Access Point Reselection and Adaptive Cluster Splitting-Based Indoor Localization in Wireless Local Area Networks

Dong Liang, Zhaojing Zhang, and Mugen Peng, Senior Member, IEEE

Abstract—Indoor localization technology has received increasing attention because the hobbies and interests of human can be mined from the location data. Wireless local area network (WLAN)-based fingerprinting localization methods have become attractive owing to their advantages of open access and low cost. However, for localization in realistic large areas, three problems persist: 1) excessive memory requirements in the offline phase; 2) high computational complexity in the online phase; and 3) how to select the access point (AP) sets with best distinction capability. A novel method of localization based on adaptive cluster splitting (ACSS) and AP reselection is proposed in this paper. The suggested method can significantly reduce the requirements of offline storage and online computing capacity while improving the localization accuracy. The expected result is demonstrated in a theoretical deduction, simulation, and with experiments in a realistic environment.

Index Terms—Access point (AP) reselection, adaptive cluster splitting (ACSS), decision tree, fingerprint (FP), indoor localization, wireless local area network (WLAN).

I. INTRODUCTION

BIG DATA, also referred to as data intensive technology, have emerged as a new technology trend in science, industry, and business [1], [2]. Big Data have become closely related to many aspects of human activity from simply recording events to research, design, production and digital services, and product delivery to the consumer [3], [4]. With the rapid development of the mobile internet, an enormous amount of data is produced by mobile terminals, representing a significant commercial value [5], [6]. Moreover, the gathering of location information of mobile terminals has attracted considerable attention because the hobbies and interests of human can be analyzed using data mining methods.

Consequently, research on indoor localization has attracted considerable attention. Localization in traditional line-of-sight (LOS) environments such as outdoors has been a success, as exemplified by the popularity of global positioning system (GPS). Extending the success of GPS to indoor environments, however, faces numerous fundamental challenges. The most notable of these are the non-LOS (NLOS) and multipath propagation problems [7].

To address these challenges, many different wireless indoor localization techniques and systems have been proposed [7]–[29]. These techniques can be categorized as nonsurvey-based and survey-based. The majority of the nonsurvey-based techniques require the position and transmitting power of the transmitter. By measuring the received signal strength (RSS) and substituting into a path loss model, the distance between the mobile terminal and transmitter can be estimated. The most common nonsurvey-based technique is the triangular method [8], which measures the distance between the mobile terminal and three transmitters positioned in different locations and then acquires the location estimation by solving algebraic functions. Clearly, nonsurvey-based techniques are constrained path loss model. Unfortunately, an accurate and universal path loss model is difficult to achieve in a complex and variable indoor environment [9].

Thus, the concept of a survey-based technique was suggested by Youssef et al. [10]. This is also known as a fingerprinting localization technique (FPLT). The basic idea is that at each location in space, there is a unique signature called a fingerprint (FP) that can be used to infer the unknown position. The FP itself is a manifestation of a physical phenomenon related to the interaction of a given signal [e.g., radio field (RF) and magnetic] with the environment. The most popular RF-based FP is RSS. The main reason is that RSS is readily available in many wireless standard implementations [e.g., cellular and wireless local area networks (WLAN)].

RSS-based FPLT is composed of two stages: 1) offline and 2) online [10]. In the offline phase, the main task is to create a database of RSS-based FPs associated with specific locations from the available access points (AP) across a given indoor environment. Location estimates are then obtained in the online phase using pattern recognition techniques where the measured FP parameters in an unknown location are compared to the database. It can be seen that FPLT is virtually unrelated to the path loss model and can supply more accurate localization estimations.

Theoretically, FPs are unique to individual locations. In reality, however, factors such as multipath propagation,
fading, time-varying artifacts (moving people/objects), and system/signal parameters (e.g., bandwidth and signal-to-noise ratio) reduce the uniqueness of the FP and create ambiguities in the position estimation stage. Thus, numerous researchers are now focusing on improving the performance from a pattern recognition perspective.

RSS-based fingerprinting localization using $k$-nearest neighbor ($k$NN) [11], [12] and probabilistic [13], [14] has been investigated. This research can improve the localization accuracy by exploiting the value of the offline database. To further improve the accuracy performance, statistical methods are being researched, where the probability of each potential position is analyzed using neural networks [15], [16] and support vector machines [17], [18], assuming that the RSS readings from different APs are independent at every instant of time. However, an explicit formulation of RSS distribution is challenging and the independence may not hold in real environments. Furthermore, these probabilistic-based systems frequently have high computational complexity, which makes it difficult to execute on mobile devices with limited processing power and small memory.

Choosing only a subset of detectable APs (also called AP selection) for the localization is an intuitive way to reduce the computational burden on and storage requirement of the devices [19]–[23]. Only the information from the selected APs is retained for the localization, whereas that from the unselected APs is discarded. The reduction of the data dimensions leads to a decrease in the computational complexity. The performance can be further enhanced when the discarded information is, in fact, a redundant noise. The discussion on AP selection has received focus in previous literature such as energy maximization [24], [25], information gain [26], [29], and principal component analysis [27], [28].

The above research can improve the localization accuracy to some extent in specific scenarios. However, for localization in realistic large areas, three problems persist the following.

1) The memory requirements for the offline FPs will be extensive in large areas. How to minimize the FP memory requirements with minimal localization accuracy lost?
2) The localization/distinction capability of different APs is somewhat different. This capability is also different in different locations considering the same AP. How to select the best AP set to perform the localization process?
3) An exhaustive search will deliver the best localization accuracy performance. However, the computational complexity is high, particularly in large areas. A fast-matching algorithm in the online phase is required.

In this paper, we propose a novel localization method that can reduce the offline storage and online computational complexity while improving the localization accuracy.

The contributions of this paper are as follows:
1) introduce a novel localization method that can reduce the offline storage and online computational complexity while improving the localization accuracy;
2) present a closed-form approximate solution of the localization error probability for both the traditional max likelihood (ML) method and the proposed method;
3) establish a demonstration application (APP) to compare the accuracy of the proposed method with a traditional method.

This paper is organized as follows. In Section II, we provide an overview of fingerprinting localization and establish the mathematical model of the indoor localization problem. In Section III, we describe the details of the proposed localization method. In Section IV, we analyze the theoretical localization error probability and propose a closed-form approximate solution for both the traditional and the proposed methods. In Sections V and VI, the simulation and demonstration results are discussed, respectively. Section VII is the conclusion of this paper.

II. FINGERPRINTING LOCALIZATION PROBLEM

Fingerprinting localization is based on a two-stage process. In the first stage, a database of offline FPs is created where each entry (FP vector) is associated with a location. In the second stage, an online FP is measured at an unknown location where it is compared to the offline database using well-known pattern recognition algorithms such as ML.

A. Offline Phase

In the offline stage, the localization area of interest is gridded as $I \times J$ squares with equal side $a$ (see Fig. 1). The offline FP of the $(i,j)$th location is denoted as

\[
\psi_w = (\psi_{w,1}, \ldots, \psi_{w,m}, \ldots, \psi_{w,M}) = (\psi_{(i,j),1}, \ldots, \psi_{(i,j),m}, \ldots, \psi_{(i,j),M})
\]

where $\psi_{w,m}$ denotes the RSS at location $w = (i,j)$ generated by the $m$th AP. For convenience of expression, we define

\[
\begin{align*}
I &= \{i | 1 \leq i \leq I\} \\
J &= \{j | 1 \leq j \leq J\} \\
W &= \{(i,j) | 1 \leq i \leq I \text{ and } 1 \leq j \leq J\} \\
M &= \{m | 1 \leq m \leq M\}
\end{align*}
\]
where $\psi_w$ is measured in offline time before online localization; unfortunately, $\psi_w$ is time-variable. Let $\mu_w$ and $\sigma_w$ represent the mean and standard deviation of $\psi_w$, where

$$\mu_w = (\mu_{w,1}, \ldots, \mu_{w,m}, \ldots, \mu_{w,M})$$

$$\sigma_w = (\sigma_{w,1}, \ldots, \sigma_{w,m}, \ldots, \sigma_{w,M}).$$

The radio map (RM) is then defined as

$$A = \begin{pmatrix} \mu_{(1,1)} & \cdots & \mu_{(1,J)} \\ \vdots & \ddots & \vdots \\ \mu_{(I,1)} & \cdots & \mu_{(I,J)} \end{pmatrix} \quad (5)$$

$$B = \begin{pmatrix} \sigma_{(1,1)} & \cdots & \sigma_{(1,J)} \\ \vdots & \ddots & \vdots \\ \sigma_{(I,1)} & \cdots & \sigma_{(I,J)} \end{pmatrix} \quad (6)$$

where $A$ is defined as the mean RM (MRM) and $B$ is defined as the standard deviation RM (SRM). RM is the database composed of all the offline FPs.

B. Online Phase

In the online stage, a mobile device at an unknown location $w_0 = (i_0, j_0)$ measures RF signatures to $M$ APs as the online FP, which is denoted as

$$\gamma = (\gamma_1, \ldots, \gamma_m, \ldots, \gamma_M).$$

The position is then estimated by comparing the online FP $\gamma$ to $A$ using well-known pattern recognition techniques. ML mapping is a popular algorithm with the theoretically smallest localization error. The ML estimation of $w_0$ is given as

$$\hat{w}_{\text{ML}} = \arg \max_{i,j} \Pr \{ \gamma|w = (i,j) \} \quad (8)$$

where $\Pr \{ X \}$ denotes the probability of event $X$. Assuming that the $M$ APs are independent of each other, (8) can be expressed as

$$\hat{w}_{\text{ML}} = \arg \max_{i,j} \left\{ \prod_{m \in M} \Pr \{ \gamma_m|w = (i,j) \} \right\}. \quad (9)$$

If the mobile device remains at location $w = (i,j)$, $\gamma_m$ will be subjected to a Gaussian distribution with mean $\mu_{w,m}$ and standard deviation $\sigma_{w,m}$ [5]. Then

$$\hat{w}_{\text{ML}} = \arg \max_{i,j} \left\{ \prod_{m \in M} \frac{1}{\sqrt{2\pi\sigma_{w,m}}} \exp \left[ -\frac{(\gamma_m - \mu_{w,m})^2}{2\sigma_{w,m}^2} \right] \right\}$$

$$= \arg \max_{i,j} \left\{ \prod_{m \in M} \frac{1}{\sigma_{w,m}} \exp \left[ -\frac{(\gamma_m - \mu_{w,m})^2}{2\sigma_{w,m}^2} \right] \right\}. \quad (10)$$

Typically, $\hat{w}_{\text{ML}}$ is difficult to achieve owing to multiple constraints; some approximate expressions are discussed in [6].

1) Assuming that $\sigma_{w,m}$ is equal for different $w$, then $\sigma_{w,m}$ can be simplified as $\sigma_m$

$$\hat{w}_{\text{ML}} = \arg \max_{i,j} \left\{ \prod_{m \in M} \frac{1}{\sigma_m} \exp \left[ -\frac{(\gamma_m - \mu_{w,m})^2}{2\sigma_m^2} \right] \right\}$$

$$= \arg \max_{i,j} \left\{ \prod_{m \in M} \exp \left[ -\frac{(\gamma_m - \mu_{w,m})^2}{2\sigma_m^2} \right] \right\}$$

$$= \arg \max_{i,j} \left\{ \exp \left[ -\frac{1}{2\sigma_m^2} \sum_{m \in M} (\gamma_m - \mu_{w,m})^2 \right] \right\}. \quad (11)$$

2) Assuming that $\sigma_m$ is equal for different $m$, then $\sigma_m$ can be simplified as $\sigma$

$$\hat{w}_{\text{ML}} = \arg \max_{i,j} \left\{ \prod_{m \in M} \frac{1}{\sigma^2} \exp \left[ -\frac{(\gamma_m - \mu_{w,m})^2}{2\sigma^2} \right] \right\}$$

$$= \arg \max_{i,j} \left\{ \exp \left[ -\frac{1}{2\sigma^2} \sum_{m \in M} (\gamma_m - \mu_{w,m})^2 \right] \right\}$$

$$= \arg \max_{i,j} \left\{ \exp \left[ -\frac{1}{2\sigma^2} \sum_{m \in M} \frac{(\gamma_m - \mu_{w,m})^2}{\sigma^2} \right] \right\}. \quad (12)$$

In reality, to eliminate the multiplication operation, $\hat{w}_{\text{ML}}$ is always approximately simplified as

$$\hat{w}_{\text{ML}} \approx \arg \max_{i,j} \left\{ \sum_{m \in M} |\gamma_m - \mu_{w,m}| \right\}$$

$$= \arg \max_{i,j} \left\{ |\gamma - \mu| \right\}. \quad (13)$$

C. AP Selection

In a general environment, $\sigma_{w,m}$ is always different based on $w$ and $m$ such that (13) cannot achieve the best localization performance.

To improve the localization performance, significant research has been addressed toward AP selection [24]–[26]. The objective of AP selection is to select APs with a stronger RSS and lower signal fluctuation. During online localization, only the FP components generated by these APs are considered. The ML estimation is then changed to

$$\hat{w}_{\text{ML+AS}} = \arg \min_{i,j} \left\{ \sum_{n \in N} |\gamma_n - \mu_{w,n}| \right\}$$

$$= \arg \min_{i,j} \left\{ |\gamma' - \mu'| \right\}. \quad (14)$$

where $N$ is a set composed of the selected APs called the best AP set. AS denotes the AP selection.

In this paper, we will not discuss the AP selection algorithm. We will simply use the mature algorithm in [29] for the proposed localization mechanism.
D. Offline Storage and Online Computational Complexity

Because there are $I \times J$ grids in the RM and each FP has $M$-dimensions, the offline storage can be calculated as

$$\alpha_{\text{ML}} = O(I \times J \times M).$$  \hfill (15)

In the online phase, to determine the ML estimation, a maximum of $I \times J$ rounds of distance calculation is required. In each distance calculation, there are $N + 1$ additions and subtractions. Consequently, the online computational complexity is approximately equal to

$$\beta_{\text{ML}} = O\left(\frac{1}{2} \times I \times J \times M\right).$$  \hfill (16)

It can be seen that $\alpha_{\text{ML}}$ and $\beta_{\text{ML}}$ are enormous, especially in a large area localization. For example, we conducted our experiments on the bottom floor in Xidan Joy-City, a shopping mall in Beijing. The localization area was approximately 40,000 m$^2$ (200 m $\times$ 200 m); approximately 200 APs were detected. In the experiment, the localization area was gridded as a maximum of 128 $\times$ 128 squares. With these specifications, $\alpha_{\text{ML}}$ and $\beta_{\text{ML}}$ can be calculated as

$$\alpha_{\text{ML}} = 128 \times 128 \times 200 = 32,768,000$$ \hfill (17)

$$\beta_{\text{ML}} = \frac{1}{2} \times 128 \times 128 \times 200 = 1,638,400.$$ \hfill (18)

Clearly, they are enormous and present a challenge for the localization calculation.

III. LOCALIZATION ALGORITHM USING AP RESELECTION AND ACS

As mentioned in Section I, for localization in realistic large areas, three problems persist: 1) extensive offline storage; 2) high online computational complexity; and 3) best AP selection. To overcome these problems, we propose a novel FP-based WLAN indoor localization mechanism called ACS. The ACS algorithm can be utilized in both 2-D and 3-D localization scenarios.

A. ACS Based on Quadtree (2-D Localization)

The basic idea of ACS is composed of three aspects.

1) In the offline phase, the traditional uniform RM (see Fig. 1) is changed to a variable-sized RM (VRM), which is composed of unisized grids (see Fig. 2).

2) The generation of the VRM is designed as an ACS process (see Figs. 2 and 3). Initially, the entire area is defined as the root cluster. Then, the root cluster is split into four subclusters equally. The splitting process is iterated until a termination condition is achieved. The iteration algorithm and termination condition are described in detail in Algorithm 1. The VRM is stored in the form of a Quadtree [30]. Similar ideas to the ACS process can be seen in [31] and [32].

3) In the online phase, the exhaustive search becomes a searching problem based on Quadtree.

ACS can improve the localization performance in three fashions.

1) The VRM can reduce the offline storage compared with the traditional uniform RM.

2) The computational complexity of the Quadtree search is significantly less than the traditional exhaustive search.

3) In the generation of the VRM, AP selection is executed in each step (called AP reselection), which means that the ACS is always utilizing the personalized best AP set for a smaller subarea (subcluster); not for the entire localization area.

The ACS algorithm is described in detail below.

1) Offline Phase: In the offline phase, the main task is to generate the VRM. The generation process is presented in Algorithm 1.

Algorithm 1. Code Generation for Offline VRM

Input: $q_0$, $A$
Output: $\Omega$

1: set empty stack $\Phi$
2: $\Phi$ $\leftarrow$ $q_0$ //push $q_0$ to $\Phi$
3: while $\Phi$ is not empty do
4: $q$ $\leftarrow$ $\Phi$ //pull a cluster out of $\Phi$, denote as $q$
5: $q_1$, $q_2$, $q_3$, $q_4$ $\leftarrow$ $q$ //pre-split $q$
6: $t_q$ $\leftarrow$ $A$
7: $v_q$ $\leftarrow$ $A$
8: $u_{q1}$, $u_{q2}$, $u_{q3}$, $u_{q4}$ $\leftarrow$ $A$

Fig. 2. Illustration of Quadtree-based VRM.

Fig. 3. Illustration of Quadtree.
9: \( \Gamma(t_q) \leftarrow t_q \)
10: \( \Omega \leftarrow q, u_{q,t}, v_q, \Gamma(t_q) \) //save cluster \( q \) to \( \Omega \)
11: if \( \Gamma(t_q) = 1 \) then
12: \( \Phi \leftarrow q_1, q_2, q_3, q_4 \)
13: end if
14: end while

I. Set a stack \( \Phi \) to store the clusters waiting to be split. Initially, there is only one cluster in the stack, called the root cluster \( q_0 \), which represents the entire localization area.

II. Presplit the current cluster \( q \), i.e., perform a virtual cross-split impartially on cluster \( q \) to generate four subclusters \( q_1, q_2, q_3, \) and \( q_4 \).

III. AP selection. Choose the best AP set that can distinguish the four subclusters. Many AP selection algorithms have been proposed. We adopt the method suggested in [29]. Its procedure is presented below.

1) Those grids with a small FP discrepancy are combined into one grid, which reduces the offline storage. Assume that \( T \) is the number of clusters stored in VRM. For each cluster, we must store \( 2N \) FPs. Therefore, the total storage for VRM is

\[
\alpha_{\text{ACS}} = 2 \times T \times N. \tag{25}
\]

2) Although some grids have been merged, the merging process is always performed in an area with high fluctuation of radio propagation. We know that localization accuracy is always inferior in these areas; hence, enlarging the grid size will not increase the localization error.

3) In the traditional localization process, the best AP set is fixed for the entire area. For ACS, each cluster has a unique best AP set. This personalized best AP set configuration can supply superior distinctions compared to the fixed. This means that an improved performance gain can be expected in the online phase. When localization is executed in a large area, the performance gain will be significant.
2) **Online Phase**: In the online phase, the localization process is designed as a searching problem based on Quadtree. The searching process is depicted below (see Algorithm 2).

**Algorithm 2. Code Generation for Online Localization**

**Input**: Ω, γ

**Output**: λ*

1: set \( k = 0, \lambda_k = 0 \)

2: if \( \Gamma(t_{\lambda_k}) = 1 \) then

3: \( \gamma(\lambda_k) \leftarrow \gamma_k, \nu_{\lambda_k} \)

4: \( u_1, u_2, u_3, u_4 \leftarrow \Omega \)

5: \( \lambda_{k+1} \leftarrow \arg\min_{\theta=1,2,3,4} \| x_{\gamma(\lambda_k)} - u_\theta \| \)

6: \( k = k + 1 \)

7: end if

8: \( \lambda^\ast = \lambda_k \)

I. Obtain the online FP vector \( \gamma = (\gamma_1, \ldots, \gamma_m, \ldots, \gamma_M) \), which is measured by the mobile device.

II. Quadtree searching. This process is presented in Algorithm 2. For convenience of expression, we use variable \( k \) to denote the tree level, and \( \lambda_k \) to denote the survive node in level \( k \). The detail searching process is presented below (also see Algorithm 2):

1) start with root node \( \lambda_0 \) and \( k = 0 \);

2) read \( \Gamma(t_{\lambda_k}) \) from the offline FP database \( \Omega \) to determine whether current node \( \lambda_k \) can be split. If yes, go to step 3); otherwise, go to step III;

3) read the filtered offline FP of the \( \lambda_k \)'s four subnodes from the offline FP database and denote as \( u_\theta \), where \( \theta = 1, 2, 3, \) and 4;

4) using AP reselection, the online vector \( \gamma \) is reselected and reordered by \( \nu_k \) as \( \gamma(\lambda_k) = (\gamma_{v_{k,1}}, \ldots, \gamma_{v_{k,j}}, \ldots, \gamma_{v_{k,N}}) \), where \( \nu_k \) is the best AP set for node \( \lambda_k \);

5) calculate the Euclidean distance between \( u_\theta \) and \( \gamma(\lambda_k) \) as

\[
d_\theta = \left\| \gamma(\lambda_k) - u_\theta \right\| = \frac{1}{N} \sqrt{\sum_{j=1}^{N} (u_{\theta,j} - \gamma_{v_{k,j}})^2} \tag{26}
\]

or simplified as

\[
d_\theta = \left\| \gamma(\lambda_k) - u_\theta \right\| = \frac{1}{N} \sqrt{\sum_{j=1}^{N} |u_{\theta,j} - \gamma_{v_{k,j}}|^2} \tag{27}
\]

The \( k + 1 \) level node is then chosen as

\[
\lambda_{k+1} = \arg\min_{\theta=1,2,3,4} |d_\theta| . \tag{28}
\]

6) \( k = k + 1 \). Go to step II. 2).

III. Output localization results.

1) The location estimation \( \lambda^\ast \) is chosen as the geometric center of current node \( \lambda_k \).

2) The estimation uncertainty is defined as the side of \( \lambda_k \).

ACS has three merits in the online phase.

1) Fast localization. Only a maximum of \( 4 \times K \) rounds of distance calculation are required. In each distance calculation, there are \( N + 1 \) additions and subtractions. Therefore, the online computational complexity is approximately equal to

\[
\beta_{ACS} = O\left(\frac{1}{2} \times 4 \times K \times N \right) = O\left(2 \times K \times N \right) . \tag{29}
\]

2) Better localization accuracy. In the flowchart of ACS, the location estimation proceeds step by step and the best AP set is replaced step by step. The ultimate location estimation is, therefore, determined from the best AP set according to the actual location \( w_0 \). The traditional ML estimation is selected from the best AP set based on
the entire localization area. Clearly, the former can provide superior localization accuracy. When localization is executed in a large area, the performance gain will be considerable.

3) ACS not only provides the location estimation but also calculates the estimation uncertainty, which is not available in traditional algorithms.

B. ACS Based on Octree (3-D Localization)

In a 3-D space localization scenario, the above ACS algorithm will function with minor modifications. The main modification is presented below.

1) Offline Phase: I. In 3-D space, the cluster is split as eight subclusters (see Fig. 6).
   II. Each subcluster contains an equal cuboid space.
   III. The length, width, and height may be unequal for each cuboid.

IV. The FPs are stored in Octree (see Fig. 7).

2) Online Phase: I. In the online phase, the localization process is designed as a searching problem based on Octree.

   II. If the searching process is terminated at node $\lambda_k$ (the last node that cannot be further split), then the location estimation is chosen as the geometric center of the cuboid represented by $\lambda_k$.

   III. The estimation uncertainty is defined as the longest side of $\lambda_k$.

IV. The online computational complexity is approximately equal to

$$\beta_{\text{ACS}} = O\left(\frac{1}{2} \times 8 \times K \times N\right) = O(4 \times K \times N).$$

IV. PERFORMANCE ANALYSIS BY THEORETICAL DETECTION

A. Comparison of Offline Storage

ACS has the benefit that it reduces the number of FPs that must be stored in the RM, i.e., it avoids storing those grids that cannot be distinguished by their best AP set. The compression ratio of offline storage between ACS and traditional ML can be calculated by (15) and (25) as

$$\xi_1 = O\left(\frac{2 \times T \times N}{I \times J \times M}\right).$$

In a realistic environment, $T$ is normally considerably less than $I \times J$. This is especially true in the case of finer granularities and stronger signal fluctuations. In this situation, more grids will be merged into one cluster, meaning a greater reduction of storage (see the illustration in Fig. 2).

Furthermore, $N$ is normally noticeably less than $M$. This is particularly true when there are many APs that can be detected by the mobile devices and there is considerable difference in the signal fluctuation among the different APs. Assuming the conditions $T \ll I \times J$ and $N \ll M$, we can multiply these two factors to obtain $T \times N \ll I \times J \times M$. As a result

$$\xi_1 \ll 1.$$  

B. Comparison of Online Computational Complexity

The compression ratio of online computational complexity between ACS and traditional ML can be calculated by (16) and (29) as

$$\xi_2 = O\left(\frac{2 \times K \times N}{\frac{1}{2} \times I \times J \times M}\right) = O\left(\frac{4 \times K \times N}{I \times J \times M}\right).$$

For a square area ($I = J$), it is simple to deduce that $K \leq \lfloor \log_2 I \rfloor$, therefore,

$$\xi_2 \leq O\left(\frac{4 \times \lfloor \log_2 I \rfloor \times N}{I \times J \times M}\right) \approx O\left(\frac{4N\log_2 I}{M}I^2\right) \ll 1.$$  

C. Comparison of Localization Error

1) Assumption: We mentioned, in Section II, that in reality, multipath propagation, fading, time varying artifacts (moving people/objects), and system/signal parameters (e.g., bandwidth and SNR) reduce the uniqueness of the FP and create ambiguities in the position estimation stage. Considering the complex and flexible environment, the analysis of localization error is difficult. In this section, we analyze the localization error probability under three assumptions.

   1) The signal propagation of different APs is independent.
   2) For the $n$th AP, the signal propagation follows the universal path loss model

$$\text{RSS} = \text{TSS} - \text{pathloss} + \chi = s_m - \alpha_{1,m} - \alpha_{2,m}\log_{10}d_m + \chi_m$$

where TSS denotes the transmitted signal strength, $d_m$ denotes the distance between the $n$th AP and the receiver, and $\chi_m$ denotes a random variable.
3) $\chi_m$ follows a Gaussian distribution with zero mean and variance $\sigma^2_{\chi_m}$ ($\sigma^2_{\chi_m}$ is unequal for different locations).

2) Localization Error Probability of Traditional ML Without AP Selection: In this paper, when the location estimation $\hat{w}$ is not equal to the actual position $w_0$, we call this event a localization error. For traditional ML estimation, the probability of localization error can be expressed as

$$P_{ML} = \Pr \{ \hat{w}_{ML} \neq w_0 \}$$

$$= \Pr \left\{ \sum_{m \in M} | \gamma_m - \mu_{w,m} | < \sum_{w \in W} \sum_{m \in M} | \gamma_m - \mu_{w_0,m} | \right\}$$

$$= \Pr \left\{ \sum_{m \in M} \left[ | \gamma_m - \mu_{w,m} | - | \gamma_m - \mu_{w_0,m} | \right] < 0 \right\}.$$ \hspace{1cm} (36)

Unfortunately, the closed form solution of $P_{ML}$ is difficult to achieve. We classify the location error into several categories. The most possible wrong location estimation is the grid (denoted as $w_1$) adjacent to the actual grid $w_0$. A second possible wrong location estimation is in the second circle (denoted as $w_2$) around $w_0$. A detailed explanation of $w_1$ and $w_2$ is presented in Fig. 8.

We define suberror probability $P_{ML(1)}, P_{ML(2)}, \ldots$ as

$$P_{ML(1)} = \Pr \left\{ \sum_{m \in M} | \gamma_m - \mu_{w_1,m} | < \sum_{m \in M} | \gamma_m - \mu_{w_0,m} | \right\}$$

$$P_{ML(2)} = \Pr \left\{ \sum_{m \in M} | \gamma_m - \mu_{w_2,m} | < \sum_{m \in M} | \gamma_m - \mu_{w_0,m} | \right\}.$$ \hspace{1cm} (37)

Without loss of generality, assuming $\mu_{w_0,m} < \mu_{w_1,m}$, $P_{ML(i)}$ can be simplified as

$$P_{ML(i)} = \Pr \left\{ \sum_{m \in M} \left[ \frac{\mu_{w_0,m} + \mu_{w_1,m}}{2} - \gamma_m \right] < 0 \right\}.$$ \hspace{1cm} (38)

Define $\rho_{w_i} = \sum_{m \in M} \left[ \frac{\mu_{w_0,m} + \mu_{w_i,m}}{2} - \gamma_m \right]$. Because $\gamma_m$ subjects to a Gaussian distribution, it is straightforward to demonstrate that $\rho_{w_i}$ also subjects to a Gaussian distribution.

Thus, the expectation of $\rho_{w_i}$ can be deduced as

$$\varphi_{w_i} = E (\rho_{w_i})$$

$$= \sum_{m \in M} \left[ \frac{\mu_{w_0,m} + \mu_{w_i,m}}{2} - E (\gamma_m) \right]$$

$$= \sum_{m \in M} \left[ \frac{\mu_{w_0,m} + \mu_{w_i,m}}{2} - \mu_{w_0,m} \right]$$

$$= \sum_{m \in M} \frac{\mu_{w_i,m} - \mu_{w_0,m}}{2}. \hspace{1cm} (39)$$

Using assumptions (1) and (3), the variance of $\rho_{w_i}$ can be deduced as

$$\delta^2_{w_i} = D (\rho_{w_i})$$

$$= D \left\{ \sum_{m \in M} \left[ \frac{\mu_{w_0,m} + \mu_{w_i,m}}{2} - \gamma_m \right] \right\}$$

$$= \sum_{m \in M} D (\gamma_m)$$

$$= \sum_{m \in M} \sigma^2_{\chi_m}.$$ \hspace{1cm} (40)

where $\sigma^2_{\chi_m}$ denotes the RSS variance of the $m$th AP in grid $w_0$. Because $\delta^2_{w_i}$ is unrelated to $i$, we denote

$$\delta^2 = \delta^2_{w_i} = \sum_{m \in M} \sigma^2_{\chi_m}. \hspace{1cm} (41)$$

Similarly, if we assume $\mu_{w_0,m} > \mu_{w_1,m}$, $P_{ML(i)}$ can also be expressed as $P_{ML(i)} = \Pr \{ \rho_{w_i} < 0 \}$. However

$$\varphi_{w_i} = \sum_{m \in M} \frac{\mu_{w_0,m} - \mu_{w_i,m}}{2}. \hspace{1cm} (42)$$

Combining (39) and (42), the general expression of $\varphi_w$ can be rewritten as

$$\varphi_{w_i} = \frac{1}{2} | \mu_{w_0} - \mu_{w_i} | = \sum_{m \in M} \frac{\mu_{w_0,m} - \mu_{w_i,m}}{2}. \hspace{1cm} (43)$$

Using assumption (2), $\varphi_{w_i}$ can be deduced as

$$\varphi_{w_i} = \sum_{m \in M} \frac{\alpha_{2,m} (\log_{10} d_{m,w_i} - \log_{10} d_{m,w_0})}{2}$$

$$= \sum_{m \in M} \frac{\alpha_{2,m} \log_{10} d_{m,w_i}}{2}$$

$$= \sum_{m \in M} \alpha_{2,m} \log_{10} \left( 1 + \frac{d_{m,w_i} - d_{m,w_0}}{d_{m,w_0}} \right). \hspace{1cm} (44)$$

where $\alpha_{2,m}$ denotes the propagation parameter of the $m$th AP, $d_{m,w_i}$ denotes the distance between the $m$th AP and position $w_i$, and $d_{m,w_0}$ denotes the distance between the $m$th AP and
where $\beta$ is a constant,

$$\beta = \sum_{m \in M} \frac{\alpha_{2,m} \sin \theta_m \log_{10} e}{2d_{m,w_0}}. \tag{46}$$

Now, $\rho_{w_1}$ subject to a Gaussian distribution with mean $\varphi_{w_1} \beta a$ and variance $\delta^2$ (denoted as $\rho_{w_1} \sim \pi (\beta a, \delta^2)$). Then, $P_{ML}(1)$ can be deduced as

$$P_{ML}(1) = \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi} \delta_w} \exp \left[ -\frac{(t - \beta a)^2}{2\delta^2} \right] dt$$

$$= \frac{1}{2} \text{erfc} \left( \frac{\beta a}{\sqrt{2}\delta} \right) \tag{47}$$

where $\text{erfc}(\cdot)$ is the complementary error function, $\text{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_{x}^{\infty} \exp(-t^2) dt$. In general scenarios, we can obtain $\frac{\beta a}{\sqrt{2}\delta} \gg 1$. Using $\text{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_{x}^{\infty} \exp(-t^2) dt \approx \frac{1}{\sqrt{\pi}x} e^{-x^2}$, thus

$$P_{ML}(1) \approx \frac{\delta}{\sqrt{2\pi} \beta a} \exp \left( -\frac{\beta^2 a^2}{2\delta^2} \right). \tag{48}$$

Similarly, $P_{ML}(2)$ can be deduced as

$$P_{ML}(2) \approx \frac{\delta}{\sqrt{2\pi} \beta (2a)} \exp \left( -\frac{\beta^2 (2a)^2}{2\delta^2} \right) \tag{49}$$

and we obtain

$$\frac{P_{ML}(2)}{P_{ML}(1)} = \frac{1}{2} \exp \left( -\frac{3\beta^2 a^2}{2\delta^2} \right) \ll 1. \tag{50}$$

Furthermore, we can obtain $P_{ML}(1) \gg P_{ML}(2) \gg P_{ML}(3) \gg \cdots$, which means that $P_{ML}(1)$ is the closest lower bound of $P_{ML}$. We call $P_{ML}(1)$ as one step error probability and regard it as the approximate solution of $P_{ML}$.

$$P_{ML} \approx P_{ML}(1) \approx \frac{\delta}{\sqrt{2\pi} \beta a} \exp \left( -\frac{\beta^2 a^2}{2\delta^2} \right). \tag{51}$$

3) Localization Error Probability of Traditional ML With AP Selection: If AP selection is utilized, (51) continues to be valid; however, $\delta^2$ and $\beta$ must be replaced as

$$\delta_{ML+AS}^2 = \sum_{n \in N} \sigma_n^2 \tag{52}$$

$$\beta_{ML+AS} = \sum_{n \in N} \frac{\alpha_{2,n} \sin \theta_n \log_{10} e}{2d_{n,w_0}}. \tag{53}$$

Such that

$$P_{ML+AS} \approx \frac{\delta_{ML+AS}}{\sqrt{2\pi} \beta_{ML+AS} a} \exp \left( -\frac{\beta_{ML+AS}^2 a^2}{2\delta_{ML+AS}^2} \right) \tag{54}$$

where $N$ is the selected best AP set. The principle of the AP selection process is to select the best AP sets with stronger RSS and lower signal fluctuation (which means smaller $d_{m,w_0}$ and $\sigma_n^2$), such that $P_{ML+AS}$ will be smaller than $P_{ML}$. 

4) Localization Error Probability of ACS: For ACS, the online localization process is accomplished by $K$-step searching. If there is a localization error in an arbitrary step, the ACS estimation will be incorrect. The probability of localization error can be expressed as

$$P_{ACS} = 1 - \prod_{k=1}^{K} (1 - P_{ACS-k}) \tag{55}$$

where $P_{ACS-k}$ is the error probability in the $k$th step. $P_{ACS-k}$ can be deduced by a similar manner to $P_{ML}(1)$ as

$$P_{ACS-k} \approx \frac{\delta_{ACS-k}}{\sqrt{2\pi} \beta_{ACS-k} b_k} \exp \left( -\frac{\beta_{ACS-k}^2 b_k^2 a^2}{2\delta_{ACS-k}^2} \right) \tag{56}$$

where

$$\delta_{ACS-k} = \sum_{n \in N_k} \sigma_n^2, \quad \beta_{ACS-k} = \sum_{n \in N_k} \frac{\alpha_{2,n} \sin \theta_n \log_{10} e}{2d_{n,w_0}}$$

$N_k$ denotes the best AP set of the $k$th step, and $b_k$ denotes the cluster size of the $k$th step. We can obtain $b_1 = 2^{K-1}$. Therefore,

$$P_{ACS-k} \approx \frac{\left( \frac{1}{2} \right)^{K-k} \delta_{ACS-k}}{\sqrt{2\pi} \beta_{ACS-k} a} \exp \left( -\frac{4^{K-k} \beta_{ACS-k}^2 b_k^2 a^2}{2\delta_{ACS-k}^2} \right). \tag{57}$$

Using the assumption that $\frac{\delta_{ACS-k}}{\sqrt{2\delta_{ACS-k}}} \gg 1$, we can see

$$\frac{P_{ACS-k}}{P_{ACS-k+1}} = \frac{\exp \left( -\frac{4^{K-k} \beta_{ACS-k}^2 b_k^2 a^2}{2\delta_{ACS-k}^2} \right)}{2 \exp \left( -\frac{4^{K-k-1} \beta_{ACS-k}^2 b_k^2 a^2}{2\delta_{ACS-k}^2} \right)}$$

$$= \frac{1}{2} \exp \left( -\frac{3 \times 4^{K-k+1} \beta_{ACS-k}^2 b_k^2 a^2}{2\delta_{ACS-k}^2} \right) \ll 1. \tag{58}$$

This means the localization error occurs primarily in the $K$th step.
Because $P_{ACS-1} \ll P_{ACS-2} \ll \cdots \ll P_{ACS-K} \ll 1$, $P_{ACS}$ can be approximated as

$$P_{ACS} = 1 - \prod_{k=1}^{K} (1 - P_{ACS-k})$$

$$\approx 1 - \left(1 - \sum_{k=1}^{K} P_{ACS-k}\right)$$

$$= \sum_{k=1}^{K} P_{ACS-k}$$

$$\approx \sum_{k=1}^{K} \frac{(\frac{1}{2})^{K-k} \delta_{ACS-k}}{\sqrt{2\pi} \beta_{ACS-k}} \exp \left(-\frac{4^{K-k} \beta_{ACS-k}^{2} \alpha^{2}}{2 \delta_{ACS-k}^{2}}\right)$$

$$\approx \frac{\delta_{ACS-K}}{\sqrt{2\pi} \beta_{ACS-K}} \exp \left(-\frac{\beta_{ACS-K}^{2} \alpha^{2}}{2 \delta_{ACS-K}^{2}}\right).$$

(D. Comparison of Traditional ML and ACS)

1) AP Selection is not Utilized in Either Traditional ML and ACS: The ratio of $P_{ML}$ and $P_{ACS}$ can be calculated by (51) and (59) as

$$\eta_1 \approx \frac{\delta_{ML+AS}^{2}}{\beta_{ML+AS}} \exp \left(-\frac{\beta_{ML+AS}^{2} \alpha^{2}}{2 \delta_{ML+AS}^{2}}\right).$$

Because AP selection is not utilized, $\beta_{ACS-K} = \beta$ and $\delta_{ACS-K} = \delta$. As a result,

$$\eta_1 \approx 1.$$

This means that when AP selection is not utilized, ACS can obtain approximately the same localization accuracy as ML.

2) AP Reselection is Utilized in Both Traditional ML and ACS: The ratio of $P_{ML+AS}$ and $P_{ACS}$ can be calculated by (54) and (59) as

$$\eta_2 \approx \frac{\delta_{ML+AS}^{2}}{\beta_{ML+AS}} \exp \left(-\frac{\beta_{ML+AS}^{2} \alpha^{2}}{2 \delta_{ML+AS}^{2}}\right)$$

$$= \frac{\beta_{ACS-K}}{\delta_{ACS-K}} \left(\frac{\delta_{ML+AS}}{\beta_{ML+AS}} \exp \left[\frac{\alpha^{2}}{2} \left(\frac{\beta_{ACS-K}^{2} - \beta_{ML+AS}^{2}}{2 \delta_{ML+AS}^{2}}\right)\right]\right)$$

where

$$\frac{\beta_{ACS-K}}{\delta_{ACS-K}} = \sum_{n \in N_{K}} \alpha_{2,n} \sin \theta_{n} \log_{10} e \sqrt{\sum_{n \in N_{K}} \sigma_{n}^{2}}$$

$$\frac{\beta_{ML+AS}}{\delta_{ML+AS}} = \sum_{n \in N} \alpha_{2,n} \sin \theta_{n} \log_{10} e \sqrt{\sum_{n \in N} \sigma_{n}^{2}}.$$

The ACS mechanism selects the best AP set $N_{K}$ within the neighboring grids of actual location $n_{0}$. The traditional method selects the best AP set $N$ within the entire localization area. In general, $N_{K}$ is better than $N$ such that $\frac{\beta_{ACS-K}}{\delta_{ACS-K}} > \frac{\beta_{ML+AS}}{\delta_{ML+AS}}$ and

$$\eta_2 = \frac{P_{ML+AS}}{P_{ACS+AS}} > 1.$$

This means that when AP selection is utilized, ACS can obtain superior localization accuracy compared to ML.

V. PERFORMANCE EVALUATION AND SIMULATION RESULTS

It can be seen that the localization error probability of ACS is constrained by several conditions. Therefore, a complete evaluation of the algorithm performance is difficult to realize. In this section, numerical simulation results under typical conditions are provided.

A. Case 1: AP Selection is not Utilized in Either Traditional ML or ACS

In this case, the simulation parameters were set as follows:

1) The localization area was gridded as $16 \times 16$ equal squares and $K = 4$ (square side $a$ was variable).
2) There were 12 APs uniformly positioned at the edge of the localization area.
3) Each AP had the same transmitting power and signal fluctuation variance, which means that AP selection was unnecessary.
4) The path loss model was set as $48 + 30 \log_{10} d$.

The simulation results of the localization error probability of the 4th level online search (denoted as $P_{ACS-k}$) are presented in Fig. 9. The signal fluctuation standard deviation (denoted as $\sigma_x$) was chosen as 4 dB and the grid size (denoted as $a$) was chosen from 1 to 5 m. It can be seen that:

1) for a certain $a$, $P_{ACS-1} \ll P_{ACS-2} \ll P_{ACS-3} \ll P_{ACS-4}$, which demonstrates that the total localization error was primarily determined by the last level;
2) with the increase in $a$, $P_{ACS-k}$ decreased rapidly.

The simulation results of the localization error probability of ACS and ML without AP selection (denoted as $P_{ACS}$ and $P_{ML}$, respectively) are displayed in Fig. 10. $\sigma_x$ was chosen as 2, 4, 6 dB and $a$ was chosen from 1 to 5 m. It can be seen that:

1) with an increase in $a$, both $P_{ACS}$ and $P_{ML}$ decreased rapidly;
2) $P_{ACS}$ was close to $P_{ML}$, especially when $a$ was large or $\sigma_x$ was small.
3) for over 90% of data points, the localization error was within the range of 1.2, 2.3, and 3.4 m based on $\sigma = 2$, 4, and 6 dB.

B. Case 2: AP Selection is Utilized in Both Traditional ML and ACS

In this case, the simulation parameters were set as follows.
1) The localization area was gridded as $16 \times 16$ equal squares and $K = 4$ (square side $a$ was variable).
2) There were 12 APs uniformly positioned at the edge of the localization area.
3) Each AP had the same transmitting power; however, different signal fluctuation variances were included, which means that AP selection was meaningful.
4) The AP’s signal fluctuation variance $\sigma_m$ followed a uniform distribution. The mean of $\sigma_m$ (denoted as $\bar{\sigma}_m$) was chosen as the same as $\sigma$ in case 1.
5) The path loss model was set as $48 + 30\log_{10}d$.

The simulation results of localization error probability of ACS and ML with AP selection (denoted as $P_{ACS+AS}$ and $P_{ML+AS}$, respectively) are displayed in Fig. 11. $\bar{\sigma}_m$ was chosen as 2, 4, and 6 dB and $a$ was chosen from 1 to 5 m. It can be seen that:
1) $P_{ACS+AS}$ was significantly smaller than $P_{ML+AS}$, especially when $a$ or $\bar{\sigma}_m$ was large;
2) for over 90% of data points, the localization error of ACS + AS was within the range of 1.1, 2.1, and 3.1 m based on $\bar{\sigma}_m = 2$, 4, and 6 dB. For the traditional ML + AS method, the results were 1.4, 2.8, and 4.5 m, respectively, i.e., the localization accuracy was increased by 0.3, 0.7, and 1.4 m, respectively.

We suggest that this performance improvement is a result of the AP reselection. Because the location estimation in the ACS method proceeded by step and the best AP set was replaced for each step, the ultimate location estimation was determined from the best AP set according to the actual location $w_0$. The traditional ML estimation was selected from the best AP set based on the entire localization area. Clearly, the former can provide superior localization accuracy. When localization is executed in a large area or under large wireless signal fluctuation, the performance gain will be considerable.

VI. EXPERIMENTAL RESULTS

The experiment was conducted on the bottom floor in Xidan Joy-City, a shopping mall in Beijing (see Fig. 12). The localization area was approximately $40\,000 \text{ m}^2$ ($200 \times 200 \text{ m}$). We employed an Android-based Lenovo smart phone S920 as the test mobile terminal and developed an application equipped with two localization algorithms: 1) traditional ML + AP selection and 2) ACS + AP selection.

In the offline phase, RSSs were measured 100 times to calculate the offline FP in each grid. In the online phase, the mobile terminal was randomly positioned within the localization area and the RSS was measured 10 times. The average was used to calculate the online FP. In our experiment, there were 203 APs detected ($M = 203$) and we chose $N = 20$. The localization area was gridded at different sizes for different localization accuracy requirements (see Table I). The actual offline storage compression ratio $\xi_1$ and online computational complexity compression ratio $\xi_2$ are given in Table I. It can be seen that ACS can significantly reduce the offline storage and online computational complexity, especially when the number of grids is large.

Given $a = 1.56 \text{ m}$, the cumulative distribution function (cdf) curves for the localization error of ACS + AS and traditional
TABLE I
ξ₁ AND ξ₂ UNDER DIFFERENT GRID SIZES

<table>
<thead>
<tr>
<th>Grid size</th>
<th>Total grids</th>
<th>a (m)</th>
<th>ξ₁</th>
<th>ξ₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large</td>
<td>32 × 32</td>
<td>6.25</td>
<td>0.19</td>
<td>2.0 × 10⁻³</td>
</tr>
<tr>
<td>Middle</td>
<td>64 × 64</td>
<td>3.12</td>
<td>0.15</td>
<td>5.9 × 10⁻¹</td>
</tr>
<tr>
<td>Small</td>
<td>128 × 128</td>
<td>1.56</td>
<td>0.09</td>
<td>1.7 × 10⁻³</td>
</tr>
</tbody>
</table>

Fig. 13. CDF curve (a = 1.56 m).

Fig. 14. CDF curve (a = 6.25 m).

ML + AS are presented in Fig. 13. For over 90% of the data points, the localization error of ACS + AS falls within the range of 2.1 m; the result for the traditional ML + AS is 5.8 m. In this case, ACS + AS improved the localization accuracy by 3.7 m.

Given a = 6.25 m, the cdf curves for the localization error of ACS + AS and traditional ML + AS are displayed in Fig. 14. For over 90% of the data points, the localization error of ACS + AS falls within the range of 5.1 m. The result for the traditional ML + AS is 6.3 m. In this case, ACS + AS improved the localization accuracy by 1.2 m.

More conclusion can be found by comparing the results presented in Figs. 13 and 14.

1) The performance gain of ACS + AS is more notable when a is small.
2) For ML + AS, the localization result of a = 1.56 m is somewhat better than that of a = 6.25 m (improved by 0.5 m). We suggest that there are two reasons for this.

First, in this experiment, the majority of the APs have a strong signal fluctuation. The limitation of the distinction capability of these APs is approximately 6 m. Therefore, simply reducing a cannot improve the localization performance significantly. Second, there are some APs that have a stronger distinction capability than 6 m. However, the traditional ML + AS method is not capable of determining these APs. Therefore, for ML + AS, setting the grid side a = 6 m is sufficient.

3) For ACS + AS, the localization result for a = 1.56 m is significantly better than that of a = 6.25 m (improved by 3.0 m). We suggest that the performance gain can be primarily attributed to the fact that ACS + AS is capable of determining the best AP set to accomplish the localization process. For ACS + AS, setting the grid side a = 1.56 m is meaningful.

VII. CONCLUSION

Localization is an appealing application and has become increasingly common in our daily life. RSS-based schemes have been widely used to provide location-aware services in WLANs. However, in realistic large area localization, numerous APs exist providing extensive amounts of location information. There are three problems with this scenario: 1) excessive memory requirements for the offline FPs; 2) high computational complexity for the online localization process; and 3) because signal strength and propagation fluctuation are meaningfully different among different APs and locations, how to determine the AP sets with best distinction capability remains a consideration.

In this paper, we proposed a novel localization mechanism. In the offline phase, we suggested that the FP storing density should consider the AP’s distinction capability. Then, we proposed a new FP generation algorithm called ACS. From the analysis results of (31) and (32), we can see that the offline storage can be significantly reduced using ACS.

Then, in the online phase, the localization process was modeled as a searching problem over Quadtree or Octree. From the analysis results using (34), we can see that the online localization computation can be reduced from \(O(I^2)\) to \(O(\log I)\), which is a significant reduction.

Finally, the traditional localization algorithm only stores the RSS as the FP. In ACS, a unique AP set with best distinction capability for a given position is also stored as the FP. This personalized best AP set configuration can improve the localization accuracy significantly, especially in a large area. The conclusion made in this paper was theoretically deduced in Section IV, verified by simulation results in Section V, and demonstrated by experimental results in a realistic environment in Section VI.

Future research in the new and largely open area of wireless technologies should be conducted along the following directions. First, we can leverage the available multiple APs to improve the location accuracy. Second, in this paper, we only considered the RSS to be the FP. However, considering channel state information, packet loss rate, and other factors to be FPs merits attention in future research [33].
REFERENCES


