}}) = \frac{\text{card} \left( \Delta p \right)}{x \cdot \text{area} \left( \Delta p \right)} \cdot \frac{z_{sp}^{\text{off}} \in \Delta p}{z_{sp}^{\text{off}} \notin \Delta p}
\]

(8)

where \( \text{card} \left( \Delta p \right) \) represents the number of elements (in the set \( \{ z_{1p}^{\text{off}}, \cdots, z_{\gamma p}^{\text{off}} \} \) in \( \Delta p \).

For each scale \( r \times c (< m^{0.6}) \) [28], we calculate the corresponding mutual information for the \( p \)-th and the \( q \)-th AP under different divisions, and then define the maximum of the calculated mutual information as the mutual information with
the scale $r \times c$, i.e., $I_{rs} (p, q)$. Besides, for the sake of comparing the difference of the mutual information with different scales, we normalize $I_{rs} (p, q)$ as $I_{rs} = I_{rs} (p, q) / \log_2 \min (r, c)$, and then obtain the corresponding mutual information feature set as $I (p, q) = \{ I_{rs} (p, q) / r < m^{0.6} \}$. Here, the maximum value of elements $I (p, q)$ is defined as the MIC between the $p$-th and the $q$-th AP, given as $m_{pq} = \max (I (p, q))$. Thus, the larger value $m_{pq}$ indicates the higher correlation between the $p$-th and the $q$-th AP, which means the higher mutual substitutability between these two APs for positioning.

3) NON-REDUNDANT AP SELECTION

According to the MIC between different APs, we construct the MIC matrix (or called the fuzzy similarity matrix) with respect to the AP correlation as

$$\mathbf{M} = \begin{bmatrix} m_{11} & m_{12} & \cdots & m_{1n} \\ m_{21} & m_{22} & \cdots & m_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ m_{n1} & m_{n2} & \cdots & m_{nn} \end{bmatrix}$$

(9)

Based on the equivalence theory of the fuzzy similarity matrix [29], we start from $\mathbf{M}$ to calculate the second power successively, i.e., $\mathbf{M}^2 = \mathbf{M} \circ \mathbf{M}$ ($l = 1, \ldots, \log_2 n$), where the notation “$\circ$” represents the Zadeh synthesis operation, such that

$$\mathbf{M} \rightarrow \mathbf{M}^2 \rightarrow \cdots \mathbf{M}^l \rightarrow \cdots \rightarrow \mathbf{M}^{\log_2 n} \quad (10)$$

When the relation $\mathbf{M} \circ \mathbf{M} = \mathbf{M}$ first appears, $\mathbf{M}^l$ is defined as the fuzzy equivalent matrix of $\mathbf{M}$, denoted as

$$\mathbf{M}^* = (\tilde{m}_{pq}) = \begin{bmatrix} \tilde{m}_{11} & \tilde{m}_{12} & \cdots & \tilde{m}_{1n} \\ \tilde{m}_{21} & \tilde{m}_{22} & \cdots & \tilde{m}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{m}_{n1} & \tilde{m}_{n2} & \cdots & \tilde{m}_{nn} \end{bmatrix} = \mathbf{M}^l \quad (11)$$

Then, we construct the $\beta$-cut matrix of $\mathbf{M}^*$, i.e., $\mathbf{M}^*_{\beta} = (\tilde{m}_{pq}^{(\beta)})$ ($\beta \in [0, 1]$), in which the element on the $p$-th row and $q$-th column is

$$\tilde{m}_{pq}^{(\beta)} = \begin{cases} 1, & \tilde{m}_{pq} \geq \beta \\ 0, & \tilde{m}_{pq} < \beta \end{cases} \quad (12)$$

After obtaining $\mathbf{M}^*_{\beta}$, we continue to perform fuzzy clustering [30] on APs to obtain $C_1, \ldots, C_K$, where $C_k (k = 1, \ldots, K)$ represents the $k$-th set of the correlated APs and $K$ represents the number of clusters. Finally, to conduct the non-redundant AP selection, we select the AP with the highest location resolution (which will be illustrated in detail in the later Section IV(a)) in each cluster for positioning.

B. LOCATION RESOLUTION IDENTIFICATION

The second task in the offline phase is to evaluate the location resolution of RSS features, which can help in selecting APs with high location resolution for positioning. To achieve this goal, we calculate the RSS feature fuzzy weight from the AP information gain ratio, and then use it during the online AP fuzzy membership calculation to obtain APs with the high location resolution.

1) AP INFORMATION GAIN RATIO CALCULATION

According to $z_{s}^{\text{off}}$, we divide RPs into U sets, denoted as $U$ sets, denoted as $C_{i1}, \ldots, C_{iL}$, and RPs in each set, i.e., $C_{ui} (u = 1, \ldots, U)$, have the same offline RSS feature from the $s$-th AP. Also, we divide RPs into $V$ sets according to $z_{s}^{\text{off}} (s = 1, \ldots, \chi)$, denoted as $D_{s1}, \ldots, D_{sV}$, and RPs in each set, i.e., $D_{sv} (v = 1, \ldots, V)$, have the same $s$-th offline RSS feature from the $i$-th AP. Then, we calculate the uncertainty of RPs with respect to the $i$-th AP as

$$H(P | i) = - \sum_{s=1}^{\chi} \sum_{v=1}^{V} \frac{|D_{sv}|}{m} \log_2 \frac{|D_{sv}|}{m} \quad (13)$$

where $|D_{sv}|$ represents the number of RPs in $D_{sv}$, $H(D_{sv}) = \sum_{u=1}^{U} \frac{|D_{su}|}{m} \log_2 \frac{|D_{su}|}{m}$ represents the uncertainty of RPs in $D_{sv}$, and $D_{sv} = D_{sv} \cap C_{ui}$ and $H(P | i)$ represents the uncertainty of RPs classification according to $\chi$ multi-dimensional RSS features about the $i$-th AP. Moreover, according to $z_{s}^{\text{off}}$, the uncertainty of RPs for the $s$-th offline RSS feature from the $i$-th AP is calculated as

$$H(P | i)_s = - \sum_{v=1}^{V} \frac{|D_{sv}|}{m} \log_2 \frac{|D_{sv}|}{m} \quad (14)$$

and $H(P | i)_s$ represents the uncertainty of RPs classification according to the $s$ ($s = 1, \ldots, \chi$)-th feature of the $\chi$ multi-dimensional RSS features about the $i$-th AP.

Finally, we calculate the information gain ratio [30] for the $s$-th offline RSS feature from the $i$-th AP as $\psi_s^i = \frac{H(P) - H(P | i)}{H(P | i)}$, where $H(P) = \log_2 m$ under the equal prior probability of each RP condition. In this case, the AP information gain ratio matrix for $\chi$ offline RSS features is constructed as

$$\Phi = [\psi_1, \ldots, \psi_\chi]^T = \begin{bmatrix} \psi_1^1 & \psi_1^2 & \cdots & \psi_1^\chi \\ \psi_2^1 & \psi_2^2 & \cdots & \psi_2^\chi \\ \vdots & \vdots & \ddots & \vdots \\ \psi_\chi^1 & \psi_\chi^2 & \cdots & \psi_\chi^\chi \end{bmatrix} \quad (15)$$

where $\psi_s = (\psi_s^1, \ldots, \psi_s^\chi)$ and $\psi_s^i$ represents the AP information gain ratio for the $s$-th offline RSS feature from the $i$-th AP. The larger value $\psi_s^i$ indicates a significant decrease in the uncertainty of RPs when the $s$-th offline RSS feature from the $i$-th AP is considered.

2) RSS FEATURE FUZZY WEIGHT CALCULATION

After the AP information gain ratio is obtained, we continue to calculate the RSS feature fuzzy weight. In concrete terms, first, we perform the normalization of $\Phi$ to obtain the offline AP fuzzy membership matrix as $\tilde{b}_{i}^{\text{off}} = (\tilde{b}_{i1}^{\text{off}}, \ldots, \tilde{b}_{i\chi}^{\text{off}})$, where $\tilde{b}_{i}^{\text{off}} = \frac{\sum_{s=1}^{\chi} \psi_s^i}{\sum_{s=1}^{\chi} \sum_{i=1}^{\chi} \psi_s^i}$ represents the $i$-th offline AP.
fuzzy membership. According to the definition of the AP information gain ratio matrix, the larger value \( b_{ij}^{off} \) indicates the higher location resolution of the offline RSS features from the \( i \)-th AP.

Second, we calculate the fuzzy membership of the \( i \)-th AP for the \( s \)-th offline RSS feature as \( r_{si}^{off} = \frac{a_{si}^{off}}{\sum_{i=1}^{\chi} a_{si}^{off}} \), and then construct the corresponding fuzzy membership matrix as

\[
R^{off} = \begin{bmatrix}
r_{11}^{off} & \cdots & r_{1n}^{off} \\
r_{21}^{off} & \cdots & r_{2n}^{off} \\
\vdots & \ddots & \vdots \\
r_{\chi 1}^{off} & \cdots & r_{\chi n}^{off}
\end{bmatrix}
\]  

(16)

Third, by supposing that the RSS feature fuzzy weight matrix is \( A^{off} = (a_{1}^{off}, \ldots, a_{\chi}^{off}) \), where \( a_{i}^{off} \) represents the \( s \)-th RSS feature fuzzy weight, the fuzzy relation equation for offline RSS features can be given as \( \hat{B}^{off} = A^{off} \circ R^{off} \), such that

\[
\left(b_{1}^{off}, \ldots, b_{n}^{off}\right) = \left(a_{1}^{off}, \ldots, a_{\chi}^{off}\right) \circ \begin{bmatrix}
r_{11}^{off} & \cdots & r_{1n}^{off} \\
r_{21}^{off} & \cdots & r_{2n}^{off} \\
\vdots & \ddots & \vdots \\
r_{\chi 1}^{off} & \cdots & r_{\chi n}^{off}
\end{bmatrix}
\]  

(17)

where \( b_{j}^{off} = \bigvee_{i=1}^{s} \left(r_{si}^{off} \wedge a_{i}^{off} \right) = \max \left\{ \left( a_{i}^{off} \wedge r_{si}^{off} \right), \ldots, \left( a_{\chi}^{off} \wedge r_{si}^{off} \right) \right\} \) and \( d_{s}^{off} = \min \left\{ a_{s}^{off}, r_{si}^{off} \right\} \). For further clarity, we rewrite (17) into

\[
\begin{align*}
\left( a_{1}^{off} \wedge r_{11}^{off} \right) \vee \cdots \vee \left( a_{\chi}^{off} \wedge r_{\chi 1}^{off} \right) & = b_{1}^{off} \\
\left( a_{1}^{off} \wedge r_{12}^{off} \right) \vee \cdots \vee \left( a_{\chi}^{off} \wedge r_{\chi 2}^{off} \right) & = b_{2}^{off} \\
\vdots & \\
\left( a_{1}^{off} \wedge r_{1n}^{off} \right) \vee \cdots \vee \left( a_{\chi}^{off} \wedge r_{\chi n}^{off} \right) & = b_{n}^{off}
\end{align*}
\]  

(18)

At this point, the sufficient and necessary condition to provide the solution to the \( i \)-th equation in (18), i.e., \( (a_{1}^{off} \wedge r_{11}^{off}) \vee \cdots \vee (a_{\chi}^{off} \wedge r_{\chi 1}^{off}) = b_{1}^{off} \), is “at least one of these \( \chi \) fuzzy equations \((a_{i}^{off} \wedge r_{ij}^{off}) = b_{i}^{off}\), \ldots, \((a_{\chi}^{off} \wedge r_{ij}^{off}) = b_{i}^{off}\) and these \( \chi \) fuzzy inequalities \((a_{i}^{off} \wedge r_{ij}^{off}) \leq b_{i}^{off}\), \ldots, \((a_{\chi}^{off} \wedge r_{ij}^{off}) \leq b_{i}^{off}\) hold simultaneously”. Then, we obtain the \( s \)-th solution to the \( i \)-th equation in (18) as

\[
da_{i} = \left(b_{1}^{off} \cdot r_{1i}^{off}, \ldots, b_{\chi}^{off} \cdot r_{\chi i}^{off}\right)
\]

(19)

Finally, from (19), we obtain the solution to the \( i \)-th equation in (18) as \( A^{(i)} = A^{off} \cup \cdots \cup A^{(i)} \). Then, the solution to (18) is given as \( A^{off} = A^{(1)} \cap \cdots \cap A^{(m)} \), which is also recognized as the RSS feature fuzzy weight matrix. Here, it is noteworthy that the larger value \( a_{i}^{off} \) indicates the higher location resolution of the \( s \)-th offline RSS feature.

IV. ONLINE PHASE

A. ONLINE AP FUZZY MEMBERSHIP CALCULATION

By supposing that there are \( w \) Target Positions (TPs) to be located in the environment, we set \( rss_{\omega}^{\text{onli}} (\omega = 1, \ldots, w) \) as the \( s \)-th online RSS feature from the \( i \)-th AP at the \( \omega \)-th target position, and then construct the online RSS feature matrix as \( Z^{\text{onli}} = (z_{1}^{\text{onli}}, \ldots, z_{\chi}^{\text{onli}}) \) in (22), where \( z_{\omega}^{\text{onli}} = (\xi_{\omega 1}^{\text{onli}}, \ldots, \xi_{\omega \chi}^{\text{onli}}) \) and \( z_{\omega s}^{\text{onli}} \) represents the \( s \)-th online RSS feature from the \( i \)-th AP, which is given by \( z_{\omega s}^{\text{onli}} = \sum_{i=1}^{\chi} rss_{\omega}^{\text{onli}} \).

\[
Z^{\text{onli}} = \begin{bmatrix}
z_{1}^{\text{onli}} \\
z_{2}^{\text{onli}} \\
\vdots \\
z_{\chi}^{\text{onli}}
\end{bmatrix} = \begin{bmatrix}
z_{11}^{\text{onli}} & \cdots & z_{1\chi}^{\text{onli}} \\
z_{21}^{\text{onli}} & \cdots & z_{2\chi}^{\text{onli}} \\
\vdots & \ddots & \vdots \\
z_{\chi 1}^{\text{onli}} & \cdots & z_{\chi \chi}^{\text{onli}}
\end{bmatrix}
\]  

(22)

After that, we construct the fuzzy mapping from \( Z^{\text{onli}} \) to the set of online AP fuzzy membership \( \phi = \left( \xi_{1}^{\text{onli}}, \ldots, \xi_{\chi}^{\text{onli}} \right) \) in (23), where \( r_{s}^{\text{onli}} = (r_{s1}^{\text{onli}}, \ldots, r_{s\chi}^{\text{onli}}) \) and \( r_{s}^{\text{onli}} = \sum_{i=1}^{\chi} z_{s i}^{\text{onli}} \) represents the \( i \)-th online membership for the \( s \)-th online RSS feature.

\[
\hat{f} : Z^{\text{onli}} \rightarrow \phi
\]

\[
\begin{align*}
\xi_{1}^{\text{onli}} & \mapsto \hat{f}(\xi_{1}^{\text{onli}}) = \xi_{1}^{\text{onli}} \\
\xi_{2}^{\text{onli}} & \mapsto \hat{f}(\xi_{2}^{\text{onli}}) = \xi_{2}^{\text{onli}} \\
\vdots & \\
\xi_{\chi}^{\text{onli}} & \mapsto \hat{f}(\xi_{\chi}^{\text{onli}}) = \xi_{\chi}^{\text{onli}}
\end{align*}
\]  

(23)

From (23), the fuzzy judgment matrix for online RSS features can be constructed as

\[
R^{\text{onli}} = \begin{bmatrix}
r_{11}^{\text{onli}} & \cdots & r_{1\chi}^{\text{onli}} \\
r_{21}^{\text{onli}} & \cdots & r_{2\chi}^{\text{onli}} \\
\vdots & \ddots & \vdots \\
r_{\chi 1}^{\text{onli}} & \cdots & r_{\chi \chi}^{\text{onli}}
\end{bmatrix}
\]  

(24)

Finally, we use \( A^{off} \) and \( R^{\text{onli}} \) to construct the fuzzy relation equation for online RSS features as

\[
\phi = A^{off} \circ R^{\text{onli}} = \begin{bmatrix}
r_{11}^{off} & \cdots & r_{1\chi}^{off} \\
r_{21}^{off} & \cdots & r_{2\chi}^{off} \\
\vdots & \ddots & \vdots \\
r_{\chi 1}^{off} & \cdots & r_{\chi \chi}^{off}
\end{bmatrix} \circ \begin{bmatrix}
r_{11}^{\text{onli}} & \cdots & r_{1\chi}^{\text{onli}} \\
r_{21}^{\text{onli}} & \cdots & r_{2\chi}^{\text{onli}} \\
\vdots & \ddots & \vdots \\
r_{\chi 1}^{\text{onli}} & \cdots & r_{\chi \chi}^{\text{onli}}
\end{bmatrix}
\]  

(25)
where $\tilde{b}_{\text{onli}} = (b_{\text{onli}}^1, \ldots, b_{\text{onli}}^n)$, $S_h_{\text{onli}} = \bigwedge_{s=1}^{K} (a_{\text{off}}^s \land r_{\text{off}}^s)$ represents the $i$-th online AP fuzzy membership, and $a_{\text{off}}^s \land r_{\text{off}}^s = \min\{a_{\text{off}}^s, r_{\text{off}}^s\}$. Here, the larger value $b_{\text{onli}}^i$ indicates the higher location resolution of the $i$-th AP and thereby the more contribution of the $i$-th AP to the positioning accuracy is higher.

### B. TARGET POSITION ESTIMATION

We suppose that coordinates of the $j$-th RP and $\omega$-th TP are $(x_j, y_j)$ and $(x_\omega, y_\omega)$ respectively. By selecting the AP with the highest location resolution in each cluster of the correlated APs for positioning, the $\omega$-th TP is estimated based on the Bayesian positioning algorithm, such that

$$
(x_\omega, y_\omega) = \arg \max_{j=1, \ldots, m} \prod_{k=1}^{K} p(leftRSS_{k\omega}|(x_j, y_j))
$$

$$
= \arg \max_{j=1, \ldots, m} \prod_{k=1}^{K} p\left(\frac{1}{\sqrt{2\pi} \delta_{kj}} e^{-\frac{(x_{\text{onli}} - x_j)^2}{2\delta_{kj}^2}}\right)
$$

where $RSS_{k\omega}$ represents the $s$-th online RSS feature from the $k$-th selected AP at the $\omega$-th TP, $p\left(\frac{1}{\sqrt{2\pi} \delta_{kj}} e^{-\frac{(x_{\text{onli}} - x_j)^2}{2\delta_{kj}^2}}\right)$ represents the conditional probability density of $RSS_{k\omega}$ at $(x_j, y_j)$, and $\mu_{kj}$ and $\delta_{kj}$ stand for the mean and standard deviation of the $s$-th offline RSS feature from the $k$-th selected AP at the $j$-th RP respectively.

### V. EXPERIMENTAL RESULTS

In an actual indoor environment with dimensions of $56.93$ m by $20.08$ m, two straight corridors and one lab, namely Areas 1, 2, and 3, are selected as the testing area shown in Fig. 2. There are total 8 APs (noted from AP 1 to 8) deployed in this area, in which 88 RPs and 25 TPs are uniformly calibrated. Besides, at each RP and TP, we do the RSS data collection for $10$ s and $5$ s respectively with the sampling rate $1$ Hz.

### A. ANALYSIS OF SYSTEM PARAMETERS

Fig. 3 shows the variation of the average positioning error with the increase of the online AP fuzzy membership under different noise power conditions. From this figure, we can find that as the online AP fuzzy membership increases, the curve of the average positioning error shows a downward trend, which indicates that the online AP fuzzy membership is positively correlated with the contribution to the positioning accuracy of the corresponding AP.

For the better illustration of the impact of the online AP fuzzy membership on the positioning accuracy, Fig. 4 shows the Cumulative Density Function (CDF) of positioning errors by using different number of APs (from 3 to 8) with the largest online AP fuzzy membership for positioning. As the number of APs increases from 3 to 8, the sum of the corresponding online AP fuzzy membership, i.e., the value $\lambda$, increases from 2.79 to 5.36, and the CDF curve of positioning errors shows an upward trend. By taking the CDF of positioning errors in the Area 1 as an example, the confidence probability of positioning errors within $4$ m is $73.88\%$, $82.60\%$, $86.72\%$, $89.27\%$, $93.12\%$, and $94.85\%$ when the value $\lambda$ equals to $2.79$, $3.56$, $4.19$, $4.73$, $5.10$, and $5.36$ respectively.

Besides, Fig. 5 shows the variation of the average positioning error with the increase of the value $\omega$, from which we can find that as the value $\omega$ increases, i.e., the increase in of the number of APs, the curve of the average positioning error shows a downward trend and gradually converges. This result further verifies that the proposed method is able to achieve the high positioning accuracy and low positioning overhead by selecting APs with the high location resolution for positioning.

Fig. 6 shows the average positioning error under different values of $\beta$ condition. In general, the average positioning error decreases by increasing the value $\beta$, which is due to the reason that by increasing of the value $\beta$, the number of clusters by fuzzy clustering increases, and thereby the number of APs used for positioning also increases, thus resulting in a downward trend of the curve of the average positioning error. However, when the value $\beta$, increases to a certain extent (e.g., $\beta \geq 0.7$), any further the increase of the value $\beta$, cannot significantly affect the average positioning error since the redundant APs may be included for positioning. In addition, when $\beta = 0.7$ and $\beta = 0.8$, the clustering results of APs are the same, thus the APs used for positioning are also the same, which results in the same average positioning error.

### B. ANALYSIS OF POSITIONING OVERHEAD

The variation of the positioning overhead with the increase of the value $\lambda$ under different noise power conditions is shown in Fig. 7. From this figure, we can find that with the increase of the value $\lambda$, the number of APs used for positioning increases, and thereby the positioning overhead is on the rise as expected. By taking the value $\lambda$ increasing from $0.93$ to $5.13$ as an example, the positioning overhead for TPs in Areas 1, 2, and 3 increases by $6.94$ s, $6.98$ s, and $5.72$ s respectively. Besides, in comparison with Areas 2 and 3, the positioning overhead for TPs in the Area 1 involves more number of TPs, and thereby higher positioning overhead is required for the AP redundancy reduction and location resolution calculation.
Besides, Fig. 8 compares the positioning overhead with different values of $\beta$ under different noise power conditions. In this figure, we can find that as the value $\beta$ increases, the number of APs recognized as the non-redundant ones also increases, and thereby the positioning overhead required for target position estimation increases as expected. By taking the noise power $2 \text{ dBm}^2$ in the Area 1 as an example, the positioning overhead increases by 8.10s for an increase of $\beta$ increases from 0.1 to 1.

Furthermore, Fig. 9 shows a comparison of the positioning overhead of the proposed positioning method and four other existing ones, i.e., T-test [22], maximum likelihood [32],
InforGain [33], and max-mean [34] positioning methods, under different noise power conditions. From this figure, it is observed that compared with the T-test positioning method (the one having the highest positioning overhead), the positioning overhead of the proposed one is reduced by a maximum of 5.41 s. Comparatively speaking, the positioning overhead of the max-mean positioning method is the lowest since it depends on simply calculating the RSS mean at each RP to conduct AP optimization. However, from the positioning result in the following section, we can find that the positioning accuracy of this method is still low, which is due to the reason that it only considers the RSS mean for AP optimization while ignoring the diversity of RSS features.

C. ANALYSIS OF POSITIONING ACCURACY
Fig. 10 shows the bitmap of the average positioning error by different positioning methods. As can be seen from this figure, compared with the InforGain, max-mean, all-AP (i.e., using all APs for positioning) [35], strongest-AP (i.e., selecting the AP with the strongest RSS for positioning) [35], and
max-variance (i.e., selecting the AP with the maximum RSS variance for positioning) [35] positioning methods, the proposed one achieves the smallest average positioning error. That is because it calculates the AP correlation and RSS feature fuzzy weight to reduce the AP redundancy and identify the location resolution of APs used for positioning.

Fig. 11 shows the CDF of positioning errors by the T-test, maximum likelihood, InforGain, and max-mean positioning methods as well as the proposed one. As expected, the confidence probability of positioning errors within 4 m by the proposed method, i.e., 90.65%, is higher than the ones by the other four methods, i.e., 89.89%, 80.43%, 77.01%, and 68.30% respectively. To further demonstrate the effectiveness of conducting AP optimization, Fig. 12 compares the average positioning error by the T-test, maximum likelihood, InforGain, weighted least square [36], median KNN [37], and mean KNN [21] positioning methods with and without AP optimization which refers to selecting the non-redundant APs with high position resolution, from which the promising effect of AP optimization in the positioning accuracy is verified.

VI. CONCLUSION

In complex environments, due to the widely-deployed Wi-Fi APs at various indoor locations, the AP redundancy and the location resolution of different APs used for positioning are often ignored in the conventional indoor positioning methods. To deal with such problems, this article proposes a new multi-dimensional RSS feature fuzzy mapping and clustering-based AP optimization for Wi-Fi positioning. Specifically, in the offline phase, after the pre-processing of the offline RSS data, the MIC between different APs is calculated as the corresponding AP correlation, based on which the fuzzy clustering algorithm is used to select non-redundant APs for positioning. After that, the AP information gain ratio is obtained from the offline RSS features to calculate the RSS feature fuzzy weight. Then, in the online phase, the non-redundant APs with the high location resolution are selected for target position estimation based on the Bayesian positioning algorithm. Experimental results demonstrate that the proposed method can not only eliminate the redundant APs with the high substitutability to reduce the positioning overhead, but also select APs with the high location resolution for positioning to guarantee the high positioning accuracy.
accuracy. However, how to achieve the effective and efficient AP optimization in a more complicated and dynamic indoor environment will be an interesting work in the future.

REFERENCES

[1] S. Li, L. Xu, and S. Zhao, “The Internet of Things: A survey,” *Inf. Syst. Frontiers*, vol. 17, no. 2, pp. 243–259, 2015.

[2] I. Tomkos, D. Klonidis, E. Pikasis, and S. Theodoridis, “Toward the 6G network era: Opportunities and challenges,” *IT Prof.*, vol. 22, no. 1, pp. 34–38, Jan. 2020.

[3] J. Su, Z. Sheng, A. X. Liu, Z. Fu, and Y. Chen, “A time and energy saving-based frame adjustment strategy (TES-FAS) tag identification algorithm for UHF RFID systems,” *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 2974–2986, May 2020.

[4] L. Zhu, S. Sun, and W. Menzel, “Ultra-wideband (UWB) bandpass filters using multiple-mode resonator,” *IEEE Microw. Wireless Compon. Lett.*, vol. 15, no. 11, pp. 796–798, Nov. 2005.

[5] M.-G. Li, H. Zhu, S.-Z. You, and C.-Q. Tang, “UWB-based localization system aided with inertial sensor for underground coal mine applications,” *IEEE Sensors J.*, vol. 20, no. 12, pp. 6652–6669, Feb. 2020.

[6] X. Liu, C. Sun, M. Zhou, C. Wu, B. Peng, and P. Li, “Reinforcement learning-based multitool double-threshold spectrum sensing with Bayesian fusion for industrial big spectrum data,” *IEEE Trans. Ind. Informat.*, early access, Apr. 15, 2020, doi: 10.1109/TIIT.2020.2987421.

[7] A. A. Asatryan and Y. A. Kravtsov, “Localization of complex rays using Gaussian windows and the limits of applicability of complex geometric optics,” *Radiophys. Quantum Electron.*, vol. 31, no. 9, pp. 746–750, Sep. 1988.

[8] J. Su, A. X. Liu, Z. Sheng, and Y. Chen, “A partitioning approach to RFID identification,” *IEEE/ACM Trans. Netw.*, early access, Jul. 10, 2020, doi: 10.1109/TNET.2020.3004852.

[9] J. Su, Z. Sheng, A. X. Liu, Y. Han, and Y. Chen, “A group-based binary splitting algorithm for UHF RFID anti-collision systems,” *IEEE Trans. Commun.*, vol. 68, no. 2, pp. 998–1012, Feb. 2020.

[10] J. Su, Y. Chen, Z. Sheng, Z. Huang, and A. X. Liu, “From M-ary query to bit query: A new strategy for efficient large-scale RFID identification,” *IEEE Trans. Commun.*, vol. 68, no. 4, pp. 2381–2393, Apr. 2020.

[11] C. Manuel, L. Jesus, and B. Gussuppe, “An empirical study of the transmission power setting for Bluetooth-based indoor localization mechanisms,” *IEEE Sensors J.*, vol. 17, no. 6, pp. 1318–1332, 2017.

[12] L. B. Zuev, B. S. Semukhin, and N. V. Zarikovskaya, “Deformation localization and ultrasonic wave propagation rate in tensile al as a function of grain size,” *Int. J. Solids Struct.*, vol. 40, no. 4, pp. 941–950, Feb. 2003.

[13] K. Zheng, H. Wang, H. Li, W. Xiang, L. Lei, J. Qiao, and X. S. Shen, “Energy-efficient localization and tracking of mobile devices in wireless sensor networks,” *IEEE Trans. Veh. Technol.*, vol. 66, no. 3, pp. 2714–2726, Mar. 2017.

[14] S.-H. Yang, H.-S. Kim, Y.-H. Son, and S.-K. Han, “Three-dimensional visible light indoor localization using AOA and RSS with multiple optical receivers,” *J. Lightw. Technol.*, vol. 32, no. 14, pp. 2480–2485, Jul. 1, 2014.

[15] M. Zhou, Y. Wang, Y. Liu, and Z. Tian, “An information-theoretic view of WLAN localization error bound in GPS-denied environment,” *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 4089–4093, Apr. 2019.

[16] M. Zhou, Y. Liu, Y. Wang, and Z. Tian, “Anonymous crowdsourcing-based WLAN indoor localization,” *Digit. Commun. Netw.*, vol. 5, no. 4, pp. 226–236, Nov. 2019.

[17] X. Li, A. K. Sangaijah, S. Kumari, F. Wu, J. Shen, and M. K. Khan, “An efficient authentication and key agreement scheme with user anonymity for roaming service in smart city,” *Pers. Ubiquitous Comput.*, vol. 21, no. 12, pp. 1–15, 2017.

[18] K. Chen, Y. Fu, E. Cheng, and L. Huang, “Performance evaluation of an improved APIT localization algorithm for underwater acoustic sensor networks,” *J. Comput.*, vol. 29, no. 1, pp. 132–142, 2018.

[19] G. Wang, K. Wang, C. Gong, D. Zou, Z. Jiang, and Z. Xu, “A 1 Mbps real-time NLOS UV scattering communication system with receiver diversity over 1km,” *IEEE Photon. J.*, vol. 10, no. 2, pp. 1–13, Apr. 2018.

[20] E.-W. Bai and L. Ding, “Robust target localization in the absence of signal propagation models,” *Automatica*, vol. 76, no. 2, pp. 127–137, Feb. 2017.

[21] P. Bahl and V. N. Padmanabhan, “RADAR: An in-building RF-based user location and tracking system,” in *Proc. IEEE INFOCOM*, Mar. 2000, pp. 775–784.
WEI NIE received the Ph.D. degree from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2015. From July 2015 to July 2016, he worked as an Engineer in Huawei. Since August 2016, he has been working as a Lecturer with the Chongqing University of Posts and Telecommunications (CQUPT), Chongqing, China. His current research interests include microwave and mm-wave passive devices and antenna technology.

WEI HE (Graduate Student Member, IEEE) received the B.S. degree in telecommunication engineering from the Chongqing University of Posts and Telecommunications (CQUPT), China, in 2017, where she is currently pursuing the M.S. degree in information and communication engineering. Her current research interests include indoor localization, gesture recognition, and machine learning.

QIAOLIN PU received the B.S. and M.S. degrees in information and communication engineering from the Chongqing University of Posts and Telecommunications (CQUPT). She is currently pursuing the Ph.D. degree in communication and information engineering with the Hong Kong Baptist University (HKBU). Her research interests include indoor localization, location privacy, and network optimization.