Article

Coordinate-Based Clustering Method for Indoor Fingerprinting Localization in Dense Cluttered Environments

Wen Liu, Xiao Fu * and Zhongliang Deng

School of Electronic Engineering, Beijing University of Posts and Telecommunications, No. 10 Xitucheng Road, Haidian District, Beijing 100876, China; liuwen@bupt.edu.cn (W.L.); dengzhl@bupt.edu.cn (Z.D.)
* Correspondence: Xiaofu@bupt.edu.cn; Tel.: +86-130-5139-0935

Abstract: Indoor positioning technologies has boomed recently because of the growing commercial interest in indoor location-based service (ILBS). Due to the absence of satellite signal in Global Navigation Satellite System (GNSS), various technologies have been proposed for indoor applications. Among them, Wi-Fi fingerprinting has been attracting much interest from researchers because of its pervasive deployment, flexibility and robustness to dense cluttered indoor environments. One challenge, however, is the deployment of Access Points (AP), which would bring a significant influence on the system positioning accuracy. This paper concentrates on WLAN based fingerprinting indoor location by analyzing the AP deployment influence, and studying the advantages of coordinate-based clustering compared to traditional RSS-based clustering. A coordinate-based clustering method for indoor fingerprinting location, named Smallest-Enclosing-Circle-based (SEC), is then proposed aiming at reducing the positioning error lying in the AP deployment and improving robustness to dense cluttered environments. All measurements are conducted in indoor public areas, such as the National Center For the Performing Arts (as Test-bed 1) and the XiDan Joy City (Floors 1 and 2, as Test-bed 2), and results show that SEC clustering algorithm can improve system positioning accuracy by about 32.7% for Test-bed 1, 71.7% for Test-bed 2 Floor 1 and 73.7% for Test-bed 2 Floor 2 compared with traditional RSS-based clustering algorithms such as K-means.

Keywords: indoor positioning; fingerprinting; access point deployment; clustering algorithm; K-means; coordinate based clustering; smallest enclosing circle

---

1. Introduction

Indoor location-based service (ILBS) has gained considerable attention in recent years due to its social and commercial values, with market value predicted to worth US $10 billion by 2020 [1]. Meanwhile, the demand for accurate localization in indoor environments, such as large public places, office buildings with mass obstacles and military facilities, has increased dramatically [2,3]. Due to the coverage limitation of satellite signals, which are widely used in Global Navigation Satellite Systems (GNSS), various indoor positioning technologies, consisting of Infrared, Ultrasonic, Wireless Local Area Network (WLAN), Bluetooth, ZigBee, Radio Frequency Identification Devices (RFID), Pseudolite and Ultra Wideband (UWB), have been proposed to provide better performance in indoor localization. Infrared and Ultrasonic positioning technology could achieve centimeter-level accuracy in Line-Of-Sight environments [4,5], however, these two technologies both need the implementation of dense access points (AP) and the obstacles in indoor environments would have a considerable influence on the accuracy and robustness of the system. Pseudolite and UWB positioning technology can perform a centimeter-level accuracy at the cost of expensive special devices and high system

complexity [6,7]. Among WLAN, Bluetooth, ZigBee and RFID positioning technologies, WLAN has aroused researchers' interests because of its wide implementation, high mobility, low networking cost and high compatibility in dense cluttered indoor environments [8–11].

Although indoor positioning systems based on WLAN have advantages over the systems based on other technologies in dense cluttered environments, there are still some factors that can influence the systems' performance in positioning. Positioning accuracy, robustness, computational burden, energy consumption, and the cost of the system implementation are the main factors which should to be considered when evaluating the performance of the whole system. Gu, Y. et al. [12] have given an overview and comparison on the existing WLAN-based indoor positioning systems in terms of positioning accuracy, robustness and cost. Bisio, I. et al. [13] have given a performance comparison of a probabilistic fingerprint-based indoor positioning utilizing two smartphones in terms of positioning accuracy, robustness and consistency, and has achieved relatively high accuracy about 1.2 m in a real test environment. Bisio, I. et al. [14] have proposed a novel probability computation method for indoor WiFi-based positioning to reduce the system's computational burden, and has achieved a significant 90% decrease on energy/time consumption with mobile devices. Related works concerning the evaluation of the systems' accuracy, computational burden and energy efficiency have also been conducted in [15,16], and these works have given important guidance for researchers on deploying fingerprint-based positioning systems. This paper mainly considered the positioning accuracy and robustness when evaluating indoor fingerprint-based positioning systems, and has given focus on the influence of AP deployment.

Indoor positioning technology based on WLAN can be divided into two different schemes: time delay-based and radio map-based. Time delay-based methods, such as Time of Arrival (TOA), Time Difference of Arrival (TDOA) and Round Trip Time (RTT), could achieve relatively higher accuracy in a relatively clean indoor environment, compared to radio map-based method. However, time delay-based methods require network synchronization and the delay should be measured exactly, which brings the localization system more complexity and is hard to realize in dense multipath and NLOS environments. Radio map-based methods, for example Received Signal Strength (RSS) based method, utilize matching algorithms to locate unknown points (UPs). Although, to a great extent, the RSS database of radio map-based methods depends on environment, these methods can achieve a relatively robust and well positioning accuracy in dense cluttered environments once the database is established.

Traditionally, RSS-based positioning methods have two stages: off-line stage and on-line stage. In the first stage, which can also be called training stage, the data collection for training is performed semi-automatically or automatically. The collected data can also be called fingerprints, which contain information representing the indoor environment. The database, also named radio map, is built based on these fingerprints. A full database should contain the coordinates (x, y, z) of each fingerprint, together with the Medium Access Control (MAC) addresses for each hearable AP and the corresponding RSS [17]. In the second stage, also called estimation stage, the localization of UPs is estimated by matching the RSS between UPs and fingerprints in database utilizing matching algorithms, e.g., K Nearest Neighbor (KNN). The on-line matching stage can proceed both on the server and the mobile devices, and it demands real-time signal processing. When the accuracy demand is high but the positioning area is huge, the database may be so large that the computational complexity of the matching algorithm is very high, which will bring more resource consumption and high power dissipation. Therefore, in order to reduce the computational complexity of the whole system, clustering methods are always necessary in the off-line training stage [18,19].

The deployment of APs in WLAN based positioning system has a significant influence on the accuracy of the system for most indoor environments, especially for large-scale indoor environments. In [20], an AP planning method for indoor positioning system has been proposed, together with the GDOP evaluation, in order to decrease the position error. In [21], a map-assisted AP placement algorithm has been proposed, and an accuracy within 3 m of 71% has been achieved. In [22],

AP topology has been evaluated and optimized utilizing Cramér-Rao Lower Bound calculation method. The deploying schemes in [20–22] have an instructional significance to the implementation of APs when a new indoor WLAN based positioning system is set up. However, the APs' deployments in most positioning conditions are generally fixed together with the construction of these infrastructures. These deployments are typically optimized for communication scope, but not for localization. This paper focused on the conditions above, analyzed the probable deployment influence on system's accuracy and proposed a novel coordinate-based clustering method to reduce positioning error instead of changing the existed AP deployments.

We had previously proposed a smallest-enclosing-circle-based clustering algorithm for indoor positioning in [23]. This paper is an extension of the work in [23], and, in this paper, the AP implementation influence is detailed analyzed. In particular, when compared to [23], in this paper, the smallest-enclosing-circle-based clustering algorithm is described and derived in detail, and new experiments are conducted in different environments to demonstrate the effectiveness of the method proposed.

2. AP Deployment Influence Analysis

The deployment of APs in indoor WLAN based positioning system has a significant influence on the positioning accuracy. This section analyzes the influence brought by the implementation of APs and gives guidance about error reduction without changing existed AP deployments.

2.1. Indoor Wireless Signal Transmission Model

Path-loss transmission model is a theoretical mathematical model, which has been widely used to represent the transmission principle of wireless signals, such as WiFi, Bluetooth, et al., in indoor environments. As shown in Equation (1), the path-loss transmission model indicates the relationship between the transmission loss and distance of wireless signals, which assumes Line-Of-Sight (LOS) along the signal transmission route in most cases. Aiming to analyzing the influence of AP deployment in this section, this paper makes the same assumption that the transmission route between APs and Reference Points (RPs) is LOS or has little obstacles. We make the assumption for the reason that the APs in most of our experimental environments are all on the ceilings and the RPs are all on the floors. In order to make our analysis comprehensive but not complicated, we utilize the path-loss model as shown in Equation (1) with the assumption elaborated above.

$$PL_d = PL_0 + 10n\log\frac{d}{d_0} + X_\sigma, \tag{1}$$

where PL_d is the path-loss in dBm of the wireless signals after transferred a distance of d. PL_0 is the path-loss after a distance of d_0 and n is an environment-depended constant which represents the loss coefficient of the indoor environment. X_σ is a normally distributed variable representing the shadowing error and its distribution function is:

$$X_\sigma \sim N\left(0, \sigma^2\right), \tag{2}$$

$$f\left(X_\sigma\right) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{X_\sigma^2}{2\sigma^2}}, \tag{3}$$

This model shows the path-loss of wireless signals in indoor environments is of direct proportion to the logarithm of transmission distance. In general, the RSS in the fingerprint database is the received signal strength of wireless signal, which can be shown as follows:

$$RSS = P + PL_d, \tag{4}$$

where P is the transmitted power of AP. Without loss of generality, we can assume that d_0 is 1 m, and it can be derived from Equations (1) and (4) that:

$$RSS = A + 10n\log d + X_\sigma, \tag{5}$$

$$A = P + PL_0, \tag{6}$$

It can be concluded from Equation (5) that RSS is also of direct proportion to the logarithm of distance, like path-loss. Therefore, if the RSS of two reference points (RPs) is similar with each other, their distances to one same AP are also similar, which can be shown as follows:

$$RSS \propto \log d, \tag{7}$$

$$\Delta RSS = 10n\left(\log d_{11} - \log d_{12}\right) = 10n\log\frac{d_{11}}{d_{12}}, \tag{8}$$

$$\Delta RSS \to 0 \Leftrightarrow d_{11} \approx d_{12}, \tag{9}$$

where d_{11} is the distance between AP_1 and RP_1, and d_{12} is the distance between AP_1 and RP_2.

2.2. Database Structure for Indoor Positioning System

Indoor RSS based positioning system locates UPs by matching their RSS to the RSS stored in the database, and calculating their positions through matching algorithm. The database is established in the off-line training stage, when the reference points' coordinates and RSS from APs with different MACs are measured and stored. The structure of the database generally used in positioning system can be shown as follows:

$$\begin{bmatrix} x_1 & y_1 & RSS_{11} & RSS_{12} & \cdots & RSS_{1p} \\ x_2 & y_2 & RSS_{21} & RSS_{22} & \cdots & RSS_{2p} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x_N & y_N & RSS_{N1} & RSS_{N2} & \cdots & RSS_{Np} \end{bmatrix}$$

where N is the number of RPs, and p is the number of APs. RSS_{ij} is the received signal strength at i-th RP from j-th AP.

Matching algorithms calculate the location of UPs by matching their RSS to the RSSs stored in the database. Traditional matching algorithms, like KNN and Weighted-KNN (WKNN), utilize Euclidean Distance between UPs and RPs as the matching principle, which can be calculated as follows:

$$dist_{Eu}\left(UP, RP_1\right) = \sqrt{\left(RSS_{01} - RSS_{11}\right)^2 + \left(RSS_{02} - RSS_{12}\right)^2 + \cdots \left(RSS_{0p} - RSS_{1p}\right)^2}, \tag{10}$$

$$dist_{Eu}\left(UP, RP_k\right) = \sqrt{\left(RSS_{01} - RSS_{k1}\right)^2 + \left(RSS_{02} - RSS_{k2}\right)^2 + \cdots \left(RSS_{0p} - RSS_{kp}\right)^2}, \tag{11}$$

$$dist_{Eu}\left(UP, RP_k\right) = \sqrt{\sum_{i=1}^{p}\left(RSS_{0i} - RSS_{ki}\right)^2}, \tag{12}$$

where $dist_{Eu}(UP, RP_k)$ is the Euclidean Distance between UP and k-th RP. KNN matching algorithm calculates the Euclidean distance between UP and RPs, and selects k RPs with minimum distance and chooses the barycenter as the location of UP.

2.3. AP Deployment Influence

In most indoor environments covered with WiFi signals, the deployment of APs (in other words, the wireless routers) is already constructed and fixed. These deployments are typically optimized

for communication scope, but not for localization. Therefore, the APs' implementation is arranged concentrated and well-organized, generally in several straight lines or Z-shaped lines above on the ceiling. This paper focuses on the situation that two RPs in the database have similar RSS and analyzed the probable reasons of this situation. This phenomenon can be divided into two different situations to analyze: the APs are close to the RPs or the APs are far from the RPs. The following subsections aim to analyze these situations.

2.3.1. APs Are Close to RPs

When the APs above on the ceiling are close to the RPs down on the floor, there is one probable situation that can lead to the phenomenon above, which can be described as follows in Figure 1.
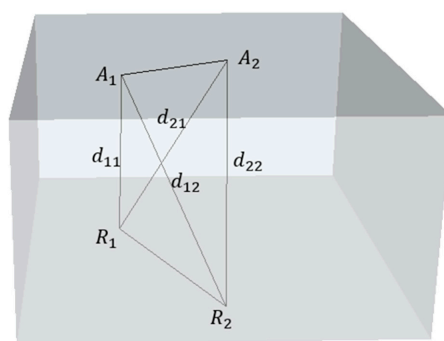
Figure 1. Access Point (AP) Deployment—Case 1.

where A_1 and A_2 are the two APs above on the ceiling, and R_1 and R_2 are the two RPs in the database. d_{11}, d_{12}, d_{21} and d_{22} are the distances between APs and RPs. When the line A_1A_2 is orthogonal to the line R_1R_2 and it is also the bisector of the line R_1R_2 in physical space, as shown in the figure, the distance between A_1 and R_1 is similar with the distance between A_1 and R_2 according to the mid-perpendicular theorem:

$$d_{11} \approx d_{12}, \tag{13}$$

Especially, d_{11} is equal to d_{12} when the line A_1A_2 is exactly the mid-perpendicular of line R_1R_2. Thus, the RSSs of R_1 and R_2 in the database are similar to each other for these APs according to Equation (9).

2.3.2. APs Are Distant from RPs

Instead of the APs close to the RPs, there are some APs deployed far from the RPs, which can be shown as follows in Figure 2.
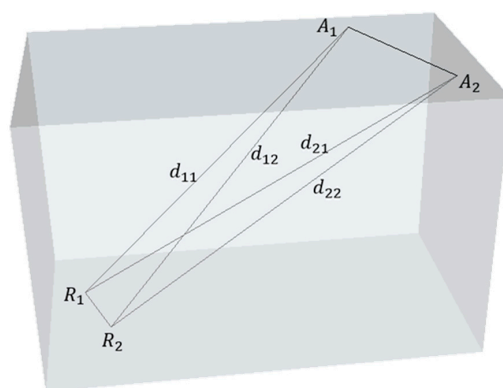

Figure 2. AP Deployment—Case Two.

where the definition of symbols are the same as the above subsection. Compared to the APs close to R_1 and R_2, A_1 and A_2 in this figure are distant from R_1 and R_2. The reason why the value of RSS is similar in this situation can be discussed as follows. Firstly, a two-dimensional model is extracted from the three-dimensional model, which can be shown in Figure 3.
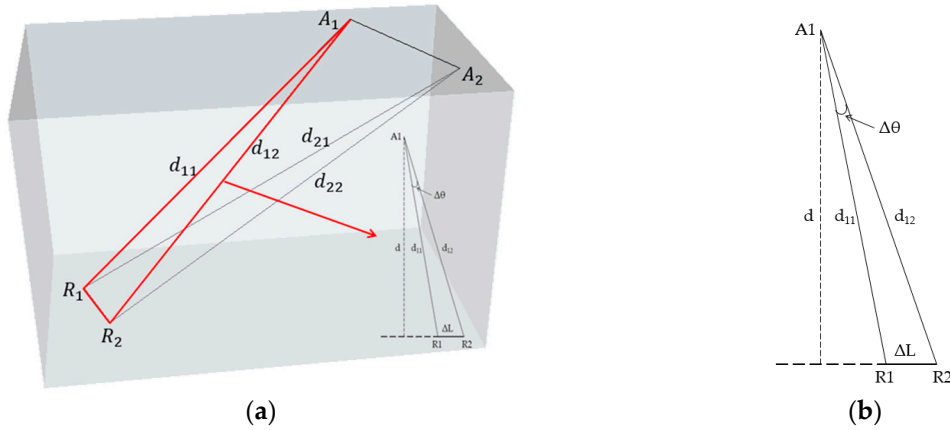

(a)　　　　**(b)**

Figure 3. (**a**) AP deployment in Case 2 in three-dimension model; and (**b**) two-dimensional model extracted from three-dimensional model for analysis.

It can be derived according to the trigonometric function theorem as follows:

$$d_{11} = \frac{d}{\cos\theta_1} \quad d_{12} = \frac{d}{\cos(\theta_1 + \Delta\theta)}, \tag{14}$$

$$\Delta L = d_{12}\sin(\theta_1 + \Delta\theta) - d_{11}\sin\theta_1 = d(\tan(\theta_1 + \Delta\theta) - \tan\theta_1), \tag{15}$$

$$\Delta d = d_{12} - d_{11} = \frac{d}{\cos(\theta_1 + \Delta\theta)} - \frac{d}{\cos\theta_1}, \tag{16}$$

$$\begin{aligned} \frac{\Delta L}{d} &= (\tan(\theta_1 + \Delta\theta) - \tan(\theta_1)) \\ &= \frac{\sin(\theta_1 + \Delta\theta)\cos(\theta_1) - \cos(\theta_1 + \Delta\theta)\sin(\theta_1)}{\cos(\theta_1 + \Delta\theta)\cos(\theta_1)} \\ &= \frac{\sin(\Delta\theta)}{\cos(\theta_1 + \Delta\theta)\cos(\theta_1)} \end{aligned}, \tag{17}$$

$$\frac{\Delta L}{d_{12}} = \frac{\sin(\Delta\theta)}{\cos(\theta_1)}, \tag{18}$$

$$\frac{\Delta L}{d_{11}} = \frac{\sin(\Delta\theta)}{\cos(\theta_1 + \Delta\theta)}, \tag{19}$$

When the APs are distant from RPs, in other words,

$$\Delta L \ll d_{11} \quad and \quad \Delta L \ll d_{12}, \tag{20}$$

$$\frac{\Delta L}{d_{11}} = \frac{\sin(\Delta\theta)}{\cos(\theta_1 + \Delta\theta)} \to 0, \tag{21}$$

$$\frac{\Delta L}{d_{12}} = \frac{\sin(\Delta\theta)}{\cos(\theta_1)} \to 0, \tag{22}$$

$$\frac{\Delta L}{d_{11}} = \frac{\sin\left(\Delta\theta\right)}{\cos\left(\theta_1 + \Delta\theta\right)} \approx \frac{\sin\left(\Delta\theta\right)}{\cos\left(\theta_1\right)} = \frac{\Delta L}{d_{12}}, \tag{23}$$

then $\Delta\theta$ is very small and, in this situation, the distance d_{11} is similar with d_{12}. Because the values of RSSs in the database are of direct proportion to the logarithm of the distance, the variation of RSS is slower than the distance. Especially when the distance between access points and reference points is very large, the variation of the RSS becomes very small. Therefore, the RSSs of RPs in the database are approximately equal according to Equation (9).

2.3.3. Influence Analysis

For the reasons analyzed in previous subsections, the phenomenon that two RPs have similar RSSs in the database generally exists in indoor fingerprint positioning system. This paper named this phenomenon Singular RSS Phenomenon (SRP) to describe the existence of two RSS-similar RPs. Meanwhile, in the off-line training stage of fingerprinting positioning, clustering algorithms are conducted to achieve sparsification of database in order to reduce the system's computation complexity of the on-line estimation stage. Traditional clustering algorithms utilized in training stage are based on the RSS in database, in other words, the algorithms cluster fingerprints into several clusters based on the comparison of their RSSs stored in database. In such conditions that SRP exists, traditional clustering algorithms would cluster these two RPs into the same cluster. In on-line estimation stage, matching algorithms are utilized to calculate the position of UPs. Matching algorithms first compare RSS between UP and cluster-heads to match the UP into one cluster, then conduct the comparison in the cluster to find the position of UP. If UP is matched into the cluster where SRP exists, the positioning results would hop between these two RPs in SRP, which brings a large bias in positioning accuracy.

3. Coordinate-Based Clustering

Since the deployment of APs in Wi-Fi-covered indoor environments has been fixed, and traditional clustering algorithms are not robust enough for indoor positioning, this paper proposed a novel clustering method based on the coordinates of RPs together with their RSS. Coordinate-based clustering methods can be divided into two schemes generally: the grid-based clustering and grid-free clustering.

3.1. Grid-Based Clustering

Grid-based clustering method has been proposed in [24], and has achieved a relatively high accuracy and robustness in indoor fingerprinting positioning. The clustering algorithms based on meshing first divide experimental environments into several grids with the same size. Then, the RPs in database are clustered into the closest grid based on the distance principle. These grid-based clustering is one of the coordinate-based clustering, and can be more robust than RSS-based clustering in environments where SRP exists. However, one challenge of the clustering algorithms based on meshing is that the size of grid should be specified before clustering and is always fixed. Therefore, grid-based clustering may not be flexible in different indoor dense cluttered environments.

3.2. Smallest-Enclosing-Circle-Based Clustering

Smallest enclosing circle (SEC) algorithm has been widely investigated in information security field [25], such as data searching and information filtering, whereas it has not been introduced into indoor fingerprint positioning system for clustering to date. This subsection analyzes the method of SEC and proposes the coordinates-based clustering algorithm, which introduces the SEC algorithm into database clustering.

3.2.1. Smallest Enclosing Circle Model

Smallest enclosing circle of a points set can be described as follows:

$$B\left(c,r\right) := \{x : \|x - c\| \leq r; x \in \Omega; r = r_{\min}\}, \tag{24}$$

where $B(c, r)$ is the circle with minimum radius that consists of all the points in the given set, and x and c represents the point x and the center of circle B, respectively. In general, the center of circle is assumed O_m with coordinates (x_m, y_m) and radius R_m. The lines connect the center of circle and the points $k_1, k_2, k_3, \ldots, k_n$ inside of the circle intersect the circle with point P_i ($i = 1, 2, \ldots, n$). The SEC problem is to estimate the (x_m, y_m) and R_m that minimize the objective function, given by Equation (10).

$$f\left(d\right) = \sum_{i=1}^{n} |O_m P_i - O_m k_i|, \tag{25}$$

where O_m is the center of circle, and the definitions of P_i and k_i have been given above.

3.2.2. Smallest Enclosing Circle Clustering Algorithm

SEC clustering algorithm, as one of the grid-free clustering algorithms, is proposed to compensate the error introduced by conventional clustering algorithms such as K-means. This algorithm clusters the reference points in fingerprint database into several clusters (this paper takes k to represent the number of clusters) based on their coordinates in physical space, instead of their RSS values. This can mitigate the positioning area deviations in dense cluttered indoor environments.

The number of clusters k in SEC clustering algorithm is selected equal to or less than the square root of RPs' number, which is the same selection principle as K-means clustering algorithm in [26]. In general, a well clustering method should follow the principle that the distances among cluster-heads are larger than the distances between RPs and their cluster-heads in its clustering results. This paper takes the mean distance of inter-cluster and intra-cluster to elaborate the selection principle of number of clusters in SEC clustering algorithm. This paper takes the symbol L and D to represent the inter-cluster and intra-cluster distance, respectively, which can be shown as follows:

$$L = \sum_{i=1}^{k} \sum_{j=1}^{k} |O_i O_j| \quad (i \neq j), \tag{26}$$

$$D = \sum_{i=1}^{k} \sum_{P \in C_i} |PO_i|, \tag{27}$$

where k is the number of clusters, O_i represents the cluster-head in the i-th cluster C_i and P represents the RPs in cluster C_i. Assuming that there are n RPs in fingerprint database, then the mean inter-cluster and intra-cluster distance can be defined as follows:

$$\bar{l} = L/k, \tag{28}$$

$$\bar{d} = D/n, \tag{29}$$

In order to obtain relatively well clustering results, the intra-structure of one cluster is similar with the inter-structure of clusters, and the inter-cluster distance is larger than the intra-cluster distance [26], which can be shown as follows:

$$\frac{\bar{d}}{D/k} \leq \frac{\bar{l}}{L}, \tag{30}$$

It can be derived from Equations (28)–(30) that the selection principle of k in SEC clustering algorithm can be shown as Equation (31).

$$k \leq \lfloor \sqrt{n} \rfloor,\tag{31}$$

The experimental evaluation of the influence of different numbers of clusters on the positioning system will be further discussed in Section 4, and without loss of generality, this paper selects the number of clusters k equal to the square root of RPs' number n. The implement of the SEC algorithm can be given as follows.

A. Symbol Definition:

The definition of symbols used in SEC clustering algorithm are as follows in Table 1.

Table 1. Symbol definition in Smallest Enclosing Circle (SEC) clustering algorithm.

Symbol	Definition	Mathematical Operators	Definition
$P = \{P_1, P_2, \cdots, P_n\}$	RPs in the database, p_k is the k-th RP	$\lfloor A \rfloor$	Integer notation, represents the maximum integer less than A.
$O = \{O_1, O_2, \cdots, O_k\}$	Circle centers in SEC, O_k is the center of k-th circle	$\forall$	Arbitrary notation, represents the selection is random.
k [1]	Number of clusters, selected as a constant generally	*min*	Minimum selection notation.
$R = \{r_1, r_2, \cdots, r_k\}$	Radiuses of SECs, r_k is the radius of k-th circle	*mid(A, B)*	Midpoint calculation, represents the midpoint of the line AB.
$C = \{C_1, C_2, \cdots, C_k\}$	SECs, C_k is the k-th circle	*dist*	Distance calculation, represents the Euclidean distance in this paper.

[1] The selection of clusters' number k depends on the number of RPs in the database. In this paper, k is selected as a constant and can be shown as Equation (32), where *Card* function is to calculate the number of elements in set P, in others words, k equals to the positive square root of RP numbers.

$$k = \left\lfloor \sqrt{Card\,(P)} \right\rfloor\tag{32}$$

B. Algorithm Process:

The proposed clustering algorithm based on smallest enclosing circle consists of four steps, which can be processed as follows:

Step 1. Select k points from set P randomly for the initialization of smallest enclosing circle-based algorithm.

Step 2. Calculate the Euclidean distances between remaining points and the selected points in database. Cluster the remaining points into k smallest enclosing circles based on the minimum-distance principle, and update the centers and radiuses of circles based on the theory of SEC.

Step 3. Evaluate the changes of circle centers before and after SEC algorithm, if the center changes are larger than the given threshold, then do Step 2; else do Step 4.

Step 4. Store the centers' coordinates and RSS information into database for on-line evaluation stage.

The pseudo code of the clustering process is as follows in Algorithm 1.

Algorithm 1: Smallest Enclosing Circle (SEC) Clustering Algorithm

---

Initialize:

$$\forall P_k = \{p_1, p_2, \cdots, p_k\} \subseteq P$$
$$for\ (i = 1, i \le k, i++)$$
$$\{O_i = p_i; r_i \le \min\left(dist\left(p_a, p_b\right)\right);$$
$$(p_a, p_b \in P; a \ne b); C_i = (O_i, r_i);\}$$

Calculation:

$$for\ (i = 1, i \le k, i++)$$
$$\{for\ (j = k+1, j \le n, j++)$$
$$\{D_{ij} = dist\left(p_i, p_j\right); D_{im} = \min_j\left(D_{ij}\right);$$
$$p_m \in C_i;\}$$
$$if\ (dist\left(p_m, O_i\right) \le r_i)$$
$$\left\{r_i' = r_i; O_i' = O_i; C_i' = C_i \triangleq \{(O_i, r_i)\};\right\}$$
$$else\ \{O_i' = mid\left(O_i, p_m\right);$$
$$r_i' = r_i + \tfrac{1}{2}dist\left(O_i, p_m\right); C_i' = \left\{\left(O_i', r_i'\right)\right\};\}$$

Verification:

$$if\ \left(dist\left(O_i', O_i\right) \le \delta\right) break;$$
$$else\ \{O = O'; R = R'; C = C';$$
$$repeat\ step\ 2\ until\ \left(dist\left(O_i', O_i\right) \le \delta\right);\}$$

---

4. Measurement Analysis

4.1. Experiments Scenario and Implementation

We conducted our experiments in two dense cluttered environments, one is the public area at the National Center For the Performing Arts (denoted as Test-bed 1), and the other is at XiDan Joy City, which is a multi-story shopping mall in Beijing (denoted as Test-bed 2).

Figure 4 shows the planar graph and the deployment of RPs in Test-bed 1 with a plotting scale of 1:2800. The area marked as green, with an area of 210 m by 140 m, is the public area where our experiment is conducted. This area is covered with IEEE 802.11b/g wireless signals from 58 wireless routers with unique MAC addresses. The deployment of these wireless routers (called Access Points) is ordered as Z-shaped lines. The points marked as black in Figure 4 represent the reference points stored in the fingerprint database, with an interval of 5 m. The number of points for training stage is 260, and the RSS information stored at one point is an average value of 100 times measurements. Hence, 26,000 pieces of RSS data are collected for calculation of average value, and 260 pieces of RSS data are stored in database finally.

Figure 5 shows the reference points' implementation of Test-bed 2 at XiDan Joy City. XiDan Joy City is a multi-story shopping mall with 12 floors (eight floors above-ground and four floors under-ground) at XiDan, Beijing. The experimental environment for Test-bed 2 is with an area of 120 m by 90 m, which is covered with IEEE 802.11b/g wireless signals. There are 4949 reference points in this experiment for all the 12 floors and 10 pieces of RSS data are collected at each point utilizing a cellphone, which runs our collecting program at RPs to collect the RSSI from different APs. The RSS data stored in the database for each point is the average of these 10 pieces of data, which are collected from four different orientations. The distribution of the total 4949 RPs at 12 floors and the number of hearable APs at each floor is shown in Table 2. We have conducted our experiments at each floor in XiDan Joy City. However, in order to validate the effectiveness of SEC algorithm analyzed above well, we selected floor-one and floor-two as the test-bed to discuss our clustering results, in consideration of the number of RPs or APs in each floor and the AP deployment in each floor. The clustering results and positioning error comparisons in other floors are also shown in this paper at the Appendix A.

Figure 5 shows the experimental scenarios of floor-one and floor-two, in which floors the APs were deployed along as several straight lines.
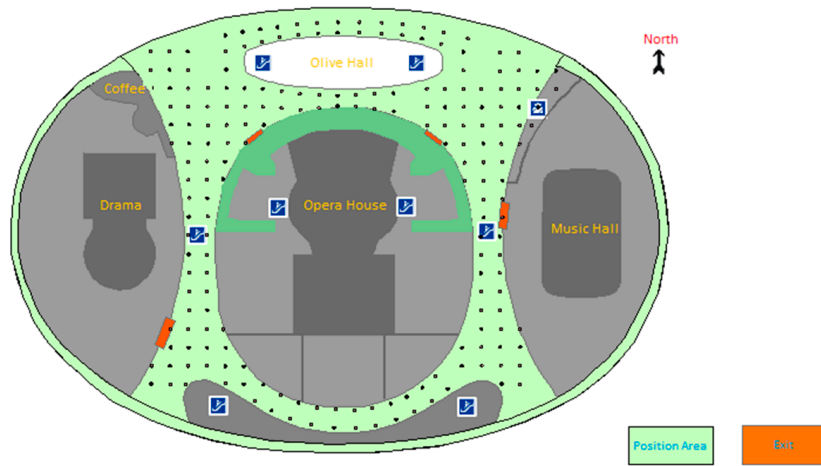

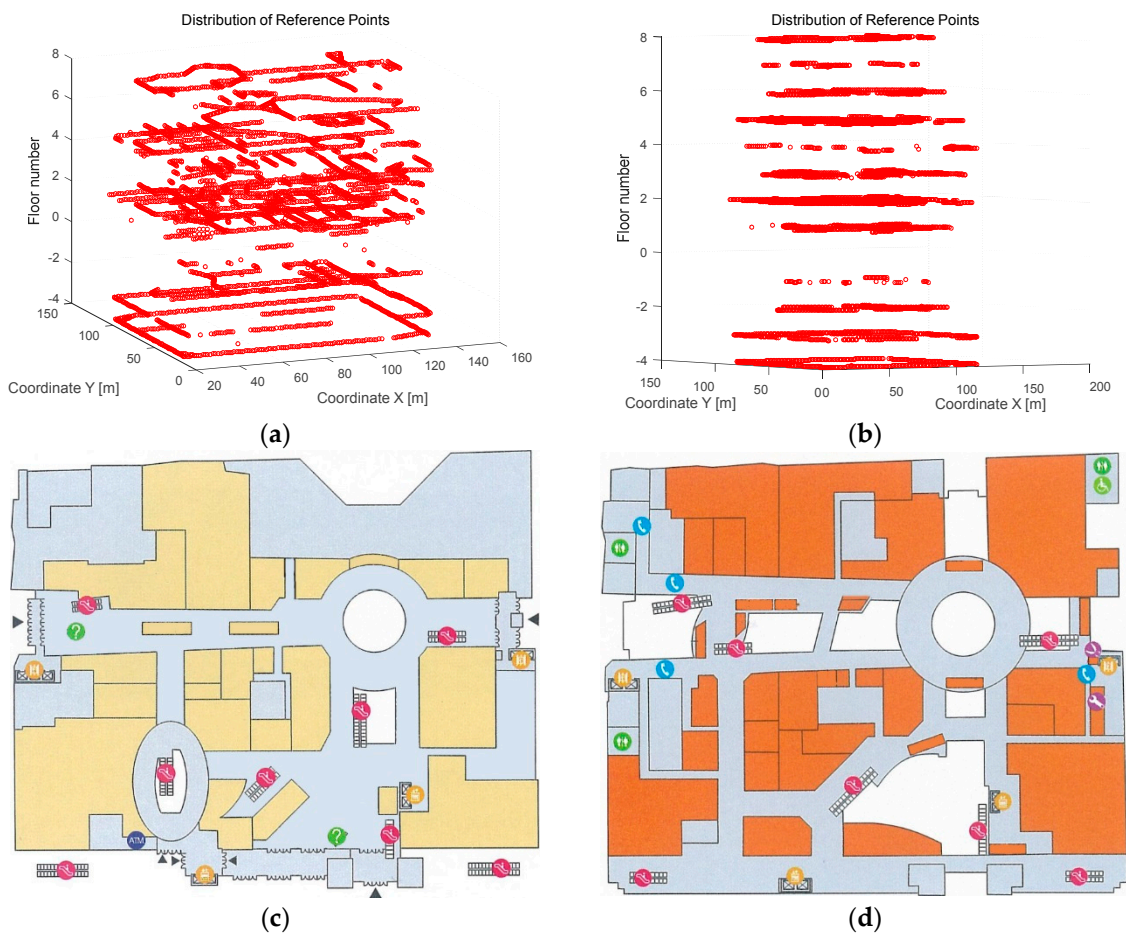**Figure 4.** Experimental environments for Test-bed 1.

Figure 5. Experimental environments for Test-bed 2: (**a**) reference points' distribution in Test-bed 2, twelve floors in total; (**b**) reference points' distribution in Test-bed 2 in side-view; (**c**) the first floor of Test-bed 2 in planar graph; and (**d**) the second floor of Test-bed 2 in planar graph.

Table 2. Distribution of RPs and the number of APs at 12 floors in Test-bed 2.

Floor Number	Number of RPs	Number of Hearable APs [1]
Floor 1	504	110
Floor 2	893	133
Floor 3	574	86
Floor 4	140	23
Floor 5	720	115
Floor 6	385	52
Floor 7	131	27
Floor 8	349	37
Floor b1 [2]	54	21
Floor b2	372	57
Floor b3	427	76
Floor b4	400	63

[1] The hearable APs represent the APs with different MACs, which can be measured RSSI at one floor; [2] Floor b1 means the first floor under ground in Test-bed 2.

4.2. Measurement Results and Analysis

4.2.1. Clustering Results and Analysis

The distributions of the RPs and their RSS values in experimental fingerprint database are investigated, and the distribution results both in scatter diagram and contour diagram are presented in Figure 6 (for Test-bed 1) and Figure 7 (for Test-bed 2). The contour diagram is based on the method of linear interpolation. It can be concluded from Figures 6 and 7 that the distribution of wireless signals in dense cluttered environments is very complex and full of randomness. In the contour diagrams, the points on the same contour line have similar RSS value in the database. Therefore, some distant points in physical space are close to each other in signal space, which is analyzed as SRP in Section 2.
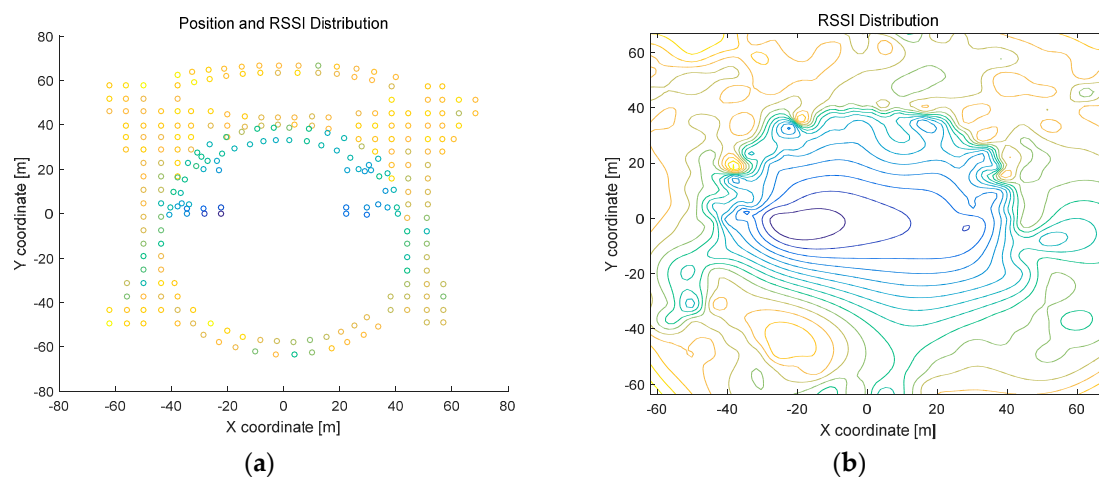
Figure 6. Reference points' Received Signal Strength Indicator (RSSI) distribution in Test-bed 1: (**a**) RSSI distribution in scatter diagram; and (**b**) RSSI distribution in contour map.
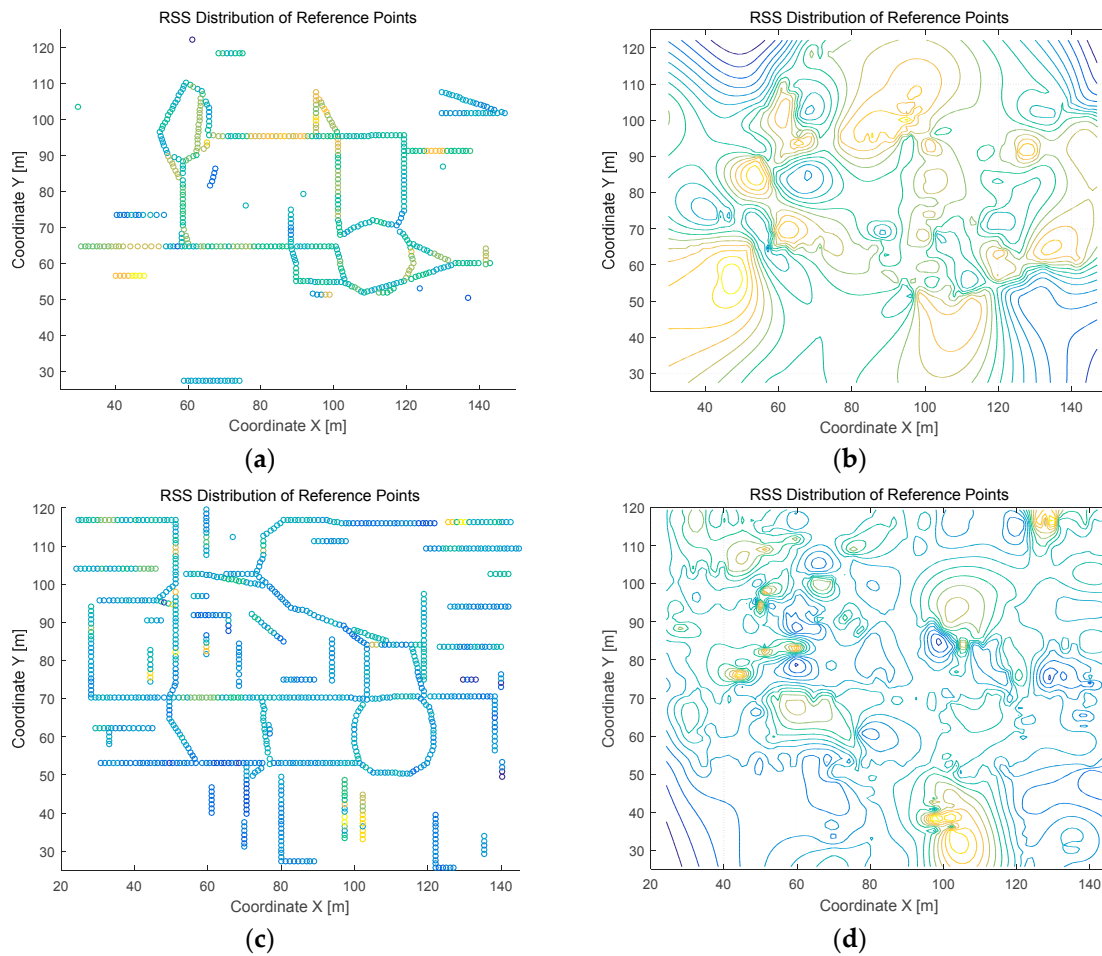
Figure 7. Reference points' RSSI distribution in Test-bed 2: (**a**) RSSI distribution in the first floor of Test-bed 2 in scatter diagram; (**b**) RSSI distribution in the first floor of Test-bed 2 in contour map; (**c**) RSSI distribution in the second floor of Test-bed 2 in scatter diagram; and (**d**) RSSI distribution in the second floor of Test-bed 2 in contour map.

Clustering comparison is conducted utilizing the methos of K-means and SEC. K-means is a classical and RSS-based clustering algorithm, and SEC is a novel and coordinate-based clustering algorithm proposed in this paper. Clustering results are shown in Figures 8 and 9 for Test-beds 1 and 2, respectively. It can be concluded that the clustering results using K-means is dispersive in physical space, and SRP has a significant influence on RSS-based clustering method. On the other hand, the clustering results utilizing SEC is compact, and SEC clustering is robust to SRP. Meanwhile, the capacity of each cluster in SEC clustering results is different, which means the flexibility of SEC is better than other coordinate-based clustering methods, like grid-based clustering.
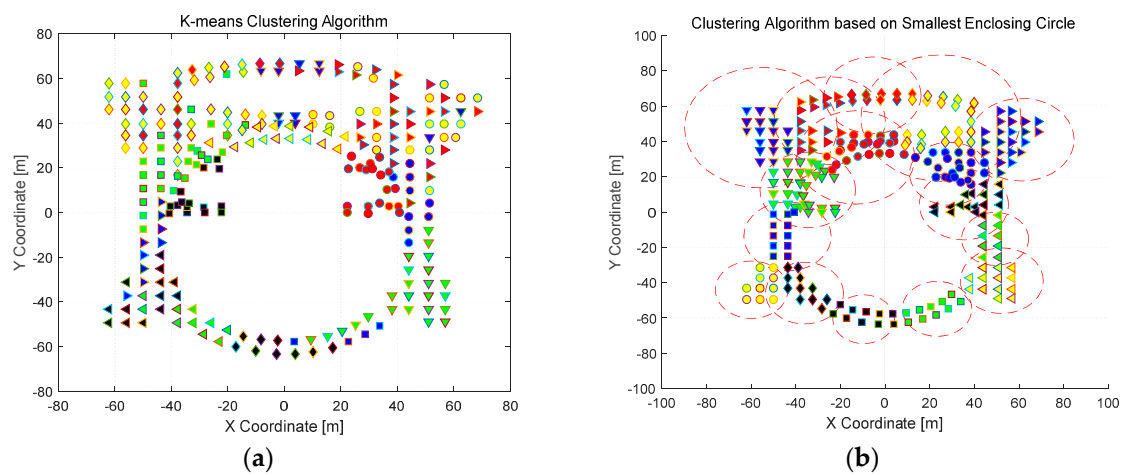
Figure 8. Clustering results comparison in Test-bed 1: (**a**) clustering results under K-means clustering algorithm; and (**b**) clustering results under Smallest Enclosing Circle (SEC)-based clustering algorithm.
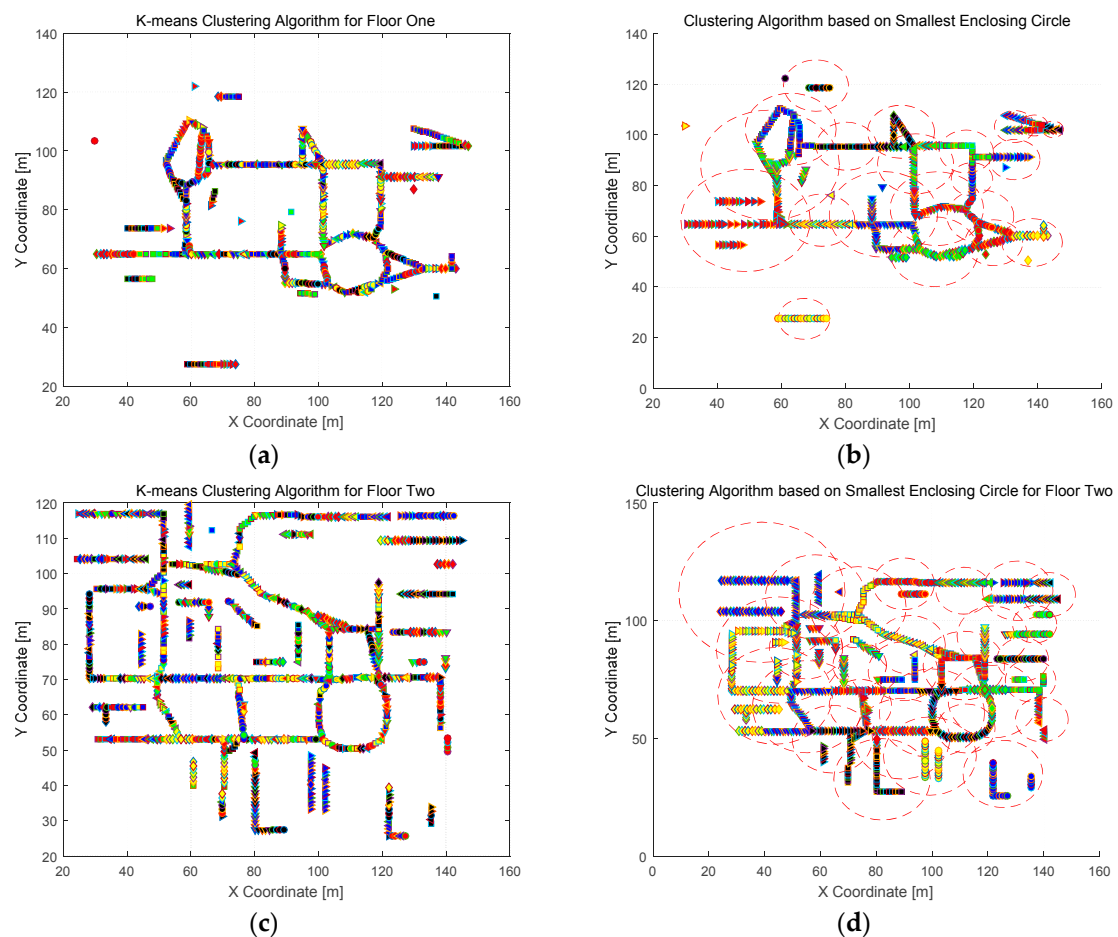
Figure 9. Clustering results comparison in Test-bed 2: (**a**) clustering results under K-means clustering algorithm in the first floor; (**b**) clustering results under SEC-based clustering algorithm in the first floor; (**c**) clustering results under K-means clustering algorithm in the second floor; and (**d**) clustering results under SEC-based clustering algorithm in the second floor.

In order to compare the clustering results between K-means and SEC algorithm more comprehensive and to validate the effectiveness of coordinate-based clustering method, a specific

metric, named Mean Intra-Cluster Distance (MICD), has been given to quantify the level of dispersion of RPs in these two clustering methods. MICD represents the mean intra-cluster distance of all the RPs in one same cluster. The calculation of MICD is shown as Equation (33) and the comparisons of MICD between K-means and SEC are shown in Figure 10.

$$MICD_k = \left(\sum_{P \in C_k} \sqrt{(x_P - x_{O_k})^2 + (y_P - y_{O_k})^2} \right) / n_k \tag{33}$$

where $MICD_k$ is the MICD of k-th cluster, (x_P, y_P) is the coordinates of the reference point P, and (x_{Ok}, y_{Ok}) is the coordinates of the k-th cluster-head O_k. C_k represents the k-th cluster. According to Equation (33), we calculated the MICD of clustering results using K-means and SEC. Figure 10 shows the comparisons of MICD in Test-beds 1 and 2. It can be concluded that the SEC clustering algorithm can cluster RPs more compact in physical space and more robust to SRP.
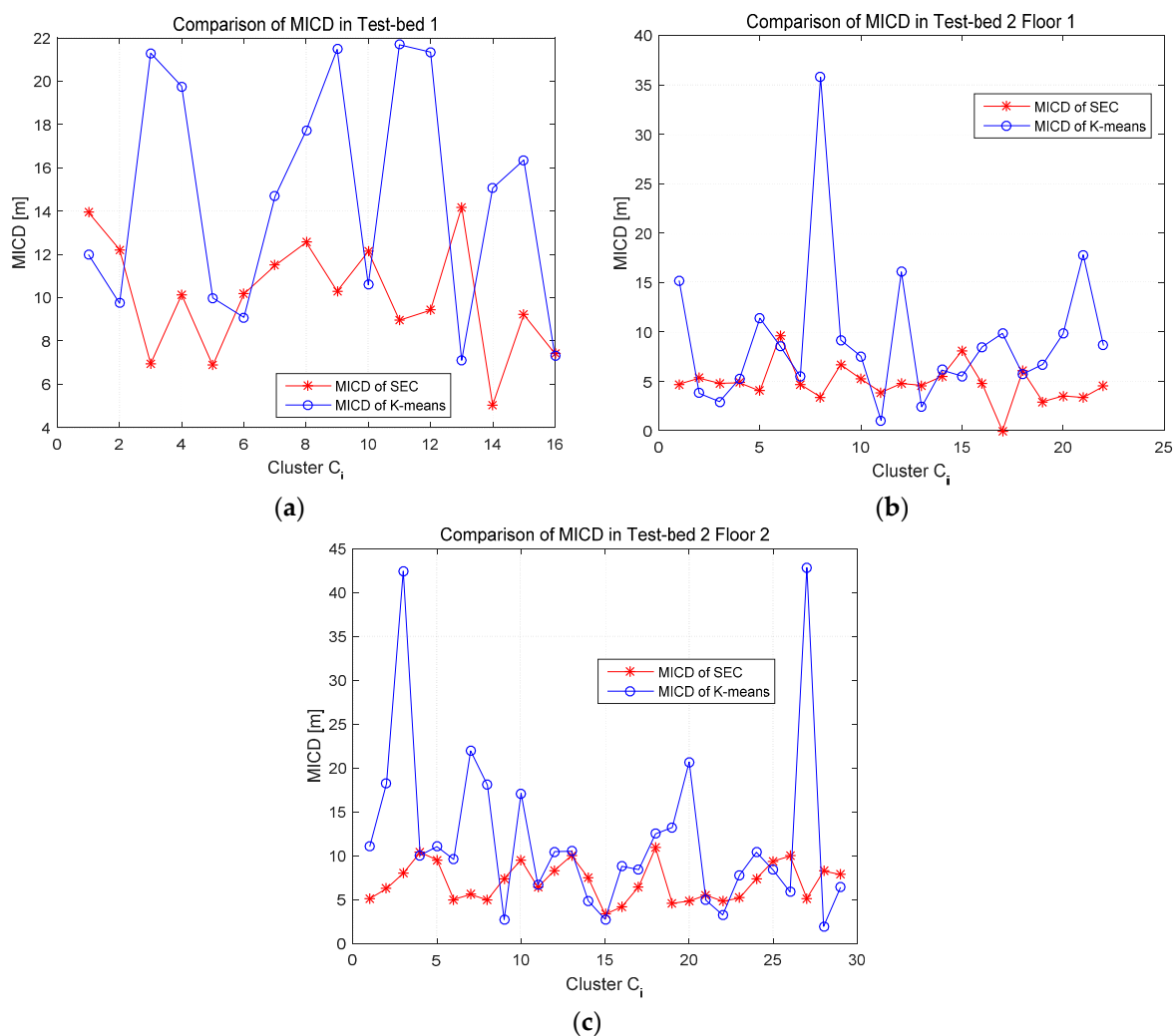
Figure 10. Comparison of Mean Intra-Cluster Distance (MICD) between K-means and SEC clustering: (**a**) MICD comparison in Test-bed 1; (**b**) MICD comparison in Test-bed 2 Floor 1; and (**c**) MICD comparison in Test-bed 2 Floor 2.

4.2.2. Positioning Accuracy Results and Analysis

Positioning accuracy is an important evaluation metric for an indoor positioning system. The accuracy of fingerprint positioning system utilizing K-means and SEC clustering algorithms

in the training stage is measured and compared for both two Test-beds. In this paper, we measured 250 test points for Test-bed 1, 500 test points for Test-bed 2 Floor 1 and 850 test points for Test-bed 2 Floor 2. All of these test points were selected randomly in the experimental scenarios, and the systems' positioning errors were measured for analysis. Regarding to the matching algorithm of fingerprint positioning system, this paper firstly conducted three matching algorithms, namely K Nearest Neighbor (KNN), Weight K Nearest Neighbor (WKNN) and Modified Weight K Nearest Neighbor (MWKNN) [23], in Test-bed 2 Floor 1 to evaluate the performance of positioning system. The comparison of positioning errors in cumulative distribution function (CDF) for these three matching algorithms has been shown in Figure 11.
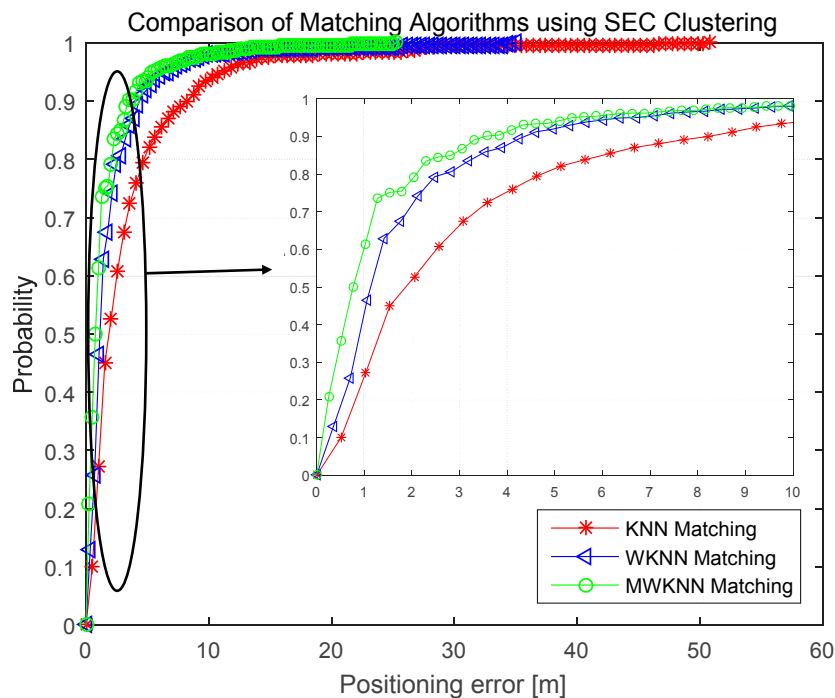
Figure 11. Comparison of matching algorithms using SEC clustering.

It can be concluded from Figure 11 that WKNN and MWKNN matching algorithms perform better positioning accuracy than KNN matching algorithm. Among all the 500 test points in Test-bed 1 Floor 1, there are 309 points (about 61.8%) with the positioning error within 1.02 m using MWKNN, 1.41 m using WKNN and 2.56 m using KNN. Meanwhile, there are 449 points (about 89.8%) with the positioning error within 3.31 m using MWKNN, 4.24 m using WKNN and 8.21 m using KNN. Without loss of generality, we select one same matching algorithm, namely MWKNN, for the purpose to compare the clustering performance utilizing both K-means and SEC. In the following discussion of this paper, the comparisons of positioning accuracy of these two clustering algorithms are all under the same MWKNN matching algorithm.

It should be pointed out that in Figure 11 there are still few location errors at test points (about 2% using MWKNN, 3% using WKNN and 6% using KNN) that are more than 10 m, and even some location errors (about 0.6% using MWKNN, 0.9% using WKNN, 2% using KNN) more than 20 m. There are two main reasons for these high location errors, one is the large difference of experimental environments between off-line training stage and on-line positioning stage at these test points, and the other is the large fluctuation in the transmitted power of some APs. Researchers who would implement this fingerprint positioning system could utilize other sensors' information, such as inertial sensors, to assist and calibrate the positioning results to achieve higher positioning accuracy.

The comparisons of positioning performance between K-means and SEC clustering algorithm utilizing MWKNN matching have been shown in Figure 12 and Table 3. It can be concluded that the positioning accuracy has been improved by about 32.7% in Test-bed 1, 71.7% in Test-bed 2 Floor 1 and 73.7% in Test-bed 2 Floor 2, using SEC compared to K-means. It can also be concluded that coordinate-based clustering algorithm, like SEC proposed in this paper, is more robust in indoor environments when SRP exists, compared with K-means.

(a)

(b)

(c)
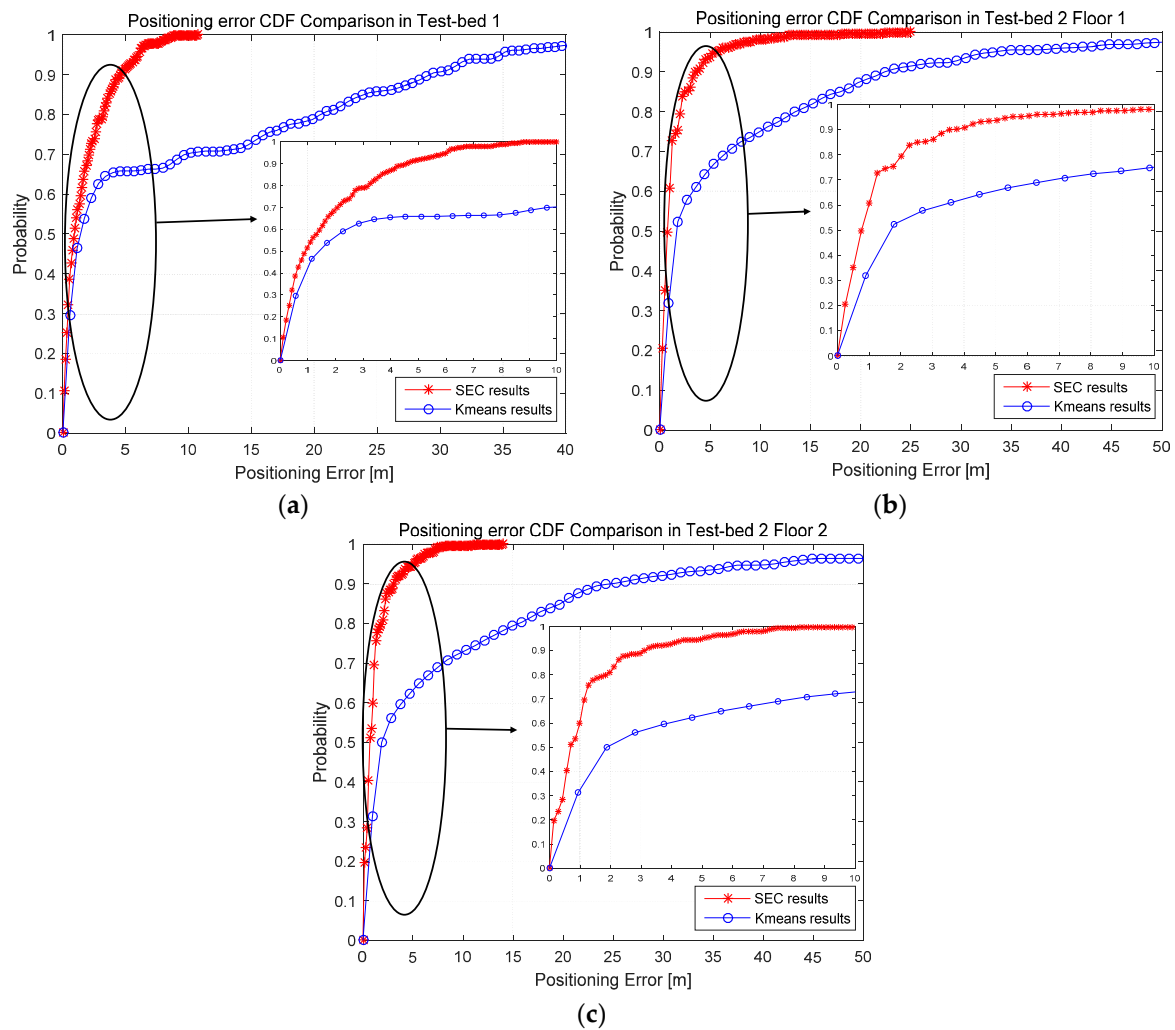
Figure 12. Positioning performance comparison between K-means and SEC in Cumulative Distribution Function (CDF) of positioning error under MWKNN matching algorithm: (**a**) CDF comparison in Test-bed 1; (**b**) CDF comparison in Test-bed 2 Floor 1; and (**c**) CDF comparison in Test-bed 2 Floor 2.

Table 3. Positioning Performance Comparison between K-means and SEC under Modified Weight K Nearest Neighbor (MWKNN) matching.

Positioning Error	Test-Bed 1	Test-Bed 2 Floor 1	Test-Bed 2 Floor 2
K-means	2.26 m (60%) 28.35 m (90%)	3.58 (60%) 21.50 m (90%)	3.73 m (60%) 22.43 m (90%)
SEC	1.52 m (60%) 4.67 m (90%)	1.07 m (60%) 3.78 m (90%)	0.98 m (60%) 3.24 m (90%)

4.2.3. Number of Clusters Evaluation

According to the analysis on the influence of number of clusters in Section 3, this paper conducted experiments using SEC clustering algorithm in Test-bed 1 and Test-bed 2 with different numbers of clusters. We selected different numbers of clusters, namely *k* in this paper, within its range given in Section 3, from one to the maximum. The performance of positioning systems with different *k* is measured utilizing the location error at 60% and 90% test points, and the comparisons are shown in Figure 13. It can be concluded that the selection of number of clusters has an influence on location error of the fingerprint positioning system, and within the selection range given in Section 3, the best positioning performance lies in the selection of maximum *k*, which equals to the square root of total number of RPs. Researchers who would implement these positioning systems could select number of clusters as the maximum *k* analyzed in Section 3.
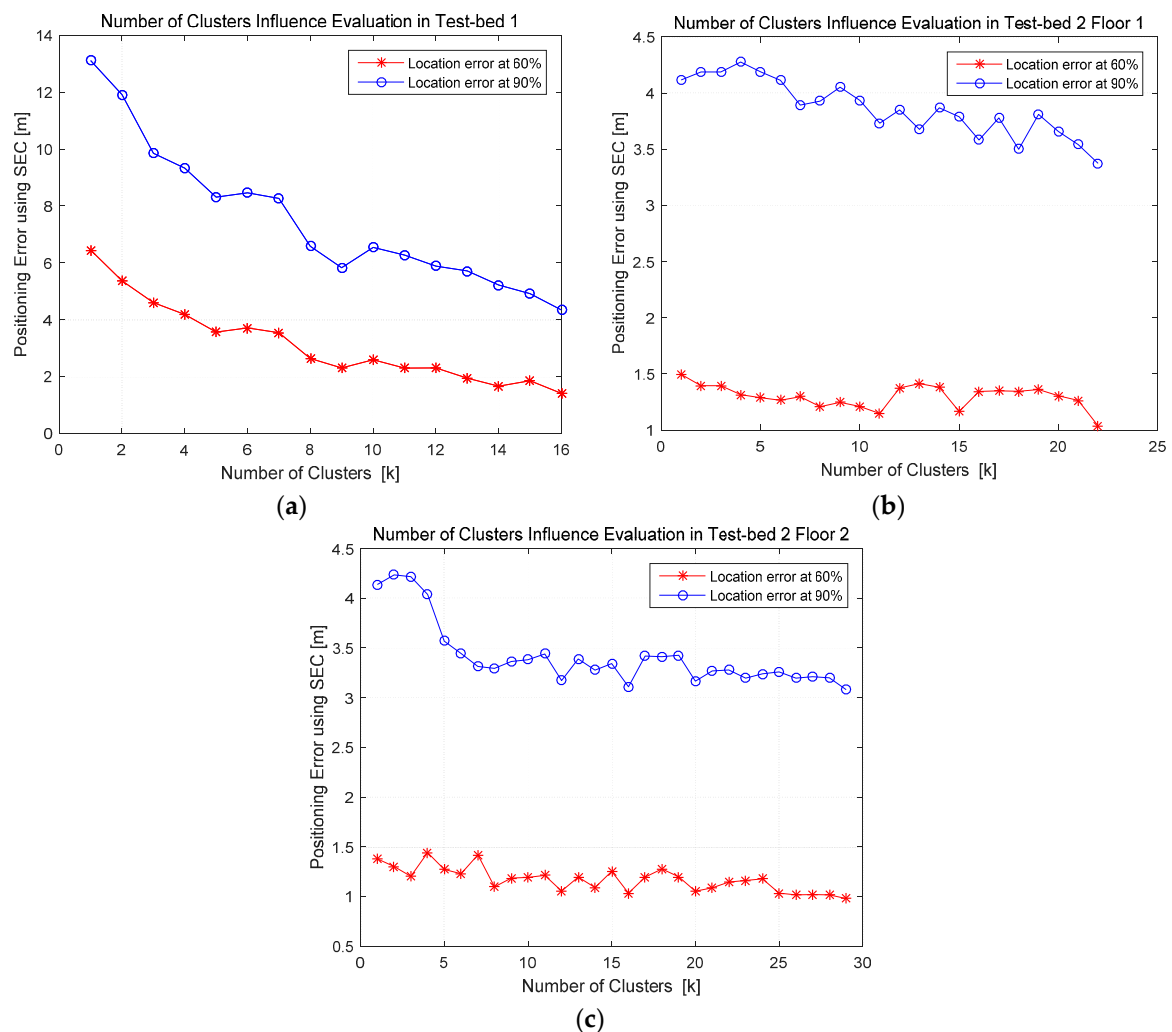
Figure 13. Number of clusters influence evaluation and comparison in Test-bed 1 and Test-bed 2 utilizing SEC clustering under MWKNN matching algorithm: (**a**) number of clusters influence evaluation in Test-bed 1; (**b**) number of clusters influence evaluation in Test-bed 2 Floor 1; and (**c**) number of clusters influence evaluation in Test-bed 2 Floor 2.

4.3. Related Works

Many indoor fingerprint-based positioning systems over the years have been proposed to improve the systems' positioning accuracy and robustness, which have been reviewed in [5]. RADAR system

has been proposed to locating and tracking users in indoor environments in [27], and has achieved a positioning accuracy of 2.37–2.65 m at 50% and 5.93–5.97 m at 90%. Horus system, which has been proposed in [28], has offered a joint clustering technique and probabilistic method for location estimation, and has achieved an accuracy of 0.86 m and 1.32 m for their test bed 1 and test bed 2, respectively. In [21], an AP placement algorithm based on map information has been proposed to improve the system's positioning performance in term of accuracy, and an accuracy within 3 m for 71% has been achieved. Our work has given focus on the influence of AP deployment for indoor fingerprint-based positioning system, and has proposed a coordinate-based clustering method to reduce these influences on positioning accuracy. Some of the performance comparisons of related works are shown in Table 4 and our work has been conducted also aiming at giving guidance for researchers who would implement a fingerprint-based positioning system in indoor dense cluttered environment.

Table 4. Comparison of related works.

Related Works	Positioning Accuracy	Experimental Environments
RADAR [27]	2.37 m (50%) and 5.93 m (90%)	Real testing environments
Horus [28]	0.86 m and 1.32 m	Test bed 1 (68.2×25.9 m^2); Test bed 2 (11.8×35.9 m^2)
Works in [21]	3 m (71%)	1200 m^2 testing environments
Our work	1.52 m (60%) and 3.24 m (90%)	Test-bed 1 (210×140 m^2); Test-bed 2 (120×90 m^2)

5. Conclusions

In this paper, a novel coordinate-based clustering method, named Smallest Enclosing Circle (SEC) based clustering, is proposed, analyzed and estimated. The influence on fingerprint database and positioning accuracy of AP deployments is studied, and a probable phenomenon named SRP in dense cluttered environments is analyzed. Traditional clustering, like K-means, used in fingerprint positioning cannot satisfy the accuracy demand in these environments because of the existence of SRP. The SEC clustering is proposed to reduce the influence of AP deployment and SRP, and to improve the accuracy and robustness of the whole system in dense cluttered indoor environments. The experiments in two different real public indoor area are conducted to validate the advantages of the proposed method. All results are based on real field measurements at China National Grand Theatre and XiDan Joy City. The results show that SEC clustering is more robust and flexible compared to RSS-based clustering, and the improvement of positioning accuracy using SEC clustering is about 32.7% for Test-bed 1, 71.7% for Test-bed 2 Floor 1 and 73.7% for Test-bed 2 Floor 2, compared to RSS-based clustering. Meanwhile, the accuracy influence of number of clusters has been analyzed and evaluated by environments, and results have shown that SEC clustering performed more robust than K-means clustering in indoor environments with SRP. Our study is useful in fingerprint positioning systems in dense cluttered environments where the AP deployment is ordered and fixed, and it is also a guid for AP deployment in a new environment to implement fingerprint positioning.

Author Contributions: Wen Liu and Zhongliang Deng are both the principal investigators; Xiao Fu conceived and conducted the experiments and analysis; and Xiao Fu analyzed the data and wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

This appendix is for the additional experimental results in Test-bed 2 at XiDan Joy City. The clustering results and positioning performance comparisons for the other floors are shown in in Figure A1.
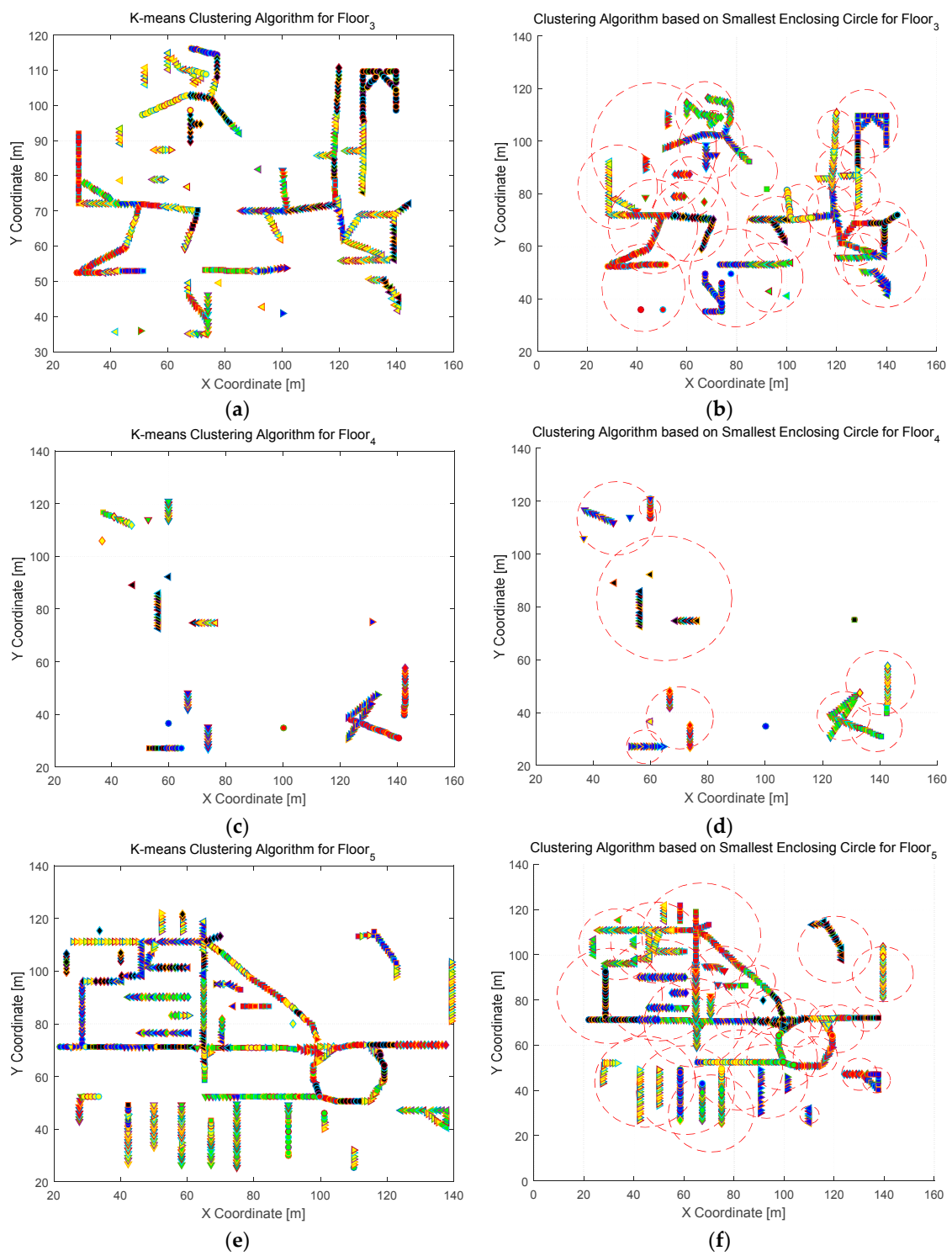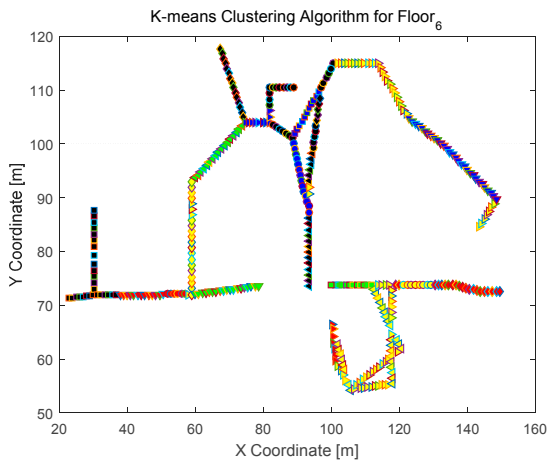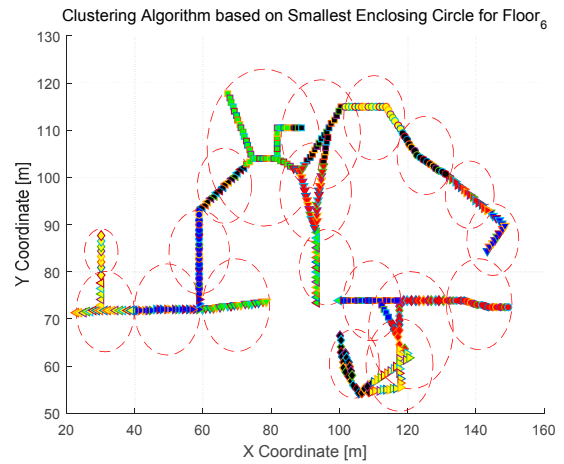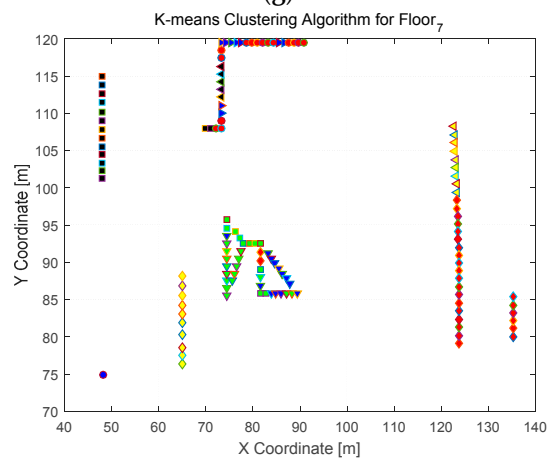
Figure A1. *Cont.*

(**g**)

(**h**)

(**i**)
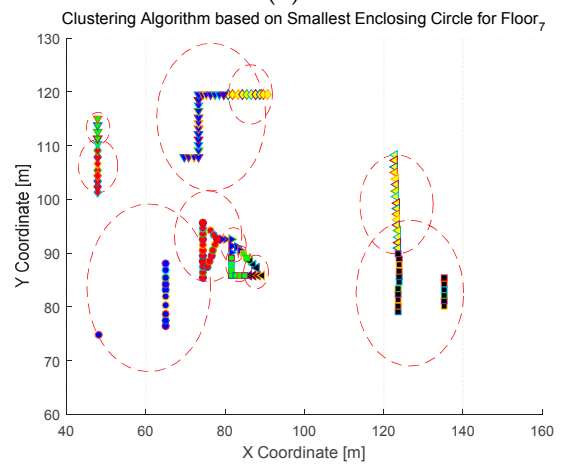
(**j**)
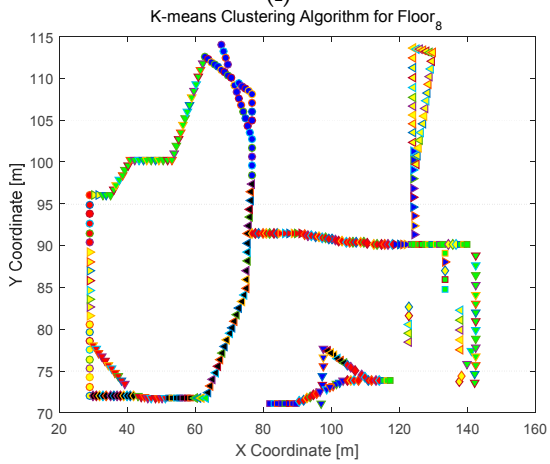
(**k**)
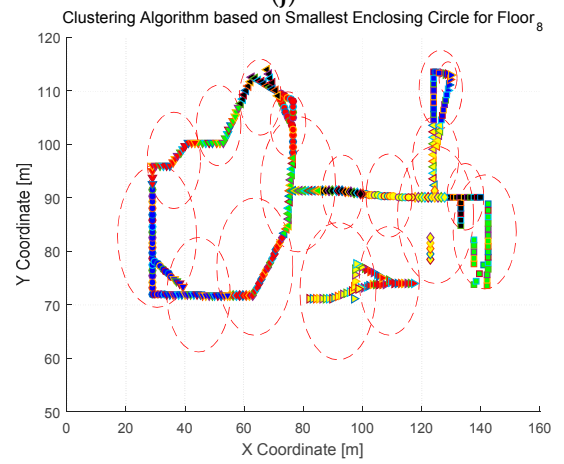
(**l**)

Figure A1. *Cont.*

(**m**)

(**n**)

(**o**)

(**p**)

(**q**)

(**r**)

Figure A1. *Cont.*

(**s**)　　　　　　　　　　　　　　　　　　　　　　　　(**t**)

(**u1**)　　　　　　　　　　　　　　　　　　　　　　　　(**u2**)

(**u3**)　　　　　　　　　　　　　　　　　　　　　　　　(**u4**)

Figure A1. *Cont.*
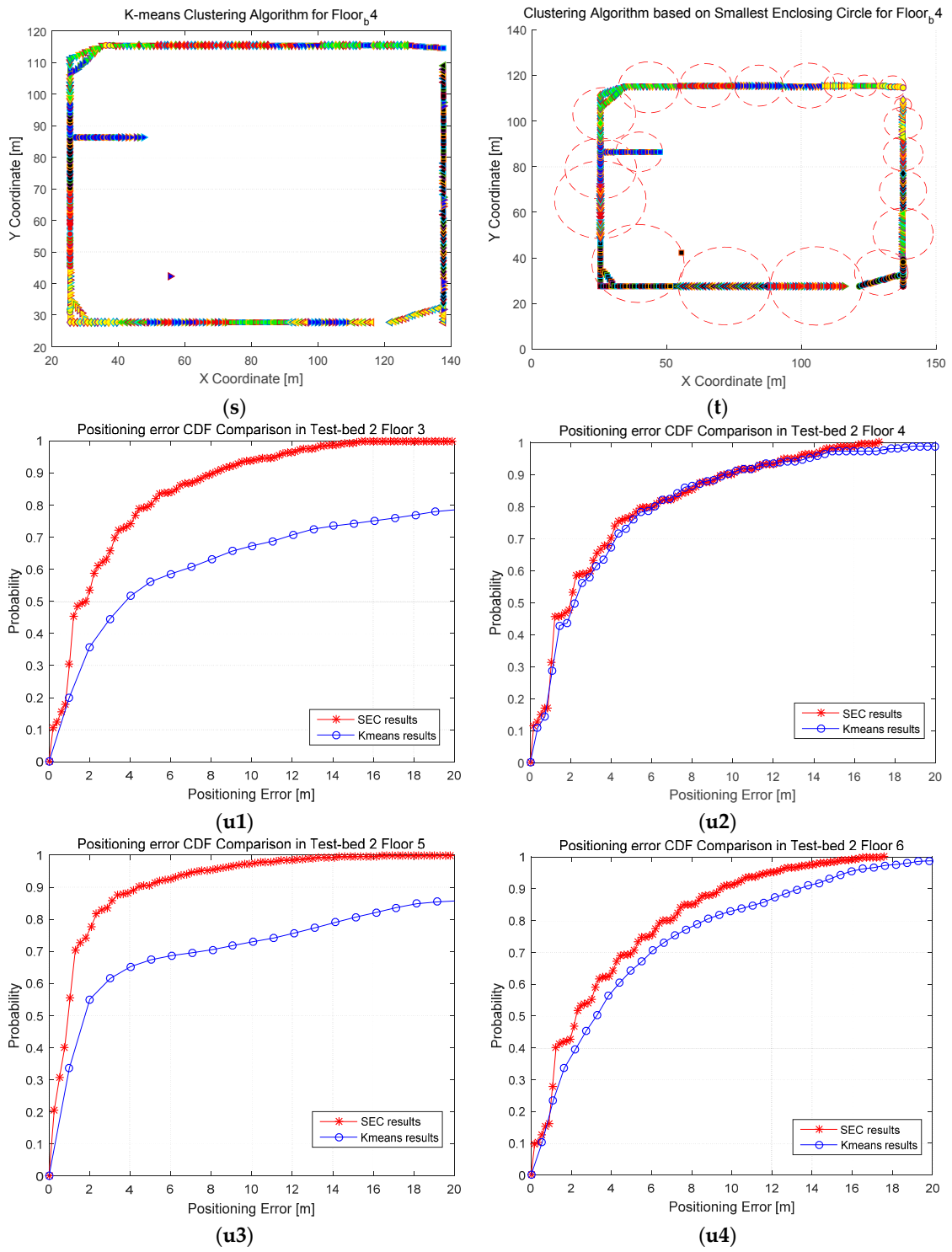
(**u5**)

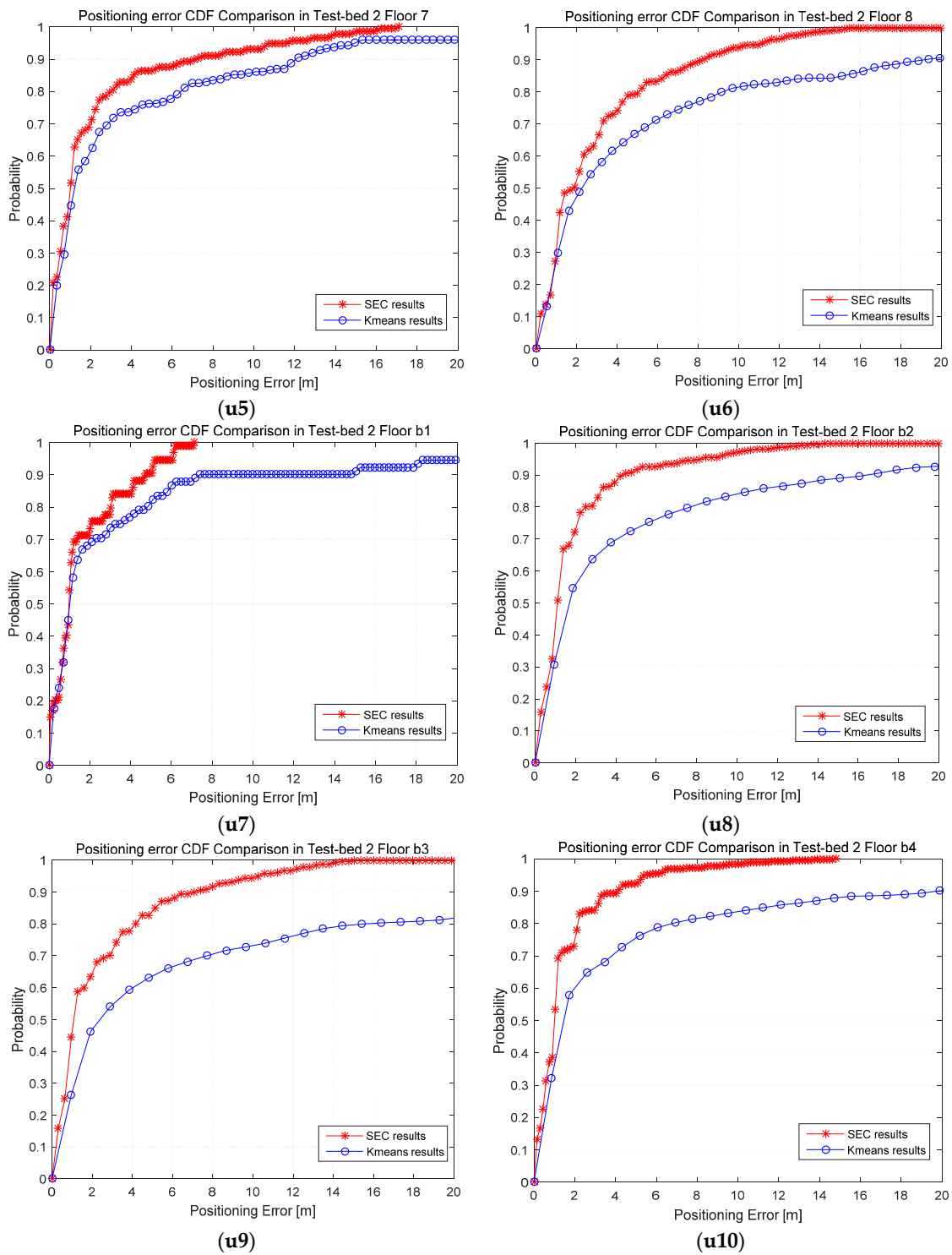(**u6**)

(**u7**)

(**u8**)

(**u9**)

(**u10**)
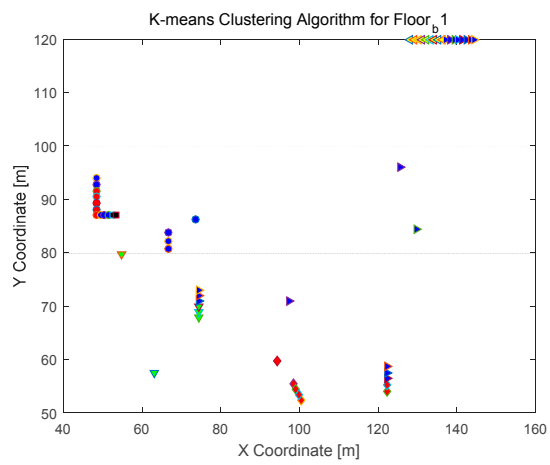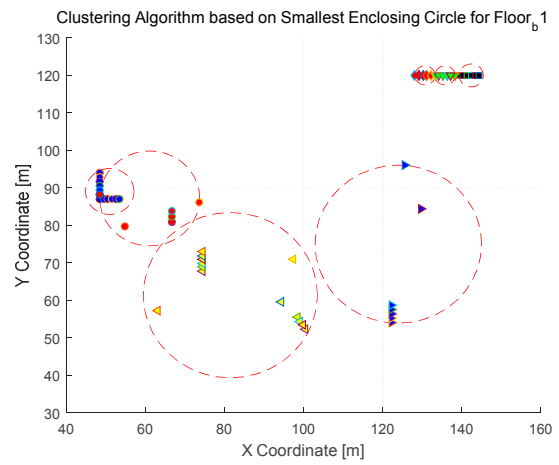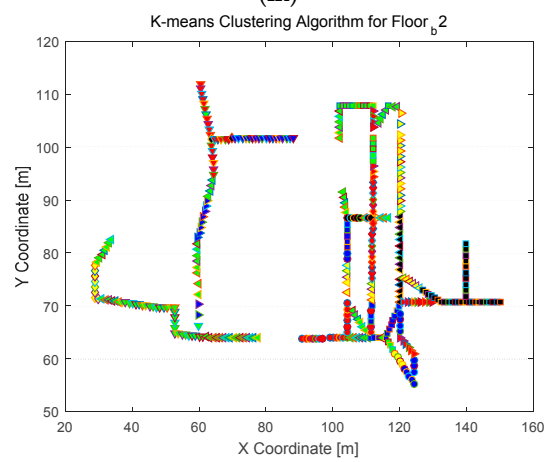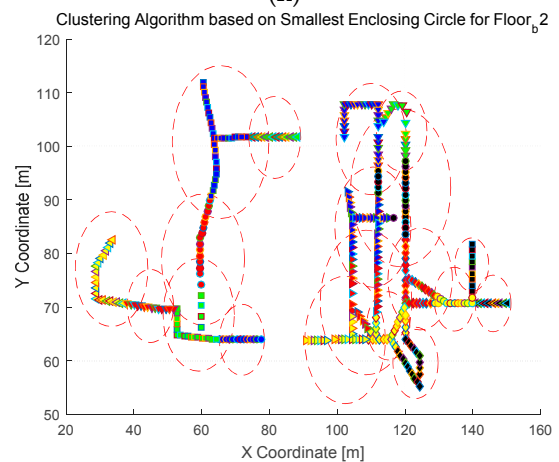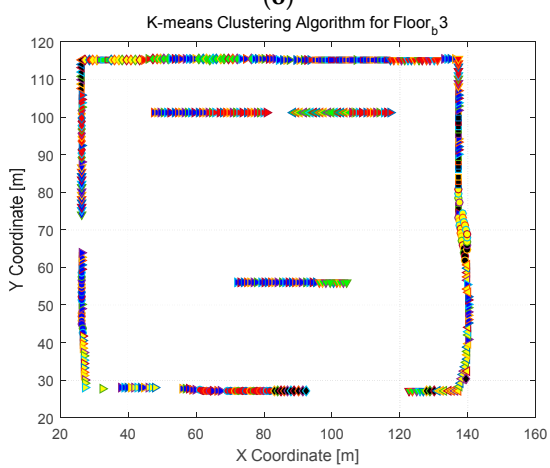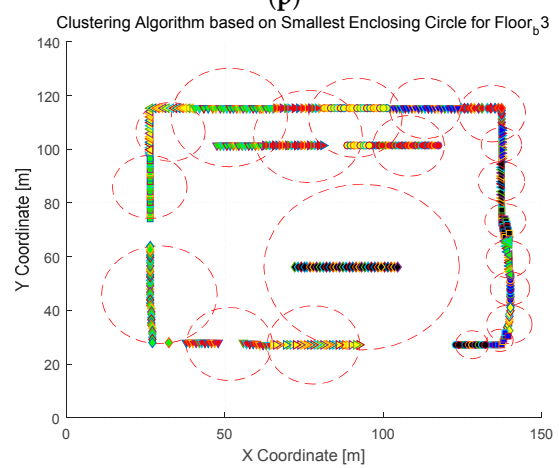
Figure A1. Clustering results and positioning performance comparisons for Test-bed 2: (**a**–**t**) the clustering results and comparisons between K-means and SEC from Floor 3 to Floor b4 in Test-bed 2; and (**u1**–**u10**) the positioning performance comparisons between K-means and SEC from Floor 3 to Floor b4 in Test-bed 2.

References

1. "Indoor Location in Retail: Where Is the Money?" ABI Research: Location Technologies Market Research. May 2015. Available online: https://www.abiresearch.com/market-research/service/location-technologies/ (accessed on 12 May 2015).

2. Deak, G.; Curran, K.; Condell, J. A survey of active and passive indoor localisation systems. *Comput. Commun.* **2012**, *35*, 1939–1954. [CrossRef]

3. Hossain, A.; Soh, W.S. A survey of calibration-free indoor positioning systems. *Comput. Commun.* **2015**, *66*, 1–13. [CrossRef]

4. Koyuncu, H.; Yang, S.H. A survey of indoor positioning and object locating systems. *IJCSNS Int. J. Comput. Sci. Netw. Secur.* **2010**, *10*, 121–128.

5. Liu, H.; Darabi, H.; Banerjee, P.; Liu, J. Survey of wireless indoor positioning techniques and systems. *IEEE Trans. Syst. Man Cybern. C Appl. Rev.* **2007**, *37*, 1067–1080. [CrossRef]

6. Jiang, W.; Li, Y.; Rizos, C. Precise indoor positioning and attitude determination using terrestrial ranging signals. *J. Navig.* **2015**, *68*, 274–290. [CrossRef]

7. Alarifi, A.; Al-Salman, A.; Alsaleh, M.; Alnafessah, A.; Al-Hadhrami, S.; Al-Ammar, M.A.; Al-Khalifa, H.S. Ultra wideband indoor positioning technologies: Analysis and recent advances. *Sensors* **2016**, *16*, 707. [CrossRef] [PubMed]

8. Yu, F.; Jiang, M.; Liang, J.; Qin, X.; Hu, M.; Peng, T.; Hu, X. 5G WiFi signal-based indoor localization system using cluster k-nearest neighbor algorithm. *Int. J. Distrib. Sens. Netw.* **2014**, *2014*. [CrossRef]

9. Kriz, P.; Maly, F.; Kozel, T. Improving indoor localization using bluetooth low energy beacons. *Mob. Inf. Syst.* **2016**, *2016*. [CrossRef]

10. Cheon, J.; Hwang, H.; Kim, D.; Jung, Y. IEEE 802.15. 4 Zigbee-based time-of-arrival estimation for wireless sensor networks. *Sensors* **2016**, *16*, 203. [CrossRef] [PubMed]

11. Montaser, A.; Moselhi, O. RFID indoor location identification for construction projects. *Autom. Constr.* **2014**, *39*, 167–179. [CrossRef]

12. Gu, Y.; Lo, A.; Niemegeers, I. A survey of indoor positioning systems for wireless personal networks. *IEEE Commun. Surv. Tutor.* **2009**, *11*, 13–32. [CrossRef]

13. Bisio, I.; Lavagetto, F.; Marchese, M.; Sciarrone, A. Performance comparison of a probabilistic fingerprint-based indoor positioning system over different smartphones. In Proceedings of the 2013 International Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS)—Part of SummerSim 2013 Multiconference, Toronto, ON, Canada, 7–10 July 2013; pp. 161–166.

14. Bisio, I.; Lavagetto, F.; Marchese, M.; Sciarrone, A. Smart probabilistic fingerprinting for WiFi-based indoor positioning with mobile devices. *Pervasive Mob. Comput.* **2016**, *31*, 107–123. [CrossRef]

15. Bisio, I.; Lavagetto, F.; Marchese, M.; Sciarrone, A. GPS/HPS-and Wi-fi fingerprint-based location recognition for check-in applications over smartphones in cloud-based LBSs. *IEEE Trans. Multimedia* **2013**, *15*, 858–869. [CrossRef]

16. Bisio, I.; Lavagetto, F.; Marchese, M.; Sciarrone, A. Energy efficient WiFi-based fingerprinting for indoor positioning with smartphones. In Proceedings of the 2013 IEEE Globecom Workshops (GC Wkshps), Atlanta, GA, USA, 9–13 December 2013; pp. 4639–4643.

17. Laitinen, E.; Lohan, E.S. On the choice of access point selection criterion and other position estimation characteristics for WLAN-based indoor positioning. *Sensors* **2016**, *16*, 737. [CrossRef] [PubMed]

18. Razavi, A.; Valkama, M.; Lohan, E.S. K-means fingerprint clustering for low-complexity floor estimation in indoor mobile localization. In Proceedings of the Globecom Workshop on Localization and Tracking: Indoors, Outdoors and Emerging Networks, San Diego, CA, USA, 6–10 December 2015; pp. 1–7.

19. Huang, Z.; Xia, J.; Yu, H.; Guan, Y.; Chen, J. Clustering combined indoor localization algorithms for crowdsourcing devices: Mining RSSI relative relationship. In Proceedings of the Wireless Communications and Signal Processing (WCSP), Hefei, China, 23–25 October 2014; pp. 1–6.

20. Zhang, Z.; Di, X.; Tian, J.; Chen, P. A WLAN planning method for indoor positioning system. In Proceedings of the 2016 International Conference on Information Networking (ICOIN), Kota Kinabalu, Malaysia, 13–15 January 2016; pp. 303–307.

21. Du, X.; Yang, K. A Map-assisted WiFi AP placement algorithm enabling mobile device's indoor positioning. *IEEE Syst. J.* **2016**, *PP*. [CrossRef]

22. Laitinen, E.; Lohan, E.S. Access point topology evaluation and optimization based on Cramér-Rao Lower Bound for WLAN indoor positioning. In Proceedings of the 2016 International Conference on Localization and GNSS (ICL-GNSS), Barcelona, Spain, 28–30 June 2016; pp. 1–5.

23. Liu, W.; Fu, X.; Deng, Z.; Xu, L.; Jiao, J. Smallest enclosing circle-based fingerprint clustering and modified-WKNN matching algorithm for indoor positioning. In Proceedings of the 2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN), Madrid, Spain, 4–7 October 2016.

24. Ilango, M.R.; Mohan, V. A survey of grid based clustering algorithms. *Int. J. Eng. Sci. Technol.* **2010**, *2*, 3441–3446.

25. Chen, J.; Fang, B.X.; Tan, J.L.; Jin, S.C. Index filtering algorithm based on minimum enclosing circle partition. *Chin. J. Comput.* **2012**, *35*, 2139–2146. [CrossRef]

26. Yang, S.L.; Li, Y.S.; Hu, X.X.; Pan, R.Y. Optimization study on k value of k-means algorithm. *Syst. Eng. Theory Pract.* **2006**, *2*, 97–101.

27. Bahl, P.; Padmanabhan, V.N. RADAR: An in-building RF-based user location and tracking system. In Proceedings of the Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies, Tel Aviv, Israel, 26–30 March 2000; Volume 2, pp. 775–784.

28. Youssef, M.; Agrawala, A. The Horus WLAN location determination system. In Proceedings of the 3rd International Conference on Mobile Systems, Applications, and Services, Seattle, WA, USA, 6–8 June 2005; pp. 205–218.