

Article

# A Drift-of-Stay Pattern Extraction Method for Indoor Pedestrian Trajectories for the Error and Accuracy Assessment of Indoor Wi-Fi Positioning

Dengbo Yu, Qingwu Hu  and Shaohua Wang \*

School of Remote Sensing and Information Engineering, Wuhan University, 129 Luoyu Road, Wuhan 430079, China; dengboyu@whu.edu.cn (D.Y.); huqw@whu.edu.cn (Q.H.)

\* Correspondence: shwang@whu.edu.cn

Received: 7 August 2019; Accepted: 21 October 2019; Published: 23 October 2019



**Abstract:** The uncertainty of indoor Wi-Fi positioning is susceptible to many factors, such as sensor distribution, the internal environment (e.g., of a shopping mall), differences between receivers, and the flow of people. In this paper, an indoor pedestrian trajectory pattern mining approach for the assessment of the error and accuracy of indoor Wi-Fi positioning is proposed. First, the stay points of the customer were extracted from the pedestrian trajectories based on the spatiotemporal staying patterns of the customers in a shopping mall. Second, the drift points were distinguished from the stay points through analysis of noncustomer behavior patterns. Finally, the drift points were presented to calculate the errors in the pedestrian trajectories for the accuracy assessment of the indoor Wi-Fi positioning system. A one-month indoor pedestrian trajectories dataset from the Xinxiang Baolong shopping mall in Henan Province, China, was used for the assessment of the error and accuracy values with the proposed approach. The experimental results were verified by incorporating the distribution of the AP sensors. The proposed approach using big data pattern mining can explore the error distribution of indoor positioning systems, which can provide strong support for improving indoor positioning accuracy in the future.

**Keywords:** indoor Wi-Fi positioning; pedestrian trajectory; walking pattern; error distribution; accuracy

## 1. Introduction

The modern rhythm of life means that most human activities, whether work, entertainment or rest, occur indoors. There is a rich diversity and a considerable amount of data from a number of sources. For instance, in large shopping malls, there are tens of thousands of customers every day, and it is very important to understand the behavior of customers to evaluate the store layouts and brand introductions to enhance the profitability of the mall [1–3]. For all types of transportation corridors, the distributions of indoor flow density and the transport efficiency are inseparable. More reasonable arrangements for the diversion of pedestrians are a priority, and in situations with high-density indoor venues, safe evacuation in the event of an emergency is particularly important. Thus, the indoor positioning data of pedestrians can be used to address these problems and implement effective solutions.

The continuous improvement of the performance of hardware sensor equipment and the ultrahigh penetration rate of smart phones lay the foundation for positioning human indoor activity patterns [4]. These technologies provide strong hardware support for understanding, forecasting, and analyzing indoor human activities. Currently, there are numerous researchers engaged in related research, whether it is the theoretical study of positioning accuracy or the analysis of human behavior patterns based on indoor positioning data with fairly objective research goals. Therefore, research on indoor

positioning is currently a popular topic. Indoor positioning technology provides the TOA (time difference of arrival) based on the electromagnetic wave propagation time positioning technology [5,6], the AOA (angle of arrival) based on the incidence angle of electromagnetic waves [7,8], the TDOA based on the arrival of the frequency of magnetic wave positioning technology [9,10], signal strength based on RSSI (received signal strength indication) positioning technology [11,12], and the AP (access point) sensor signal strength based on positioning technology [13,14].

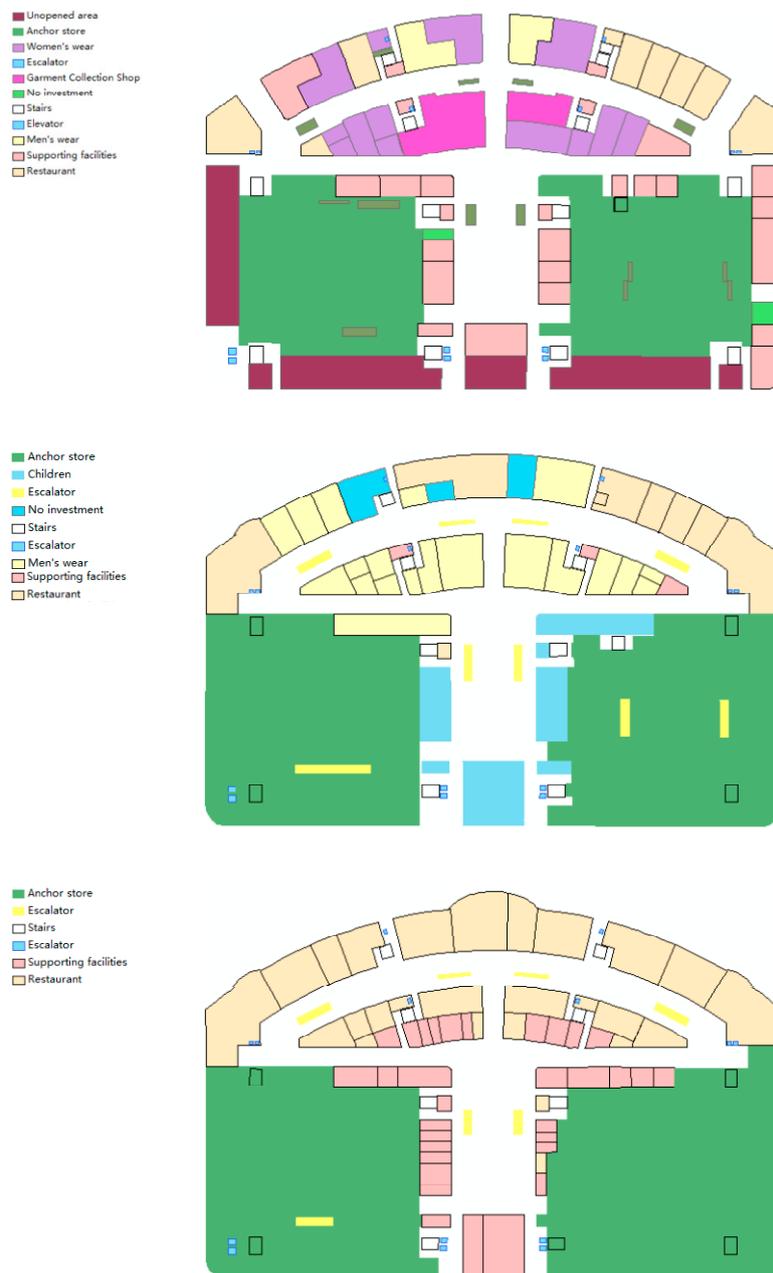
However, regardless of the type of positioning technology, there will inevitably be errors, and improved positioning accuracy first relies on identifying the factors that affect regional positioning [15–17]. The large number of localization algorithms and the complex factors influencing error in positioning data have been discussed by several authors [18–23]. In terms of indoor structures, the indoor positioning accuracy in an ideal environment would be much better than the actual situation [24–26]. Due to the diversity of indoor environments, the influences of different indoor areas on the signal are different [27–29]. The results of the statistical analysis of error reveals that the main source of error in regional positioning data is the influence of the wireless communication environment between the mobile terminals [30,31] and AP sensors [32–36], followed by the error caused by the most important module positioning algorithms in the indoor positioning system and the positioning error caused by the difference in the signal reception intensities of different mobile terminals [37–40]. Using multiple data types to improve the credibility of indoor positioning data is a good choice. For example, credit card data have been presented to analyze customer behavior [41], and Bluetooth has been used to enhance the positioning ability. The system is affected by high population densities, such as those in a mall, and the cost of operation is high [42,43]. Considering the actual environment, the positioning accuracy and the overall operating costs in malls, this type of positioning scheme has mainly been employed to locate indoor pedestrians by uniformly arranging AP sensors throughout the mall [44]. The accuracy of the Wi-Fi AP positioning systems is usually assessed by the theory model based on the signal positioning principle. Therefore, it is a challenge to propose an effective method for evaluating the quality of indoor positioning accuracy [45].

Indoor Wi-Fi positioning systems generate massive trajectory data. Indoor trajectory data are similar to outdoor GNSS (Global Navigation Satellite System) trajectories from taxis, cars and smartphones, which provide rich information for spatiotemporal data mining for pattern analysis [46–48]. The GNSS trajectory data have been used not only to explore pick-up/drop-off information but also to provide taxi drivers with cruising routes to the recommended pick-up locations [49–51]. GNSS trajectories are used for urban traffic analyses, pattern analyses of crowd motion and recommending locations to friends [52]. Thus, these spatiotemporal data mining methods can also be implemented for the trajectory data from indoor Wi-Fi positioning systems. Recently, an indoor trajectory mining algorithm was proposed to generate alternative maps for indoor localization [53,54]. The goal of trajectory processing is noise filtering. The accuracy of the indoor crowdsourced positioning data is an important factor in trajectory learning.

It is very difficult to obtain ground truth data for the indoor Wi-Fi positioning system. Our paper introduces a GIS-based spatiotemporal data-processing method for very large amounts of indoor trajectory data to extract the stay patterns of pedestrians for the accuracy assessment of indoor Wi-Fi positioning systems. In this paper, the phenomenon of positioning drift was found in the indoor pedestrian trajectory data in a shopping mall, and we used this phenomenon to evaluate the errors in the indoor Wi-Fi positioning system. The spatiotemporal attributes of the indoor pedestrian trajectories collected by the Wi-Fi access points in the mall are used to extract the stay trajectory based on the spatiotemporal staying patterns of the customers. The noncustomer behavior patterns were used to discover drift points from the stay points. The drift points can reflect the real errors in the indoor Wi-Fi positioning system. The distribution structure of the AP sensors was used to validate the error assessment results of the proposed approach. The results show that the proposed approach using big data pattern mining can explore the error distribution of indoor positioning systems, which can provide strong support for improving indoor positioning accuracy in the future.

## 2. Data and Materials

There is a large amount of pedestrian trajectory data from the indoor Wi-Fi positioning of the Baolong Shopping Mall in Xinxiang, Henan Province, P.R. China, comprising the position information of approximately 6 million customers; approximately 200 million records are generated each day. The data collection occurred from 18 September 2015, to 18 October 2015. There are three floors in the shopping mall, and the locations and types of store on each floor are completely different, as shown in Figure 1.

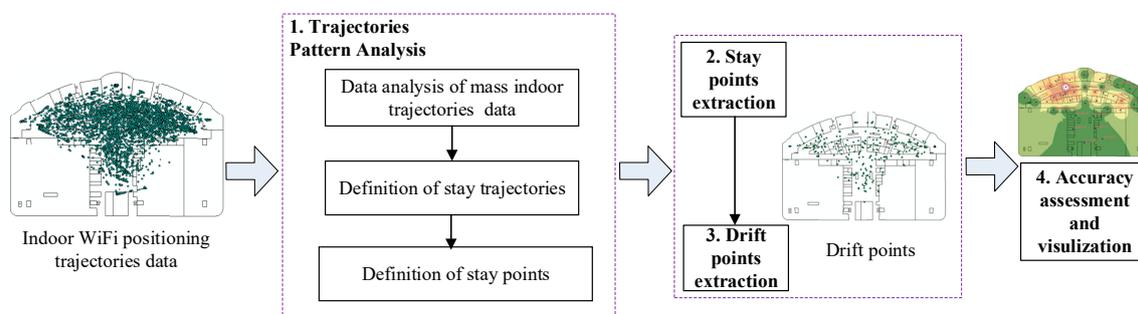


**Figure 1.** Distribution of stores in the shopping mall (floors 1, 2, and 3).

Each pedestrian positioning record included unique MAC addresses for the mobile devices, which are the basis for distinguishing the individual trajectories. The positioning data also contain the coordinates in the XY plane, floor location (FLOOR) and time (TIME).

### 3. Method

There are mass position data that come from the pedestrians, while they are walking in the indoor environment. Their smartphones interact with the Wi-Fi AP sensors, and the server of the indoor Wi-Fi position system records their trajectories. We cannot easily deduce the accuracy of each smartphone from its built-in position sensor. Thus, the mass position data of different people at different times makes possible the error analysis of the indoor positioning system. Most of the indoor position data are the moving points of the pedestrians, and they have different speeds and walking patterns. It is difficult to analyze the positioning errors via these moving points. The pedestrians have a typical customer stay behavior pattern while shopping. During the stay while shopping, the customer's location will remain relatively constant for a short period of time. However, due to the positioning errors of the indoor positioning system, the coordinates of the stay points will change beyond the relative fixed position. We defined this phenomenon as "drift". Thus, the drift points extracted from stay trajectories can be used for the assessment of the accuracy of the indoor positioning system. The framework of the proposed error analysis approach is described in Figure 2.



**Figure 2.** The technical framework of the proposed approach.

As Figure 2 shows, first, the pedestrian trajectory patterns in mass positioning data are implemented based on the analysis of the mass indoor data and the customer behavior characteristics. The stay trajectories and stay points are defined for the analysis of error and accuracy, which are described in Section 3.1. Second, the stay points are extracted from the indoor positioning trajectories in Section 3.2. Then, the pattern of the point drift not caused by customer behavior is analyzed, and the drift points are extracted from the stay points for the analysis of the accuracy of the indoor Wi-Fi positioning system, which is described in Section 3.3. Section 3.4 presents the accuracy analysis and calculation approach based on the drift points.

#### 3.1. Definition of Pedestrian Trajectory Patterns in Mass Indoor Positioning Data

The pedestrian trajectories contain different customer behavior patterns. To describe the different trajectory data, we define the following data for the model:

**Wi-Fi Point (WP)** is a positioning point of the pedestrian trajectories from the Wi-Fi positioning system, which contains XY coordinates and a timestamp, as shown in Figure 3.

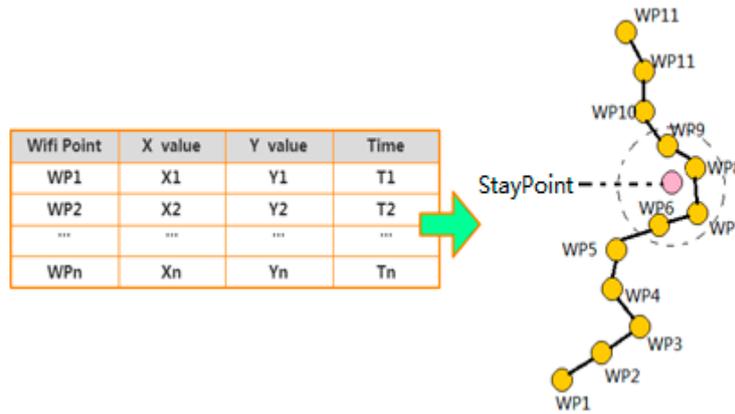


Figure 3. Wi-Fi Point (WP) and Wi-Fi Point Log (WPL).

**Wi-Fi Point Log (WPL)** is a continuous WP set where  $wpl = \{wp_1, wp_2 \dots, wp_n\}$ , in which each WP ( $WP_i$ ) belongs to WP ( $WP_i \in WP$ ), as shown in Figure 3.

**Pedestrian Trajectory (Tr)** is defined in Equation (1) as follows:

$$Tr = \{\overrightarrow{WP_i} | \overrightarrow{WP_i} = (x_i, y_i, t_i), i \in I\} \tag{1}$$

where  $\overrightarrow{WP_i}$  is defined by the  $x_i, y_i$  coordinates of the positioning point in accordance with the time-ordered  $t_i$  connection.  $Tr$  is discontinuous in time, as shown in Figure 3.

As Figure 4 shows, this visualization provides a more intuitive explanation of the concept of stay trajectories. We stipulate that, during a certain time period, customers that remain continuously within a reasonable distance and time threshold during certain activities produce a set of points called a stay trajectory ( $STr$ ), as Figure 3 shows. The centroid point in this set of dwell points in the stay trajectory is denoted as a stay point ( $SP$ ). A stay point is a collection of points representing the customer in a certain area of the mall for a continuous period of time. This includes a variety of customer behavior patterns, such as the time spent in more interesting stores. The behavior within a certain activity period is more fixed, or customers in a certain position may wait for the arrival of friends; these activities will cause behavior trajectories to form a stay, as shown in Figure 2. Therefore, we define a stay point as follows:

$$SP = \{WP_k | (x_k, y_k) \cap C_{buffer}, t_e - t_s > t_{threshold}, D < D_{threshold}, k \in I\} \tag{2}$$

$$C_{buffer} = S(CE, D_{threshold}) \tag{3}$$

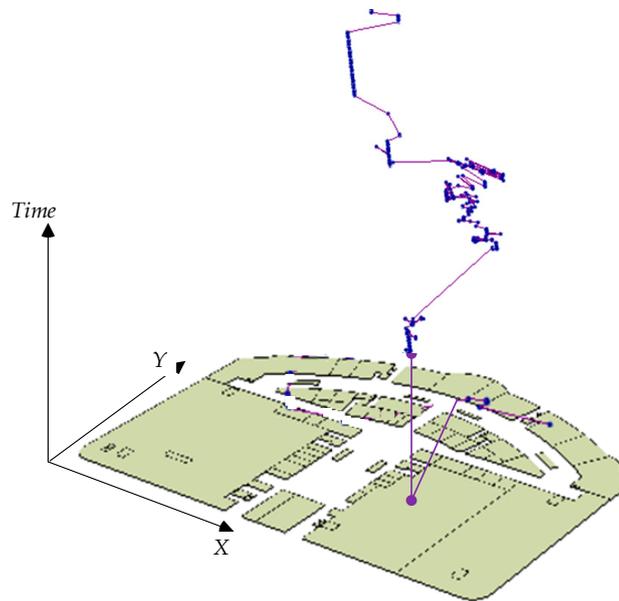
$$D = Distance \{CE_{k-1}, (x_k, y_k)\} \tag{4}$$

where  $C_{buffer}$  is a buffer constructed from the centroid (CE) that constitutes the stay point, which is the original point, and the distance threshold ( $D_{threshold}$ ) is the radius. The selection of the radius of the buffer is the result of the last convergence of the iterative time and distance thresholds where  $D$  is the distance between the updated centroid and the next anchor point,  $t_e$  is the end time of the stay trajectory, and  $t_s$  is the start time of the stay trajectory; the definition mainly includes any point in  $SP$ . In addition, the time difference must be greater than the threshold value, which is calculated as follows:

$$s_s = \{p_j, p_{j+1} | (x_j, y_j) = (x_{j+1}, y_{j+1}), t_{j+1} - t_j \gg T_s\} \tag{5}$$

where  $t_{threshold}$  is a time threshold;  $s_s$  is a set of stationary points, which appear under the same coordinates or farther apart; and the time discontinuity may be caused by a variety of factors. For example, a Wi-Fi unit may be turned off to save battery power; after a period, the staff can provide Wi-Fi access to again. Another cause could be that in the store, staff can be placed in a location in the store where they move around. Therefore, this type of stationary point will affect the

subsequent extraction work to some extent, and the stationary point set will be eliminated from the stay trajectory data.



**Figure 4.** Pedestrian Trajectory ( $Tr$ ).

### 3.2. Stay Point Extraction

There are many modes of customer behavior in shopping malls and most notably the staying mode. Thus, it is very important to extract reasonable stay trajectories. The customer activities in the mall are mainly based on the distances and times of action, so the determination of the distance and time thresholds are fundamental for extracting reasonable and effective staying trajectories and stay points. The stay point has two spatiotemporal constraints. The distance was used for the first step to preliminarily determine the potential stay points. Then, the time threshold was strictly selected to screen the final stay points.

#### 3.2.1. Distance Threshold Determination

Since the activities of the customers in the mall are complex and changeable, decisions about indoor residence are difficult. If people are engaged in a small range of activities, we can think of this behavior as a type of stay. For example, in the mall, if a store is interesting, customers in this area will find their favorite items or may wait for friends. Therefore, it is necessary to consider the two factors, distance and time, for the people who exhibit the characteristics of staying in this complex environment. Considering the pedestrian walking speed, the stay pattern should be restricted to a certain walking distance. This can be calculated by the adjacent points. The distance between the adjacent points ( $D_k$ ,  $D_i$ ) is one of the important parameters for differentiating customer behavior, as shown in Figure 5.

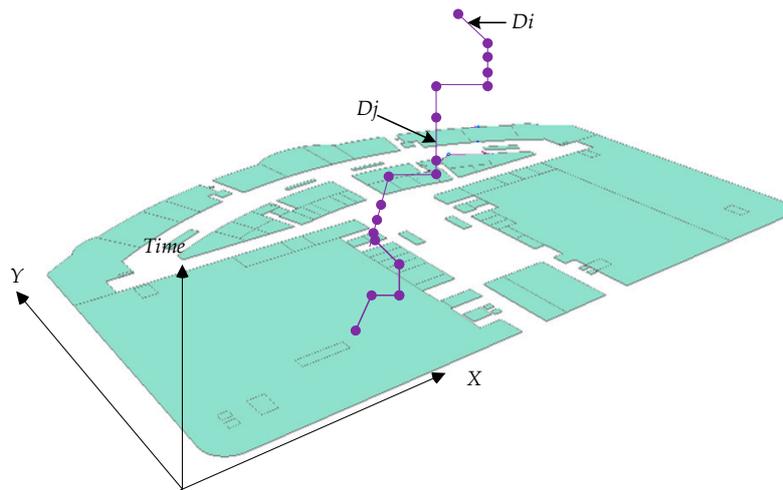


Figure 5. Distance between points.

We statistically analyzed the distances between adjacent points. We used the average distance as the threshold of the distance determination for the stay point in the first step. The distance threshold was determined by the average distance between two adjacent points in the customer trajectory dataset, as shown in Equations (6) and (7) as follows:

$$D_{mean} = \frac{\sum_{i=0}^{n-1} (D_0 + D_1 + \dots + D_{n-1})}{N - 1} \tag{6}$$

$$D_{threshold} = \frac{\sum_{i=0}^n (D_{mean0} + D_{mean1} + \dots + D_{meann})}{N} \tag{7}$$

where  $D_{mean}$  is the average of the distance between two points in a customer trajectory and  $D_{threshold}$  is the average of the extracted  $D_{mean}$ . Figure 6 shows the statistics of the distances in the whole dataset.

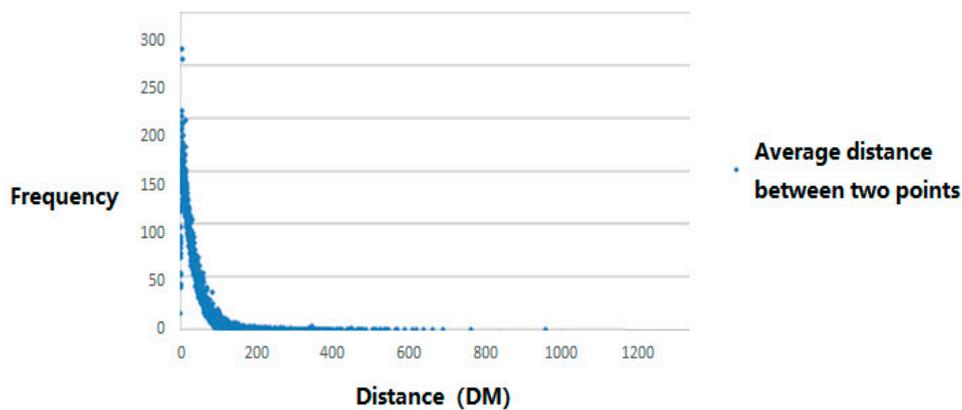


Figure 6. Distance threshold determination.

As Figure 6 shows, the distances between the two adjacent points were evenly distributed. The average distance was 3.34 m. Thus, in this paper, we chose 3.34 m as the distance threshold ( $D_{threshold}$ ) for the stay point extraction. Although this may not be the true stop point, we used further screening from the time restriction.

### 3.2.2. Time Threshold Determination

According to the results shown in the following illustration, there was a similar trend in the first 3~9 min, but the values were very different, resulting in high uncertainty. Thus, the results of the experiment cause a great deal of interference. As Figure 7 shows, the different distance thresholds showed convergence after 13 min. In addition, the trend remained steady among the thresholds. Because short periods of activity in the mall are very common and complex behaviors, it is difficult to effectively define these short-term activities; thus, a short stay is required to achieve a high level of positioning accuracy, and other auxiliary information can be used to carry out efficient extraction analysis. Therefore, we extracted the stay trajectory and the time threshold of the point of stay at 13 min.

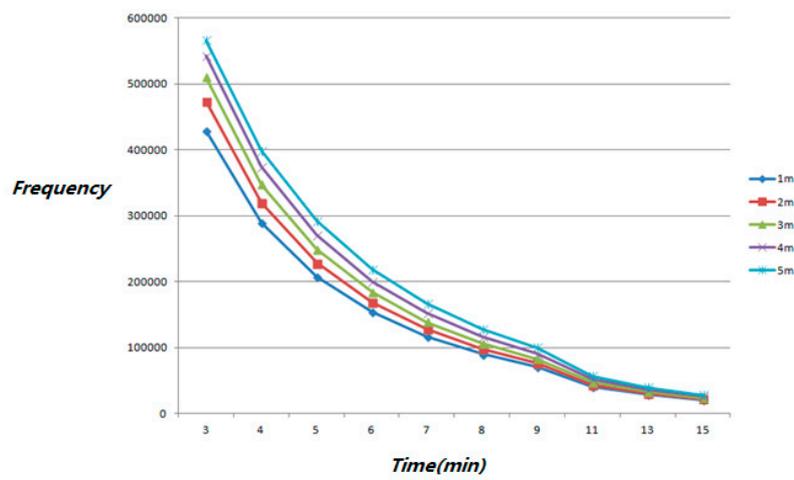


Figure 7. Time threshold determination.

### 3.2.3. The Algorithm for the Extraction of Stay Points from the Staying Trajectories

Through Algorithm 1, which is presented below, these dwell points can be extracted from the algorithm if the customer stays in an area for more than a certain time period. For example, if the customer remains within a 3.34 m range and stays more than 13 min, we record each anchor point between the start time and departure time and at the center of the centroid and then calculate the trajectory coordinates. Before the extraction of the stay trajectories extraction, some intermittent trajectories with limited travel times and large distance intervals were removed from the massive dataset of trajectories. Over repeated iterations, there was a relatively high confidence among the stay trajectory and stay points.

This algorithm for extracting the stay points considers the customer behavior in the mall and constantly updates the centroid to determine whether the customer movement during a period of action exceeds the threshold. Since time and space are continuous, the extraction of a point with this methodology can ensure that the extracted data have high accuracy.

**Algorithm 1: Stay Point\_Detection (WP, distThreh, timeThreh, CE)**

Input: A Wi-Fi point log WP, a distance threshold *distThreh* and time span threshold *timeThreh*, an updated centroid CE, the initial CE was set as the starting point of a complete pedestrian trajectory.

Output: A set of stay points  $SP = \{WP\}$

Orderby: Wi-Fi record time T

1. Loop: All pedestrian positioning records
2. **While**  $j < \text{pointNum}$  **do**
3.    $j = j + 1$
4.   **While**  $j < \text{pointNum}$  **do**
5.     Calculate the distance between CE and Wi-Fi:  $\text{Distance}(\text{CE}, p_j)$
6.     **If**  $\text{dist} < \text{distThreh}$ , **then**
7.       Update CE ( $p_0, p_1, \dots, p_j$ )
8.       Calculate the time span between two Wi-Fi points as  $\text{diffTime} = p_j.T - p_{j-1}.T$
9.       **If**  $\text{diffTime}$  is Continuous-time **then**
10.          $\text{SP.insert}(WP_j)$
11.          $j = j + 1$
12.       **Else**
13.          $\text{SP.time} = \text{SP.Endtime} - \text{SP.Starttime}$
14.         **If**  $\text{SP.time} > \text{timeThreh}$  **then**
15.           Clear CE.value
16.          $j = j + 1$
17.       **Return** SP

### 3.3. Drift Point Extraction Based on Noncustomer Behavior Patterns

#### 3.3.1. Drift Phenomenon in Stay Points

Despite the limitations of the algorithm, the stay trajectories could be obtained with a high degree of credibility, and we found that there was a high-frequency phenomenon of data coordinate drift. We called this phenomenon a drift point. A point with drift phenomenon shows a trajectory from the previous point location within a short period of time. The distance between these coordinates is large during the short stay, after which the customer returns to the normal range of customer activities. First, the existence of a drift point suggests that it is part of a stay pattern. Then, during the stay period, some points drift. In addition, the drift point is continuous in time with the positioning points that are recorded before and after its movement. This kind of drift represents a system error in the Wi-Fi positioning.

In the process of extracting the drift points, there were some obstacles and problems, including whether the drift in positioning data was actually a real behavior; for example, when purchasing a drink inside the mall, a customer will visit a vending machine, stay for a short period of time, and then return to their original position. Therefore, we distinguish between the types of drift phenomena, as shown in Figure 8.

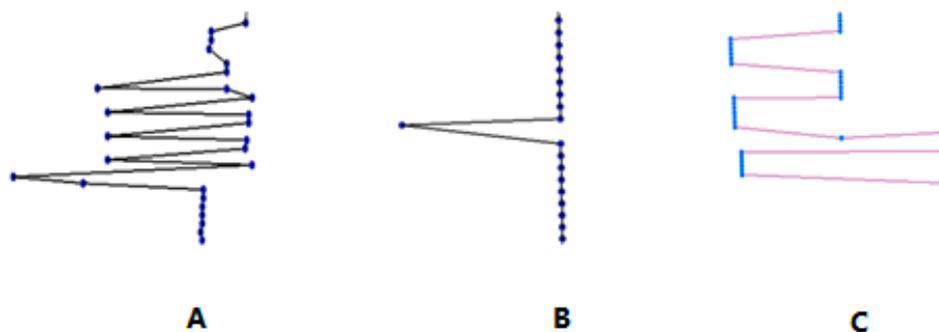


Figure 8. Types of drift phenomenon in the stay trajectories.

**Drift Phenomenon A:** Repeated jumping of the trajectory inside the mall and short stay periods account for a certain proportion of the customer positioning data, but this type of behavior trajectory cannot be used to determine whether the system caused the drift or if the true customer behavior is reflected. Thus, in the drift point extraction algorithm, such behavior characteristics are not interpreted as the characteristic of the drift points caused by Wi-Fi positioning error.

**Drift Phenomenon B:** During a stay period, there is an occurrence of abnormal data movement, but after a short time, the trajectory jumps back to the original region. This type of behavior is a non-customer behavior that is characteristic of the typical drift points caused by Wi-Fi positioning errors. Thus, some arithmetic has been added to the drift point extraction algorithm to restrict the positioning data to ensure that the extraction of the drift points is objective and real.

**Drift Phenomenon C:** This type of drift phenomenon stays in several places repeatedly. The stay time is not long; therefore, for this type of drift point characteristic pattern, it is also difficult to determine whether this activity is the customer's real behavior pattern or the result of the environmental factors. Therefore, this type of drift phenomenon is not a characteristic pattern caused by Wi-Fi positioning error.

Only points exhibiting **Drift Phenomenon B** are showing a noncustomer behavior and can be confirmed to be drift points (DP) for the accuracy assessment of the indoor Wi-Fi positioning system.

### 3.3.2. Extraction of Drift Points from Stay Points

Therefore, we extracted the drift points that showed the characteristics of **Drift Phenomenon B**. As shown in Figure 9, a customer has data records in areas A, B, and C. The areas A and C are closer together, and there are consecutive time records. The customer moved a distance between A and B, stayed for a short time at location B, and then returned to area A immediately. After a period of time, the customer moved from A to area C and stayed there for some time.

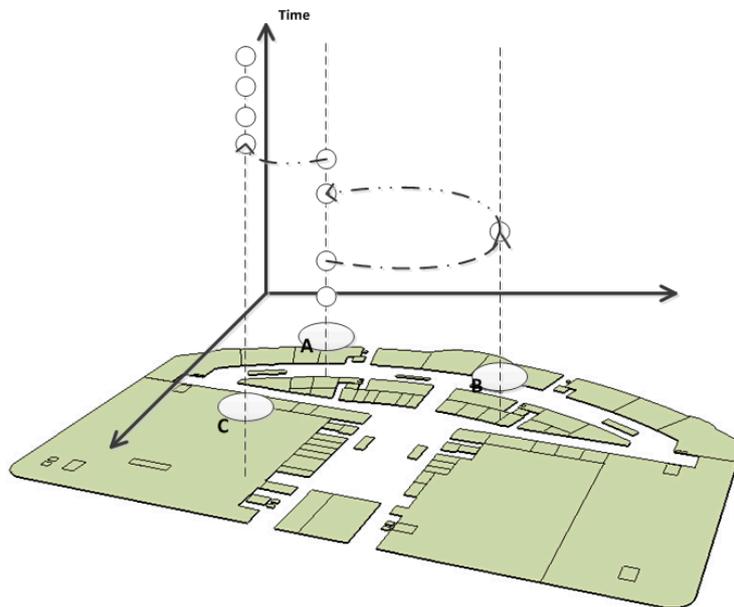


Figure 9. Sketch map of the drift phenomenon.

Therefore, we defined the drift points as follows:

$$DP = \{p_k, p_{k-1} | t_k - t_{k-1} < t_{threshold2}, Distance(p_k, p_{k-1}) > D_{threshold}, K \in I\} \quad (8)$$

where the definition of  $DP$  (drift points) creates limits for the distance and time thresholds. The time difference between  $t_k$  and  $t_{k-1}$  is used to determine the temporal continuity. The distance between  $P_k$  and  $P_{k-1}$  is compared with the action  $D_{threshold}$ . Therefore, it is found that the trajectory is a drift phenomenon, and the drift phenomenon is not the real behavior of the customer.

According to the order of the times, the data are arranged sequentially, the distance between the current positioning point and the previous anchor point is computed, and if the distance threshold is smaller than before the extraction, the point is judged as a stay and enters the trajectory judgment process. If the distance is greater than the threshold value, it enters the drift point judgment process. Finally, the staying trajectory with the drift points is extracted. The distance threshold of the drift point is selected as the time and distance between the stay point and the drift point. If the result is greater than the human walking speed of 1.5 m/s, it exceeds the drift point distance threshold, and the time threshold of less than one minute is used to extract the drift point. The detail procedure of drift points (DP) detection is given in Algorithm 2.

---

**Algorithm 2 Drift Point Detection ( $WP, distThreh, timeThreh, CE, driftdistThreh, drifttime, Threh$ )**

---

**Input:** A Wi-Fi point log  $WP$ , a distance threshold  $distThreh$  and time span threshold  $timeThreh$ , an updated centroid  $CE$ , a distance threshold of drift points and time span threshold of drift points

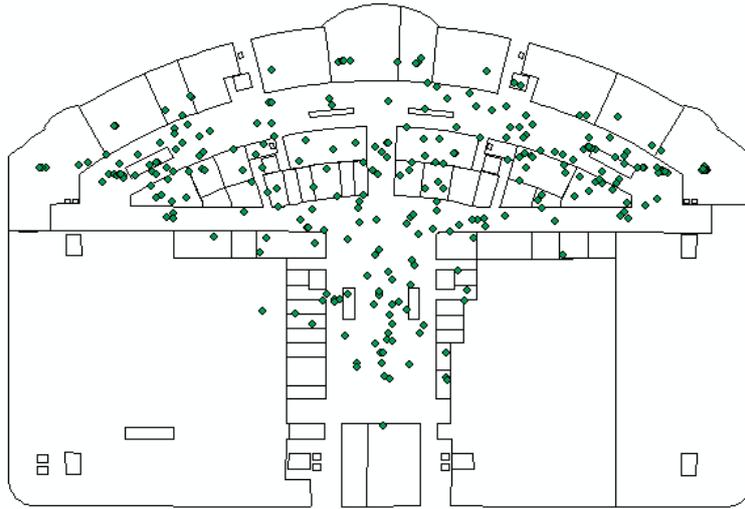
**Output:** A set of stay points and drift points

**Orderby:** Wi-Fi record time  $T$

1. Loop: All pedestrian positioning records
  2. **While**  $j < pointNum$  **do**
  3.      $J = j + 1$
  4.     **While**  $j < pointNum$  **do**
  5.         Calculate the distance between  $CE$  and Wi-Fi point:  $dist = Distance(CE, p_j)$
  6.         Calculate the time span between two Wi-Fi points:  $diffTime = p_j.T - p_{j-1}.T$
  7.         Calculate the time span between drift point and stay point:  $driftdiffTime = p_j.T - p_{j-i}.T$
  8.         Calculate the time span between drift point and stay point:  $backstayPointdist = p_j.T - p_{j-i}.T$
  9.         **If**  $dist < distThreh$  **then**
  10.             Update  $CE(p_0, p_1..p_j)$
  11.             **If**  $diffTime$  is Continuous-time **then**
  12.                  $DP.insert(WP_j)$
  13.                  $j = j + 1$
  14.             **Elif**  $dist > driftdistThreh$  **and**  $diffTime$  is Continuous-time
  15.                 **and**  $driftdiffTime < drifttimeThreh$
  16.                  $i = i + 1$
  17.             **Elif**  $backstayPointdist < distThreh$  **and**  $diffTime$  is Continuous-time
  18.                  $DP.insert(WP_j, WP_{j-1}, \dots, WP_{j-i})$
  19.                  $DP.time = DP.Endtime - DP.Starttime$
  20.                  $i = 0$
  21.             **If**  $DP.time > timeThreh$  **then**
  22.                 Clear  $CE.value$
  23.                  $j = j + 1$
  24.     **Return**  $DP$
- 

Figure 10 gives the results of the extraction of the drift points from stay points with the proposed Algorithm 2.

As Figure 10 shows, in indoor places, the drift points are more scattered. Therefore, the drift phenomenon was not limited to a particular region; rather, Drift Phenomenon B was found throughout the entire indoor area. This result was consistent with the random distribution characteristics of the errors from the indoor Wi-Fi positioning system.



**Figure 10.** Distribution of drift points (DP).

### 3.4. Accuracy Analysis Based on Drift Points

We use the extracted drift points (DP) and stay points (SP) for the accuracy assessment of the indoor Wi-Fi positioning system, as shown in Figure 11. The standard deviations (*SD*) and averages (*AVE*) of all stay points as calculated in Equations (9)–(12) are taken as the random positioning error of the indoor Wi-Fi positioning system. The average offset of drift points (*AoD*) and the standard deviation of drift points (*SDoD*) to the centroid are taken as the system error of the indoor Wi-Fi positioning system, as calculated in Equation (11). These metrics are calculated as follows:

$$SD = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2} \quad (9)$$

$$AVE = \mu = \frac{\sum_{i=1}^N x_i}{N} \quad (10)$$

$$AoD = \mu_d = \frac{\sum_{i=1}^N d_i}{N_d} \quad (11)$$

$$SDoD = \sqrt{\frac{1}{N_d} \sum_{i=1}^N (d_i - \mu_d)^2} \quad (12)$$

where  $N$  is the total number of stay points and  $N_d$  is the number of drift points,  $x_i$  is the distance from each point to the centroid of stay points,  $\mu$  is the average of  $x_i$ ,  $d_i$  is the distance from each drift point to the centroid of the stay point and  $\mu_d$  is the average offset of the drift points.

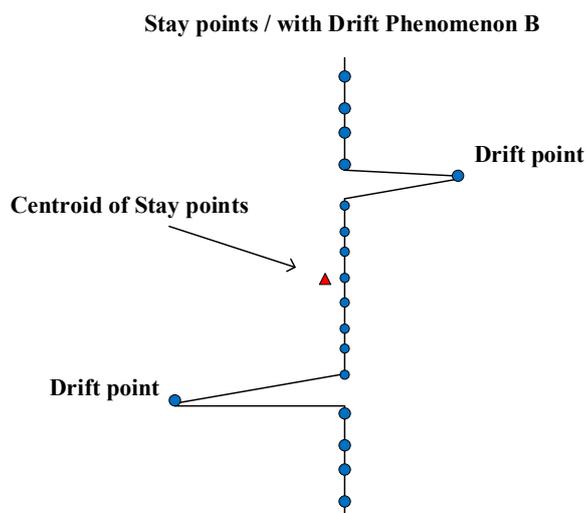


Figure 11. Error and accuracy calculations with the stay points and drift point.

### 4. Experiments and Analysis

#### 4.1. Error and Accuracy of the Indoor Wi-Fi Positioning System

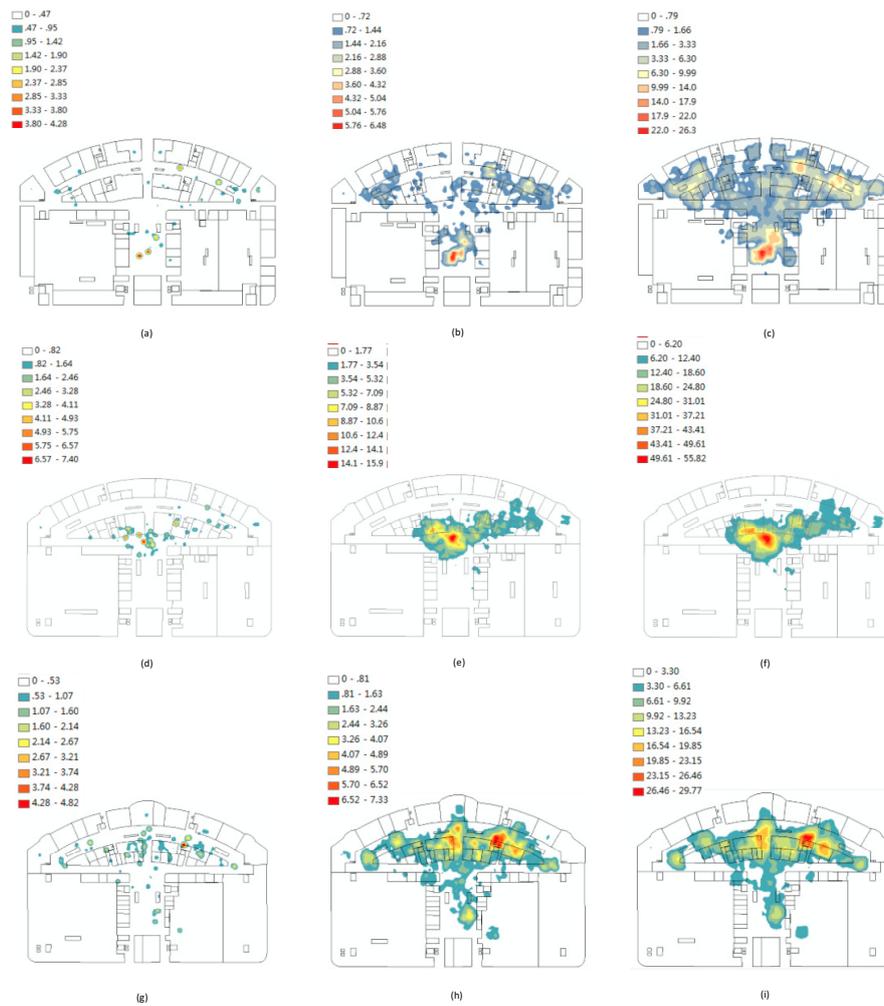
We conducted a detailed analysis of the three floors of the shopping mall and produced statistics for three periods from the dataset, including time, count of MAC (*CM*), count of drift points (*CDP*), average error (*AVE*), standard deviation (*SD*), and the average offset of drift points (*AoD*). The periods examined were one day, one week, and one month. The count of MAC (*CM*) was based on the number of users within the mall. The count of the drift points (*CDP*) was the number of drift points extracted from stay points. The statistical results of accuracy are shown in Table 1.

Table 1. Detailed statistical results of the indoor positioning accuracy.

Floor	Time	CM	CDP	AVE (m)	SD (m)	AoD (m)	SDoD (m)
Ground Floor	1 day	35,478	463	3.21	2.43	9.12	1.16
	1 week	210,562	3126	3.32	2.22	9.11	0.96
	1 month	1,015,468	13,659	3.09	2.35	8.89	1.12
Second Floor	1 day	32,156	421	3.11	2.54	9.12	0.96
	1 week	200,456	2965	3.89	2.23	9.56	0.89
	1 month	1,008,632	12,654	3.81	2.41	8.95	1.07
Third Floor	1 day	38,456	481	3.65	2.33	9.66	1.13
	1 week	226,544	3248	3.65	2.47	9.43	0.98
	1 month	1,125,986	14,025	3.42	2.69	9.13	1.03

As Table 1 shows, the average error was within 4 m, and the standard deviation was approximately 2.5 m, which indicated the accuracy of the indoor Wi-Fi positioning system. The average offset of drift points was approximately 9 m, which indicated a system error in the indoor Wi-Fi positioning system. However, the standard deviation was approximately 1 m, indicating that the system error was relatively stable for the system compensation and correction.

Figure 12 presents the kernel destiny of the drift points on different floors and during different time periods.



**Figure 12.** Results of the kernel density analysis of the indoor Wi-Fi position data.

Figure 12a–c presents the statistical analysis of the first floor for one day, one week, and one month, respectively. It is clear that the drift points on the first floor were scattered and that the peak value was near the exit. Figure 12d–f shows the statistical analysis of the second floor for one day, one week, and one month, respectively. The distribution of the points on the second floor were more centralized, and the points were mainly located in the central area of the second floor. The high flow density of the area caused the positioning errors. Figure 12g–i visualizes the statistical analysis of the third floor for one day, one week, and one month, respectively. The stores on the third floor featured a food court, so the peak of the error was concentrated on the food court in the central area.

#### 4.2. Analysis of the Spatial Accuracy of the Indoor Wi-Fi Positioning System

For the indoor surroundings, the error and accuracy of the Wi-Fi positioning system were related to the indoor structure. The indoor structure of a shopping mall determines the distribution of the shops, which leads to different crowd densities. In addition, the indoor structure determined the layout of the AP sensors, which led to different signal intensities of the indoor Wi-Fi system. Thus, the crowd density and signal intensity will affect the spatial accuracy of the indoor Wi-Fi positioning system. We further analyzed the spatial error and accuracy of the indoor Wi-Fi positioning system from two perspectives.

#### 4.2.1. Relationship between Crowd Density and Indoor Positioning Error

To analyze the correlation between the crowd density and indoor positioning error, the six million data recordings in the stores on one day on each floor were used as examples for a perform linear regression and correlation analysis. The result is shown in Table 2.

**Table 2.** Correlation analysis between the crowd density and the error in the stores.

	Model Summary		Coefficients		
	Adjusted $R^2$	Std.	Model	$T$	Sig.
1st Floor	0.977	33.698	Crowd Density	182.658	0.03
2nd Floor	0.931	38.365	Crowd Density	188.634	0.04
3rd Floor	0.948	35.291	Crowd Density	186.889	0.01

In Table 3, we can see that the value of  $T$  was the result of the  $t$ -test of the regression coefficient. The larger the absolute value was, the smaller the  $sig$  was, where  $sig$  represents the significance of the  $t$ -test. In statistics,  $sig < 0.05$  is generally considered to represent a significant test coefficient. The absolute value of the regression coefficient was significantly greater than zero, indicating that the independent variable could effectively predict the variation in the dependent variable. The statistical results showed that the volume of people in a shop affected the error value, as illustrated in Figure 13.

**Table 3.** Correlation analysis between the crowd density and error outside the store.

	Model Summary		Coefficients		
	Adjusted $R^2$	Std.	Model	$T$	Sig.
1st Floor	0	99.312	Crowd Density	0.461	0.791
2nd Floor	0	96.567	Crowd Density	0.369	0.801
3rd Floor	0	91.226	Crowd Density	0.335	0.737



**Figure 13.** Statistical results of the crowd density and the error in the stores.

As Figure 13 shows, in the stores, the crowd density and the positioning error were positively correlated. The high crowd density could easily result in error. The higher the crowd density was, the greater the error was. Thus, we conducted a statistical experiment in the area outside of the indoor stores. First, we evenly divided and orderly numbered the area outside of the indoor stores. Then, we performed a spatial join with the grid, the recorded points and the error points of the customer and performed a statistical analysis of the relationship between the flow density of the area outside the stores and the error. We also conducted a linear regression and correlation analyses with the observational data. The results are shown in Table 3.

As Table 3 shows, we can see that the correlation between the area outside the store and the error was low, so the flow density of area outside the store had less of an influence on the error.

#### 4.2.2. Relationship between the AP Sensors and Indoor Positioning Error

It is hard to obtain ground truth data of the trajectories. The method proposed in this paper was to assess the actual accuracy of the indoor Wi-Fi position and to improve the layout of the Wi-Fi sites. Thus, we introduced a theoretical model of Wi-Fi positioning based on the Wi-Fi sites to generate an error map as the ground truth data. Given that the floor plans of different stores are complex, the communication strength of the AP sensor varies greatly. Therefore, we verified the error and accuracy of the indoor Wi-Fi positioning system according to the distribution of the AP sensors.

Inside the shopping mall, the AP sensors (Aruba ap-103) were arranged uniformly. Taking the average nearest distance between the different AP sensors as a search radius, the number of other AP points appearing within the search radius of each AP point is multiplied by a coefficient (0.5 in the store and 1 outside the store). Then, the results are treated with a weight  $Z_i$ , and an inverse distance-weighted interpolation is performed using the AP locations and the  $Z_i$  weight, which is shown in Equation (13) as follows:

$$W_i = \frac{h_i^{-p}}{\sum_{j=1}^n h_j^{-p}} \times Z_i \quad (13)$$

where  $P$  is an arbitrary positive number and is usually set to 2.  $H_i$  is the distance between the AP sensor and the interpolating point, as in Equation (14) as follows:

$$h_i = \sqrt{(x - x_i)^2 + (y - y_i)^2} \quad (14)$$

where  $(X, Y)$  is the coordinate of the interpolating point, and  $(X_i, Y_i)$  is the coordinate of the AP sensor.

Figure 14 gives the predicted positioning errors with respect to the AP sensors calculated by the inverse distance-weighted method and the actual errors of the indoor Wi-Fi positioning system with our proposed approach.

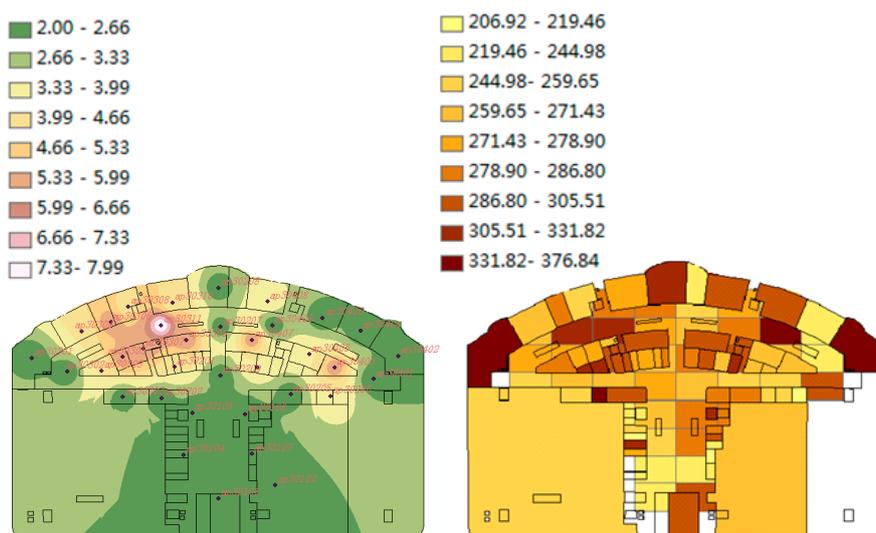


Figure 14. Comparison of the predicted and actual errors.

As Figure 14 shows, the error was small in the areas with higher scores for the sensors and was especially high at the left and right ends of the floor plan, where the error was particularly prominent. Combined with the distribution of the AP sensors, it was found that the distribution of the AP sensors had a great influence on the production of the error.

In addition, we randomly selected samples in nine different areas of the weighted distribution to study the relationship between the distances between the AP sensors and the positioning error. Figure 15 gives the statistical results of the accuracy assessment of the average distance of the sample points to the three nearby sensors.

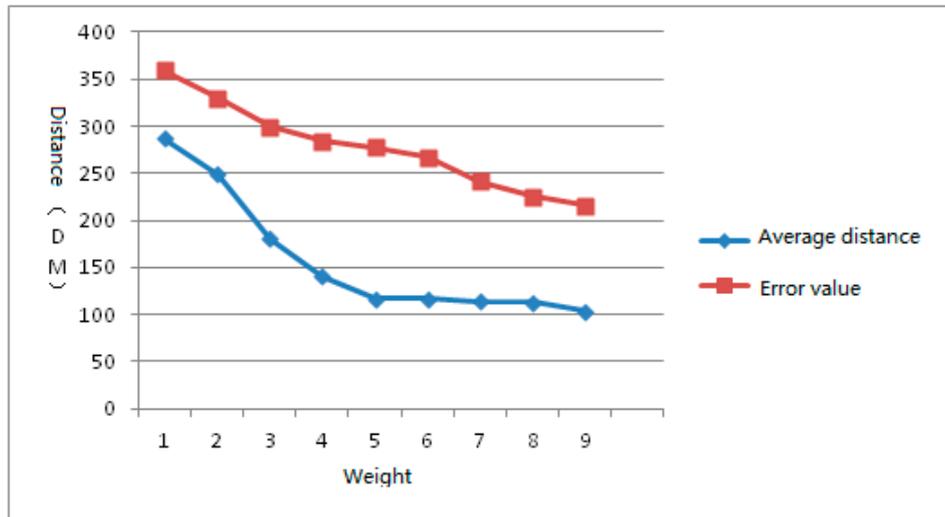


Figure 15. The average distance between the three nearby APs and error statistics of the sample points.

As Figure 15 shows, with increasing weight level, the average distance between the sample point and the nearby AP sensors as gradually reduced, and the positioning error was gradually reduced. Thus, the locations of the AP sensor will affect the error of indoor positioning systems. This result can also verify that the error and accuracy assessment with our proposed approach was correct and effective.

### 5. Discussion

According to pattern analysis of the indoor pedestrian trajectory data from the Baolong Shopping Mall in Xinxiang city, we can not only acquire the set of stay points with a high degree of confidence but also provide a scientific and effective approach to the stores in the shopping mall in the form of the stay points dataset and propose effective improvement schemes from a security standpoint. For example, we evaluated the aggregations of people staying still at different times and the flow densities in each area to avoid the casualties caused by emergencies. As shown in Figure 16, the experiments were conducted on the first floor of the shopping mall on October 1st and September 26th, and we compared the data. Because October 1st is National Day, the visitor flow rate was relatively large compared with that on nonholidays. Therefore, the kernel density analysis was conducted using the extracted drift points.

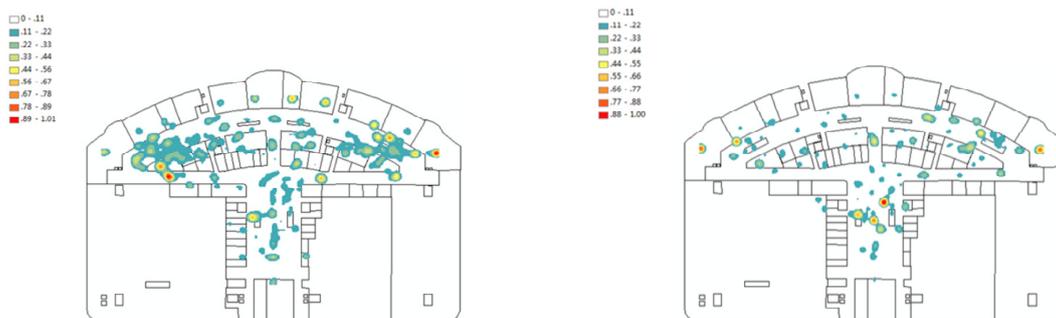


Figure 16. Record of October 1st and September 26th.

It can easily be seen that the coverage area and the numerical values of the former dates were larger than those of the latter, and the MAC record on October 1st contained nearly 69,812 records, while there were only 34,354 records for September 26th. It was clear that the customer flow on that day was different. In some particular areas, a high visitor flow rate was accompanied by a high error. Therefore, from the extraction and correction of the drift points, the stay points could be merged to improve the positioning system. It can be helpful to provide such a scientific and reasonable improvement project to businesses. For example, if businesses were allowed to improve their AP positioning to enhance the accuracy of indoor positioning, the design structure of the mall would not affect the signal transmission as much. This measure can not only guarantee data collection but also will keep the doors open. By comparing the times that customers entered and exited a store, we found that there were two main factors that affected the accuracy of the indoor positioning in the shopping mall: the internal structure and the distribution of the AP sensors. In the future, we will continue to explore the influence of the building materials on positioning accuracy. We can overlap the real accuracy map with the proposed trajectory analysis with the accuracy calculation result from the theoretical model of the indoor Wi-Fi positioning system. This comparison can provide strong support for exploring reasonable schemes to improve the error in the future.

In this paper, we analyzed the error of indoor positioning data by extracting the drift points, which was not only a good evaluation method, but also a new point of view for assessing indoor positioning error. The accuracy assessment result was related to the drift points extracted from stay points. However, customers may also move within stores when they stop while shopping; thus, the stay behavior patterns of the customers while shopping has a relatively fixed position. Although the determination of the drift points considers both the time restriction and distance restriction, the movement of some customers around the store appears similar to drift, which will affect the error calculation of the indoor positioning system. Combining more than two drifts in the stay points would increase the reliability of the precision of the computation. There is a large gap between the staying time and distance in indoor positioning data and outdoor GNSS data. The collection of outdoor GNSS positioning data is often based on travel habits. Therefore, people's action distances are usually measured from the location of their home, their office or entertainment venues and there are a variety of choices of modes of travel, and the rate of coverage of the outdoor GNSS positioning data in one day can usually cover all travel in daily life and work. Therefore, the timescale and distance of GNSS outdoor positioning is much larger than that for indoor positioning. In the analysis of the activities of people indoors, the accuracy of the extraction of the stay points from indoor positioning data needs to be more sophisticated to produce content that is more refined. It will also be important to conduct further studies in the future and to explore the trajectories of pedestrians.

In addition, we can combine this approach with GNSS to achieve seamless positioning of customer location information to obtain complete behavior trajectories as customers move from indoors to outdoors and to achieve a more detailed study of customer behavior in the future. For example, this approach would be useful in residential and retail areas, to name a few. Based on the behavior observed in the shopping mall and the outdoor behavior patterns, we can acquire abstract characteristics of and attribute information about the customer. In the future, other information can be selected to meet customer needs according to these characteristics and attributes. We can calculate when the user accesses the POI (point of interest) based on the user's trajectory data. In the process of the extraction and analysis of indoor positioning data, this information can also be combined with other data (credit card records, video surveillance, registered brand member information, etc.) to achieve precise estimations of the needs of the user. Current research shows that credit card information has potential in urban planning for combating disease: indoor positioning data, in combination with specific information on stores and local weather information, can be used to predict outbreaks of disease.

## 6. Conclusions

Although the accuracy of indoor positioning technology continues to improve, there are still many phenomena that are easily overlooked for the positioning algorithm. For example, the frequency of drift phenomena influences the real location information. In addition, we still lack effective methods and theories for evaluating positioning error. These factors affect location-based services and applications with indoor positioning big data. In this paper, we proposed an original method for evaluating the uncertainty range of indoor positioning based on the Wi-Fi access points deployed in a shopping mall. We also provide a clear definition of stay points and drift points in indoor positioning data. After preprocessing the original data, we performed a scientific and effective extraction of the set of stay points and drift points via an algorithm. The experiments showed that the above definitions and algorithms can be used to evaluate the error and accuracy of indoor Wi-Fi positioning systems and can be applied to a wider range of applications. We found that the formation of these errors was closely related to the locations of APs and to the internal structure of the shopping mall. We believe that this approach constitutes a promising research direction for improving the accuracy of indoor positioning data through the analysis of the error of indoor positioning data. In future studies, through research on improving indoor positioning accuracy, this approach can ensure that the feature extraction of behavioral patterns is accurate.

**Author Contributions:** Methodology, Qingwu Hu and Dengbo Yu; investigation, Dengbo Yu; resources, Shaohua Wang; writing, Dengbo Yu; writing—review and editing, Qingwu Hu; supervision, Qingwu Hu and Shaohua Wang.

**Funding:** This research was funded the Science and Technology Planning Project of Guangdong, China (Grand No. 2017B020218001) and the Fundamental Research Funds for the Central Universities (Grant no. 2042017kf0235).

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

## References

1. Lee, S.J.; Min, C.; Yoo, C.; Song, J. Understanding customer malling behavior in an urban shopping mall using smartphones. In Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication, Zurich, Switzerland, 8–12 September 2013; pp. 901–910.
2. Jibiki, K.; Iwata, S.; Tamagawa, S. A study of AED layout method in shopping mall based on the customers' subconsciousness. *AIJ J. Technol. Des.* **2017**, *23*, 643–648. [[CrossRef](#)]
3. Pipelidis, G.; Moslehi Rad, O.R.; Iwaszczuk, D.; Prehofer, C.; Hugentobler, U. Dynamic Vertical Mapping with Crowdsourced Smartphone Sensor Data. *Sensors* **2018**, *18*, 480. [[CrossRef](#)] [[PubMed](#)]
4. Correa, A.; Barcelo, M.; Morell, A.; Vicario, J.L. A Review of Pedestrian Indoor Positioning Systems for Mass Market Applications. *Sensors* **2017**, *17*, 1927. [[CrossRef](#)] [[PubMed](#)]
5. Guvenc, I.; Chong, C.C. A Survey on TOA Based Wireless Localization and NLOS Mitigation Techniques. *IEEE Commun. Surv. Tutor.* **2009**, *11*, 107–124. [[CrossRef](#)]
6. Zhao, Q.; Wang, J.; Liu, Y.; Yang, X. System error correction based on particle swarm optimization in TOA indoor location. *Electron. Meas. Technol.* **2017**, *40*, 189–192.
7. Akgul, F.O.; Pahlavan, K. A Novel Statistical AOA Model Pertinent to Indoor Geolocation. *J. Geogr. Inf. Syst.* **2010**, *2*, 45–48. [[CrossRef](#)]
8. Alarifi, A.; Al-Salman, A.; Alsaleh, M.; Alnafessah, A.; Al-Hadhrani, S.; Al-Ammar, M.A.; Al-Khalifa, H.S. Ultra Wideband Indoor Positioning Technologies: Analysis and Recent Advances. *Sensors* **2016**, *16*, 707. [[CrossRef](#)]
9. Bocquet, M.; Loyez, C.; Benlarbi-Delai, A. Using enhanced-TDOA measurement for indoor positioning. *IEEE Microw. Wirel. Compon. Lett.* **2005**, *15*, 612–614. [[CrossRef](#)]
10. Xie, T.; Zhang, C.; Li, Y.; Jiang, H.; Wang, Z. An enhanced TDoA approach handling multipath interference in Wi-Fi based indoor localization systems. In Proceedings of the International Midwest Symposium on Circuits and Systems, Boston, MA, USA, 6–9 August 2017; pp. 160–163.

11. Mazuelas, S.; Bahillo, A.; Lorenzo, R.M.; Fernandez, P.; Lago, F.A.; Garcia, E.; Blas, J.; Abril, E.J. Robust Indoor Positioning Provided by Real-Time RSSI Values in Unmodified WLAN Networks. *IEEE J. Sel. Top. Signal Process.* **2009**, *3*, 821–831. [[CrossRef](#)]
12. Passafiume, M.; Maddio, S.; Cidronali, A. An Improved Approach for RSSI-Based only Calibration-Free Real-Time Indoor Localization on IEEE 802.11 and 802.15.4 Wireless Networks. *Sensors* **2017**, *17*, 717. [[CrossRef](#)]
13. Zhang, W.; Hua, X.; Yu, K.; Qiu, W.; Chang, X.; Wu, B.; Chen, X. Radius based domain clustering for Wi-Fi indoor positioning. *Sens. Rev.* **2017**, *37*, 54–60. [[CrossRef](#)]
14. Qin, J.; Sun, S.; Deng, Q.; Liu, L.; Tian, Y. Indoor Trajectory Tracking Scheme Based on Delaunay Triangulation and Heuristic Information in Wireless Sensor Networks. *Sensors* **2017**, *17*, 1275. [[CrossRef](#)]
15. Pastell, M.; Frondelius, L.; Järvinen, M.; Backman, J. Filtering methods to improve the accuracy of indoor positioning data for dairy cows. *Biosyst. Eng.* **2018**, *169*, 22–31. [[CrossRef](#)]
16. Walters, C. Characterization of Smart Phone Received Signal Strength Indication for WLAN Indoor Positioning Accuracy Improvement. *J. Netw.* **2014**, *9*, 1061–1065.
17. Papaioannou, S.; Wen, H.; Markham, A.; Trigoni, N. Fusion of Radio and Camera Sensor Data for Accurate Indoor Positioning. In Proceedings of the International Conference on Mobile Ad Hoc and Sensor Systems, Philadelphia, PA, USA, 28–30 October 2014; pp. 109–117.
18. He, S.; Chan, S.H.G. Sectjunction: Wi-Fi indoor localization based on junction of signal sectors. In Proceedings of the IEEE International Conference on Communications, Sydney, NSW, Australia, 10–14 June 2014; pp. 2605–2610.
19. Yang, S.; Dessai, P.; Verma, M.; Gerla, M. FreeLoc: Calibration-free crowdsourced indoor localization. In Proceedings of the IEEE INFOCOM, Turin, Italy, 14–19 April 2013; pp. 2481–2489.
20. Lymberopoulos, D.; Liu, J.; Yang, X.; Choudhury, R.R.; Sen, S.; Handziski, V. Microsoft Indoor Localization Competition: Experiences and Lessons Learned. *GetMobile* **2015**, *18*, 24–31. [[CrossRef](#)]
21. Farshad, A.; Li, J.; Marina, M.K.; Garcia, F.J. A microscopic look at Wi-Fi fingerprinting for indoor mobile phone localization in diverse environments. In *International Conference on Indoor Positioning and Indoor Navigation*; IEEE: Busan, Korea, 2013; pp. 1–10.
22. Liu, Y.; Yang, Z.; Wang, X.; Jian, L. Location, localization, and localizability. *J. Comput. Sci. Technol.* **2010**, *25*, 274–297. [[CrossRef](#)]
23. He, S.; Chan, S.H.G. Wi-Fi Fingerprint-Based Indoor Positioning: Recent Advances and Comparisons. *IEEE Commun. Surv. Tutor.* **2017**, *18*, 466–490. [[CrossRef](#)]
24. Wallbaum, M.; Diepolder, S. Benchmarking wireless lan location systems wireless lan location systems. In Proceedings of the Second IEEE International Workshop on Mobile Commerce and Services, Munich, Germany, 19 July 2005; pp. 42–51.
25. Niu, J.; Lu, B.; Cheng, L.; Gu, Y.; Shu, L. ZiLoc: Energy efficient Wi-Fi fingerprint-based localization with low-power radio. In Proceedings of the 2013 IEEE Wireless Communications and Networking Conference, Shanghai, China, 7–10 April 2013; pp. 4558–4563.
26. Ji, Y.; Biaz, S.; Wu, S.; Qi, B. Optimal Sniffers Deployment on Wireless Indoor Localization. In Proceedings of the International Conference on Computer Communications and Networks, Honolulu, HI, USA, 13–16 August 2007; pp. 251–256.
27. Males, J.R.; Worrell, G.C. Wireless Based Positioning Method and Apparatus. U.S. Patent 8081991B2, 20 December 2011.
28. Nishida, K.; Toda, H.; Kurashima, T.; Suhara, Y. Probabilistic identification of visited point-of-interest for personalized automatic check-in. In Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Washington, DC, USA, 13–17 September 2014; pp. 631–642.
29. Nuaimi, K.A.; Kamel, H. A survey of indoor positioning systems and algorithms. In Proceedings of the International Conference on Innovations in Information Technology, Abu Dhabi, UAE, 25–27 April 2011; pp. 185–190.
30. Liu, H.; Gan, Y.; Yang, J.; Sidhom, S.; Wang, Y.; Chen, Y.; Ye, F. Push the limit of Wi-Fi based localization for smartphones. In Proceedings of the 18th Annual International Conference on Mobile Computing and Networking, Istanbul, Turkey, 22–26 August 2012; pp. 305–316.

31. Xie, H.; Gu, T.; Tao, X.; Ye, H.; Lv, J. MaLoc: A practical magneticfingerprinting approach to indoor localization using smartphones. In Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Washington, DC, USA, 13–17 September 2014; pp. 243–253.
32. Xiao, Z.; Wen, H.; Markham, A.; Trigoni, N.; Blunsom, P.; Frolik, J. Non-Line-of-Sight Identification and Mitigation Using Received Signal Strength. *IEEE Trans. Wirel. Commun.* **2014**, *14*, 1689–1702. [[CrossRef](#)]
33. Seco, F.; Jiménez, A.R.; Prieto, C.; Roa, J.; Koutsou, K. A survey of mathematical methods for indoor localization. In Proceedings of the IEEE International Symposium on Intelligent Signal Processing, Budapest, Hungary, 26–28 August 2009; pp. 9–14.
34. Sun, G.; Chen, J.; Guo, W.; Liu, K.R. Signal Processing Techniques in Network Aided Positioning. *IEEE Signal Process. Mag.* **2005**, *22*, 12–23.
35. Vaupel, T.; Seitz, J.; Kiefer, F.; Haimerl, S.; Thielecke, J. Wi-Fi positioning: System considerations and device calibration. In Proceedings of the International Conference on Indoor Positioning and Indoor Navigation (IPIN), Zurich, Switzerland, 15–17 September 2010; pp. 1–7.
36. Honkavirta, V.; Perala, T.; Ali-Loytty, S.; Piché, R. A comparative survey of WLAN location fingerprinting methods. In Proceedings of the 2009 6th Workshop on Positioning, Navigation and Communication, Hannover, Germany, 19 March 2009; pp. 243–251.
37. Sun, W.; Liu, J.; Wu, C.; Yang, Z.; Zhang, X.; Liu, Y. MoLoc: On Distinguishing Fingerprint Twins. In Proceedings of the 2013 IEEE 33rd International Conference on Distributed Computing Systems, Philadelphia, PA, USA, 8–11 July 2013; pp. 226–235.
38. Han, D.; Jung, S.; Lee, M.; Yoon, G. Building a practical Wi-Fi-based indoor navigation system. *IEEE Pervasive Comput.* **2014**, *13*, 72–79.
39. Liu, H.; Darabi, H.; Banerjee, P.; Liu, J. Survey of Wireless Indoor Positioning Techniques and Systems. *IEEE Trans. Syst. Man Cybern. Part C* **2007**, *37*, 1067–1080. [[CrossRef](#)]
40. De Montjoye, Y.A.; Radaelli, L.; Singh, V.K. Unique in the shoppingmall: On the reidentifiability of credit card metadata. *Science* **2015**, *347*, 536–539. [[CrossRef](#)] [[PubMed](#)]
41. Faragher, R.; Harle, R. An analysis of the accuracy of bluetooth low energy for indoor positioning applications. *Phys. Rev. A* **2014**, *84*, 8049–8054.
42. Khalajmehrabadi, A.; Gatsis, N.; Akopian, D. Modern WLAN Fingerprinting Indoor Positioning Methods and Deployment Challenges. *IEEE Commun. Surv. Tutor.* **2017**, *19*, 1974–2002. [[CrossRef](#)]
43. Ji, Y. *Location Determination within Wireless Networks*; VDM Verlag Dr. Müller: Saarbrücken, Germany, 2009.
44. Wan, N.; Lin, G.; Wilson, G.J. Addressing location uncertainties in GNSS-based activity monitoring: A methodological framework. *Trans. Gis* **2017**, *21*, 764. [[CrossRef](#)]
45. Mao, Y.; Zhong, H.; Qi, H.; Ping, P.; Li, X. An adaptive trajectory clustering method based on grid and density in mobile pattern analysis. *Sensors* **2017**, *17*, 2013. [[CrossRef](#)]
46. Qingwu, H.U.; Wang, M.; Qingquan, L.I. Urban Hotspot and Commercial Area Exploration with Check-in Data. *Acta Geod. Cartogr. Sin.* **2014**, *43*, 314–321.
47. Wang, M.; Li, Q.; Hu, Q.; Zhou, M.; Li, Q.Q.; Hu, Q.W. Quality Analysis on Crowd Sourcing Geographic Data with Open Street Map Data. *Geomat. Inf. Sci. Wuhan Univ.* **2013**, *38*, 1490–1494.
48. Wu, L.; Hu, S.; Yin, L.; Wang, Y.; Chen, Z.; Guo, M.; Chen, H.; Xie, Z. Optimizing Cruising Routes for Taxi Drivers Using a Spatio-Temporal Trajectory Model. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 373. [[CrossRef](#)]
49. Zheng, L.; Xia, D.; Zhao, X.; Tan, L.; Li, H.; Chen, L.; Liu, W. Spatial-temporal travel pattern mining using massive taxi trajectory data. *Physica A* **2018**, *501*, 24–41. [[CrossRef](#)]
50. Yang, C.; Shao, H.R. Wi-Fi-based indoor positioning. *Commun. Mag. IEEE* **2015**, *53*, 150–157. [[CrossRef](#)]
51. Zhang, H.; Shi, B.; Song, S.; Zhao, Q.; Yao, X.; Wang, W. Statistical analysis of the stability of bus vehicles based on GPS trajectory data. *Mod. Phys. Lett. B* **2019**, *33*, 1950015. [[CrossRef](#)]
52. Giannotti, F.; Nanni, M.; Pedreschi, D.; Pinelli, F. Trajectory pattern analysis for urban traffic. In Proceedings of the International Workshop on Computational Transportation Science, Washington, DC, USA, 3 November 2009.

53. Yoo, J.; Kim, H.J.; Johansson, K.H. Mapless indoor localization by trajectory learning from a crowd. In Proceedings of the 2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN), Alcala de Henares, Spain, 4–7 October 2016.
54. Yoo, J.; Johansson, K.H.; Kim, H.J. Indoor Localization Without a Prior Map by Trajectory Learning from Crowdsourced Measurements. *IEEE Trans. Instrum. Meas.* **2017**, *66*, 2825–2835. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).