

# An Efficient Wireless Access Point Selection Algorithm for Location Determination Based on RSSI Interval Overlap Degree Determination

Bang Wu, Zixiang Ma, Stefan Poslad

School of Electronic Engineering and Computer Science  
Queen Mary University of London  
London, United Kingdom

Email: {Bang.Wu, zixiang.ma, stefan.poslad}@qmul.ac.uk

Wei Zhang

School of Geodesy and Geomatics  
Wuhan University  
Wuhan, China

Email: 2012202140021@whu.edu.cn

**Abstract**— Optimally choosing wireless Access Points (APs) as urban areas become more densely packed with them becomes increasingly challenging. In WiFi-based Indoor Positioning Systems (IPS), Selecting Wireless Access Point (AP), namely, WiFi routers, is significant as the more APs that are selected, the higher the computation, energy and time cost. This is unsuitable for networking low-resource devices as part of an Internet of Things. In addition, selecting the optimum number of APs not only reduces redundant information but also improves the positioning accuracy. In this paper, we present a novel AP selection method that uses the RSSI Interval Overlap Degree (IOD) to discriminate between known location Reference Points. We validated our algorithm in an office-like indoor space at a Queen Mary computer science lab. The results show that our algorithm has an improved performance, which is 13.6%, 18.2%, and 7.6% better than IG (information gain), MI (mutual information), SD (standard deviation) used as baseline algorithms, respectively.

**Keywords**—WiFi RSSI; IPS; Fingerprinting; Interval Overlap; Location-based Services;

## I. INTRODUCTION

As both indoor and outdoor spaces become more complex, especially in urban areas, the need for indoor positioning and navigation services [1], [2], [3] becomes paramount. There are many other value-add location-based services (LBS) applications, e.g., commercial ads, can be made more personalized if they are location-aware [4]. Due to the increasing complexity of the indoor environment and the lack of a global positioning system, the research and development for low cost, high accuracy, real-time and high reliable LBS is still full of challenges [5].

An indoor positioning system (IPS) is a system to locate objects or people inside a building using Radio Frequency (RF), magnetic fields [7], acoustic signals and dead-reckoning approaches using inertial sensors such as accelerometers and gyroscopes integrated into mobile devices [8]. Although a range of indoor localization technologies can be used [9], [10], the use of Wireless Fidelity (WiFi) as a positioning system has become more popular in recent years, especially indoors, because of the broad deployment of APs. Local area LBS tends to be based upon the use of Received Signal Strength Indicator (RSSI) that indicates the radio signal decay with distance [11]. This requires no specialist equipment unlike other methods based upon determining the timing delay between sending versus receiving signals over local area that requires more specialized clocks and their synchronization. RSSI based WiFi-

based LBS generally includes two approaches, a simple signal decay or signal attenuation or path loss model and a fingerprint-based system. An advantage of the simple path loss model system is that it only requires off-the-shelf smartphones, tablets, or laptops to get the location of the device based on RSSI according to signal attenuation model [12]. Unfortunately the simple signal decay model assumes a free space for RF signal propagation but almost all physical space is not free it contains physical objects such as people and fixtures, which causes additional complex RF signal attenuation and reflection so reducing its accuracy. The fingerprint-based technique offers a more accurate positioning system in comparison, however, it typically works in two phases [13]. First it requires a fingerprint or radio map of RF signals to be collected in an offline phase where the RSSI signals from multiple APs are collected and mapped to locations determined by some other more accurate location determination method, e.g., laser tape, UWB, and then stored as a radio map or fingerprint database. Then during a second, operational or online phase, the location is determined through an estimation of matching the current position with the nearest known location in the pre-stored radio map.

Many RSSI based positioning systems are client-driven architectures in terms of triggering and driving indoor LBS services. This offers the clients' devices a basic privacy filter in that at the very least, users can decide whether to switch client-side location based sensors on or off, albeit some mobile devices can be configured by providers to bypass this.

Location sensing and the computation to determine the position can in general be somewhat power hungry for LBS applications in mobile client devices [14]. Hence, energy-efficient indoor location services is a core issue, else the entire client device can become unusable and non-operational because of a lack of power to drive core wireless services [15].

Due to the increasing availability and high density of heterogeneous wireless APs in indoor environments, more and more networks and APs are provided and can be detected by mobile (wireless) access devices. This however increases the power and computation efficiency to detect valid receivers, and their location, in the face of so many more competitive, possibly invalid transceivers and the need to efficiently select these. The mere fact that a wireless receiver can receive information means that it is in range of a transmitter and gives it a fix in relation to this transmitter. The use of multiple transmitters can give a more accurate fix over a single transmitter through comparing transceiver signal characteristics such as RSSI but this has limits. When

discovering the transmitters or receivers, that are available, the more there are, requires more computation and power to differentiate and to map these to locations.

Using more APs can improve the positioning accuracy due to more signal comparisons to determine the location. Unfortunately, some APs are affected more than others by dynamic effects which cause the RSSI values to fluctuate at the same location at the frequencies used, such as people moving or the proximity to other RF sources that are intermittently active in the same unlicensed RF frequencies as Wi-Fi such as Bluetooth and microwave ovens. These effects affect the location determination accuracy. Appropriate AP selection not only helps to remove the APs with a poorer location determination accuracy but also improves the computation and energy efficiency. It is a critical challenge how to choose the optimum number of APs for location determination.

The remainder of this paper is organized as follows: Section II reviews related work. Section III introduces the RSSI Interval Overlap Degree (IOD) algorithm to select APs and IV presents the experiments to determine the AP selection performance. Section V concludes with a discussion and future work.

## II. RELATED WORK

According to the wireless attenuation model [11], the WiFi signal intensity attenuation model is described by

$$P_{r,dB}(d) = P_{r,dB}(d_0) - 10n \log_{10}\left(\frac{d}{d_0}\right) \quad (1)$$

where  $P_{r,dB}(d)$  is the signal power received at a point which has a distance  $d$  to the transmitter;  $P_{r,dB}(d_0)$  is the signal power received at a reference point with a distance  $d_0$  to the transmitter;  $n$  is the pathloss factor.

In equation (1),  $P_{r,dB}(d_0)$ ,  $d_0$  and  $n$  are known in advance. The relationship between  $d$  and  $P_{r,dB}(d)$  can be determined to be

$$\log_{10}d = \frac{P_{r,dB}(d_0) - P_{r,dB}(d) + 10n \log_{10}(d_0)}{10n} \quad (2)$$

This expresses how  $P_{r,dB}(d)$  changes with  $d$ . In a real world indoor space,  $P_{r,dB}(d)$  may exhibit signal fluctuations caused by a variety of factors, e.g., multipath and non-line-of-sight (NLOS) propagation, time-variant attenuation such as people moving. Therefore, different locations often have different RSSI values from the same AP which is the foundation for fingerprint-based positioning approach.

An essential goal of indoor location estimation systems is to increase the location estimation accuracy while reducing the power consumption. Many AP selection approaches have been proposed to help establish a more efficient and reliable positioning system as follows.

### A. Statistical Analysis-Based Methods for AP selection

Statistical analysis based methods use values such as the mean and standard deviation to act as a criterion for AP selection.

Youssef et al. [16] proposed selecting APs that have a high mean value of RSSI known as MaxMean. This method allocates higher importance to the strongest RSSI. However, the mean value of RSSI doesn't always accurately reflect the dynamic variety of RSSI as RSSI can be easily affected by the environment factors described earlier.

Another method called RSSI standard deviation (SD) in [17] is used as the standard deviation of RSSI series collected at a RP to analyze the signal stability of APs. Assuming  $RSSI_i^j$  denotes the  $j$ -th RSSI value of  $AP_i$  and  $\overline{RSSI}_i$  is the mean value of RSSI from  $AP_i$ .  $SD_i$  denotes the SD of  $AP_i$  which can be calculated as follows:

$$SD_i = \sqrt{\frac{1}{n-1} \sum_{j=1}^n (RSSI_i^j - \overline{RSSI}_i)^2} \quad (3)$$

Then, sort these values of each AP's SD in ascending order and select top- $k$  APs with smallest SD as a lower SD value means the RSSI series from the AP is more stable. SD is a criterion to reflect the amount of variation of a set of data values relative to the mean value. However, each of the values in the set is included for the computation of SD, this makes SD not suitable for a situation where some outliers exist.

### B. Information Theory-Based Methods for AP selection

In information theory and machine learning, information gain (IG) is often used as a criterion for feature selection. For fingerprint-based positioning, AP selection is also regarded as feature selection. Chen et al. [14] proposed an AP selection strategy based on IG. The idea of selecting AP using IG is as follows

If  $N_t$  is the number of APs ( $AP_j, 1 \leq j \leq N_t$ ) that can be detected at each RP ( $R_i, 1 \leq i \leq S$ ) in an indoor space. The IG of  $AP_j$  can be described by

$$IG(AP_j) = H(R) - H(R|AP_j) \quad (4)$$

where  $H(R)$  is the entropy of a RP and can be demonstrated by

$$H(R) = \sum_{i=1}^S P(R_i) \log P(R_i) \quad (5)$$

where  $P(R_i)$  is the prior probability of  $R_i$ . There are  $S$  RPs in total, we set  $P(R_i) = 1/S$  for each RP since the probability of each RP is viewed within a uniform distribution.  $H(R|AP_j)$  is the conditional entropy of a RP given the RSSI of  $AP_j$ . It can be calculated accordingly as follows

$$H(R|AP_j) = -\sum_v \sum_{i=1}^S P(R_i, AP_j = v) \log P(R_i|AP_j = v) \quad (6)$$

where  $v$  denotes one possible RSSI value from  $AP_j$  and  $V$  are all the possible values from  $AP_j$ . Then the conditional probability can be shown as follows

$$P(R_i|AP_j = v, v \in V) = \frac{P(AP_j = v|R_i)P(R_i)}{P(AP_j=v)} \quad (7)$$

The discriminative ability means the ability that an AP can distinguish RPs from each other. For decision tree classifiers, the higher the IG value, the better the classification ability of this feature. Therefore, a higher IG value for the AP means a strong discriminative ability towards RPs. If IG values of all APs are sorted in descending order, the top  $k$  APs with highest IG values will be chosen.

The main merits of IG are that IG makes good use of the diversity of the data samples and takes the discriminative ability of each AP into consideration. Nevertheless, IG is more likely to choose the AP having more variables than others, some of which are unstable and more physically environmentally sensitive, leading to the degradation of positioning accuracy.

Another information theory-based method is mutual information (MI) which is used by Zou et al. [18] for selecting APs. The MI of two random variables is a measure of the mutual dependence between the two variables. The higher of

MI, the more information these variables share which also means redundant information. Redundant information doesn't bring any useful information but increases the amount of computation. Thus, the main target of this MI-based AP selection method is to pick out some APs with the least amount of redundant information.

If  $MI(AP_a, AP_b)$  denotes the MI of  $AP_a$  and  $AP_b$  and it can be described by

$$MI(AP_a, AP_b) = H(AP_a) + H(AP_b) - H(AP_a, AP_b) \quad (8)$$

where  $H(AP_a)$  and  $H(AP_b)$  are the information entropies of  $AP_a$  and  $AP_b$  respectively.  $H(AP_a, AP_b)$  expresses the joint entropy of  $AP_a$  and  $AP_b$ .

$$H(AP_a, AP_b) = \sum_{v_2} \sum_{v_1} [P(RSSI_a = v_1, RSSI_b = v_2 | v_1 \in V_1, v_2 \in V_2) \times \log P(RSSI_a = v_1, RSSI_b = v_2)] \quad (9)$$

The first step of this approach is to identify the AP pairs with the least MI. If  $N_t$  is the number of APs in the space, there exists  $N_t(N_t - 1)$  AP pairs. Assume  $(AP_a, AP_b)$  is the choice, then add another  $AP_c$  to compute the MI of  $AP_c$  and  $(AP_a, AP_b)$ . At this time,  $(AP_a, AP_b, AP_c)$  with the least MI will be chosen. The process can be described by:

$$\arg \min_{AP_z} MI(AP_a, AP_b, \dots, AP_z) \quad (10)$$

where  $MI(AP_a, AP_b, \dots, AP_z)$  is calculated by

$$MI(AP_a, AP_b, \dots, AP_z) = H(AP_a, AP_b, \dots) + H(AP_z) - H(AP_a, AP_b, \dots, AP_z) \quad (11)$$

Following this procedure, a group including  $k$  APs with the smallest  $MI(AP_1, AP_2, \dots, AP_k)$  will be chosen.

The MI approach focus on the information relevance between APs and can efficiently reduce the redundant information. However, one drawback is that MI method doesn't tell the difference between APs. And another one is that it doesn't evaluate each AP's discriminative ability to RPs independently.

### C. Group Discrimination-Based Method for AP selection

Unlike some AP selection strategies measuring the importance of each AP, Lin et al. [19] presented a group-discrimination-based AP selection method, known as Group Discrimination or GD. The basis of this approach is in using Support Vector Machines (SVMs) to assess the group performance of selected APs instead of the individual importance. However, one big problem is the large amount of computation needed. When used for a large-scale AP scenario, with requires a high amount of data training, this will consume a lot of machine memory and require much computation, which is not suitable for resource limited devices.

There are many limitations for the methods mentioned above. Statistical analysis-based methods only reflect a certain statistical index for RSSI samples. Information theory-based approaches retain the diversity of information well, but IG tends to choose APs with more variables which are unstable and more physically environmentally sensitive. MI has weak ability to discriminate between APs and evaluate each AP's discriminative ability to RPs. The group discrimination-based algorithm needs additional machine memory and a long computation time. Therefore, a novel method, called Interval Overlap Degree (IOD), is proposed to offer solution to these limitations.

There are many features about this proposed approach. First, IOD uses numerical interval as a measure to analyze the

characteristics of RSSI sample values from APs. Secondly, IOD preserves the original features of samples instead of extracting one statistical criterion for AP selection; at the same time, it inherits the rule of individual discriminative ability of APs for AP selection. Lastly, IOD achieves a higher accuracy while improving the computational efficiency.

The concept of the overlap degree has been used for localization in [20]. It is used to represent the number of regions that cover the overlap point. The overlap points with maximum overlap degrees are to act the estimated location of the unknown node. In this paper, to the best of our knowledge, we firstly introduce numerical interval overlap degree of RSSI serial values, also called IOD, into WiFi indoor positioning paradigm to achieve the AP selection.

## III. METHODOLOGY

### A. The Principle of IOD for AP selection

In the WiFi fingerprint positioning system, during the offline fingerprint collection stage, RSSI data samples from multiple APs will be collected at each RP as shown in Fig. 1.  $U_j = \{AP_1, AP_2, \dots, AP_n\}$  presents  $n$  detected APs at  $RP_j$ . In real environment, the AP list at each RP may be different, here we just show the same AP set at each RP.

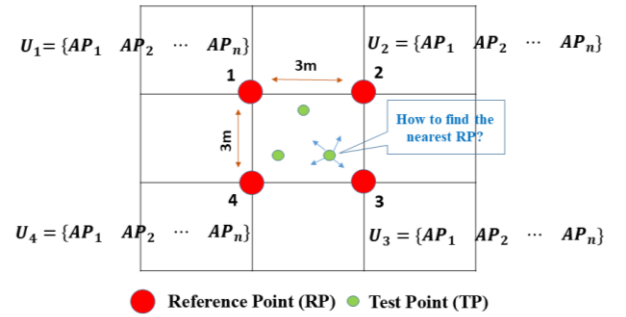


Fig. 1. The explanation for positioning process

The positioning stage is regarded as a process of pattern recognition. Each RP is divided into a class. The features, APs, have a key impact on the matching result. Thus, one aim of AP selection is to choose the most appropriate AP set to achieve a better positioning accuracy. To demonstrate the work principle of IOD for AP selection, we selected  $AP_i$  as an example as shown in Fig. 2.

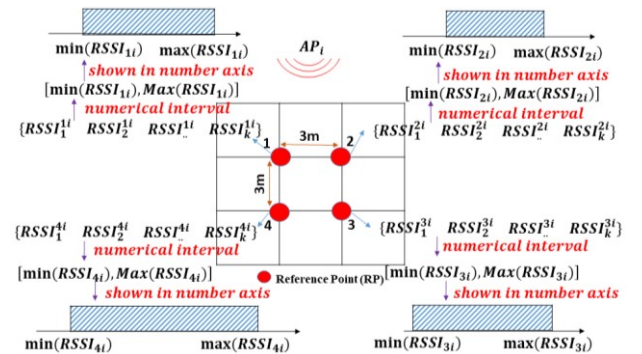


Fig. 2. The extraction of RSSI numerical interval of  $AP_i$

Assume  $RSSI_{1i} = \{RSSI_1^{1i}, RSSI_2^{1i}, \dots, RSSI_k^{1i}\}$  denotes  $k$  possible RSSI values from the  $i$ -th AP at the first RP. The numerical interval of  $RSSI_{1i}$  can be shown as:  $[\min(RSSI_{1i}), \max(RSSI_{1i})]$ . Visualize this numerical interval using number axis (the blue stripes as shown in Fig. 2). The RSSI values of  $AP_i$  at each RP all can be displayed as numerical intervals. The question is how to discriminate RPs from each other using  $AP_i$ . In other words, what rule can be used to distinguish these four types of numerical intervals as shown in Fig. 2.

IOD can be employed as a criterion to discriminate numerical intervals effectively. As shown in Fig. 3, the red part in number axis denotes the overlap length ( $OI$ ), and the gray part is the non-overlap length. Then  $IOD(NI_1, NI_2)$  can be calculated as follows:

$$IOD(NI_1, NI_2) = \frac{1}{2} \left( \frac{OI(NI_1, NI_2)}{\text{length}(NI_1)} + \frac{OI(NI_1, NI_2)}{\text{length}(NI_2)} \right) \quad (12)$$

Where  $NI_1, NI_2$  represent different numerical intervals.  $\text{length}(NI_1), \text{length}(NI_2)$  stand for the length of  $NI_1, NI_2$ , respectively.  $OI(NI_1, NI_2)$  is the overlap length between  $NI_1$  and  $NI_2$ . The value range of IOD is  $[0, 1]$ .

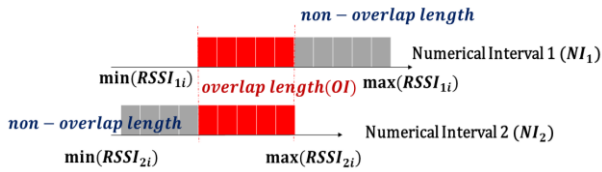


Fig. 3. The sketch of interval overlap

IOD is defined as the mean of proportion that the overlap length between two numerical intervals accounts in each of these two numerical intervals. The lower the IOD, the bigger the difference between two numerical intervals, which means these are easier to differentiate.

There are four overlap relationships between two numerical intervals. Fig. 4 shows how to compute the  $OI(NI_1, NI_2)$  in these four cases. We set the value of  $OI(NI_1, NI_2)$  zero when there is no overlap.

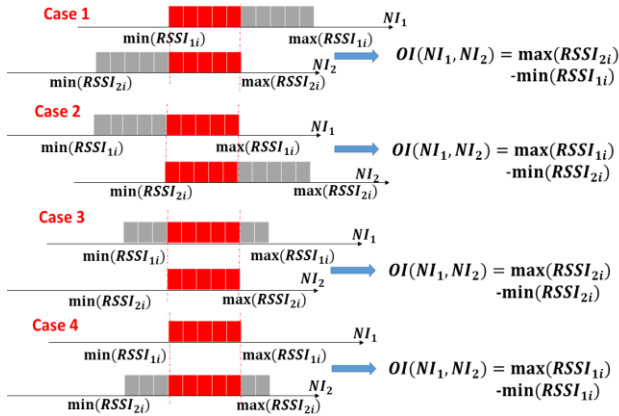


Fig. 4. The four cases of overlap length.

The difference between two numerical intervals can be measured by IOD as described above. Likewise, the RSSI serial values from  $AP_i$  at different RPs can be discriminated using the

IOD method. The following steps are how to select optimal AP set.

### B. Process of AP Selection

To achieve a high positioning accuracy, the RPs should have suitable features to distinguish them, then the Test point (TP) could match the nearest RPs more accurately. This is the reason we need to select APs with the strongest discriminative ability towards RPs.

The RSSI samples from the same AP are different in different RPs. If  $RSSI_{ia} = \{RSSI_1^{ia}, RSSI_2^{ia}, \dots, RSSI_k^{ia}\}$  denotes the RSSI sample values collected at  $i$ -th RP from  $AP_a$  and  $RSSI_{ja} = \{RSSI_1^{ja}, RSSI_2^{ja}, \dots, RSSI_k^{ja}\}$  represents the RSSI samples collected at  $j$ -th RP from  $AP_a$ .  $IOD_{ij}^a$  stands for the IOD between  $RSSI_{ia}$  and  $RSSI_{ja}$ . The higher  $IOD_{ij}^a$  is, the smaller the difference between  $RSSI_{ia}$  and  $RSSI_{ja}$ . This means it is more difficult to discriminate  $i$ -th RP and  $j$ -th RP using  $AP_a$ . So our target is to select APs with the lowest IOD.

If  $N$  is the total number of RPs, then there are  $N(N-1)$  RP combinations in all. Let  $IOD_{f-a}$  represent the final IOD of  $AP_a$ . It can be calculated as follows:

$$IOD_{f-a} = \sum_{i=1}^{N-1} \sum_{j=i+1}^N IOD_{ij}^a \quad (13)$$

After figuring out final IOD of all APs, we rank them in ascending order and select the top- $S$  APs with lowest IOD.

## IV. EVALUATION

### A. Experiment Setup

To validate the performance of the proposed AP selection method, experiments were conducted in an office-like indoor space at a Queen Mary computer science lab. Fig. 5 shows the plan of experiment scene. This room is around  $450 \text{ m}^2$  ( $30\text{m} \times 15\text{m}$ ). The RPs are in black, and the TPs are in green. There were twenty-four RPs and one hundred TPs in total. The horizontal and vertical distance between RPs is 3 meters. The horizontal and vertical distance between most TPs is 1 meter except for some locations without a TP. During the experiment, as people were working and normally moving around during daily activities, we didn't filter the effects of environmental factors e.g., signal reflection, temporal attenuation and noise.

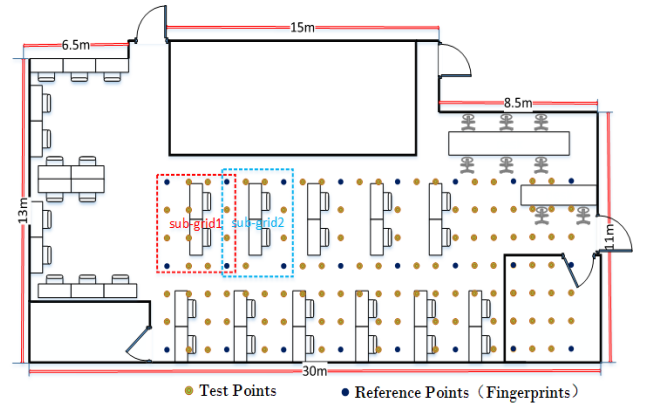


Fig. 5. The plan of experiment scene



We used the same smart phone device throughout as a RSSI collector. The size of RSSI samples at each RP is 120 and the

sampling time span is 2 seconds. Over 90 APs were detected in this experiment.

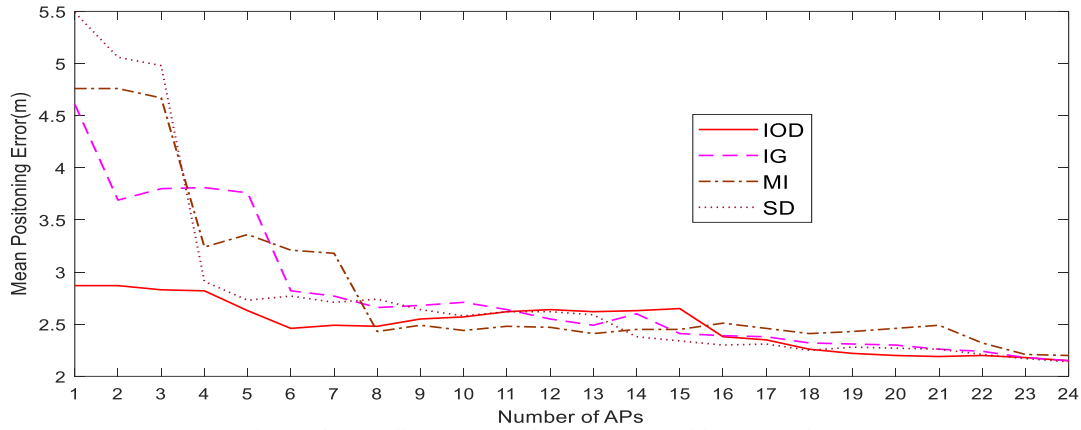


Fig. 6. The overall mean positioning error (MPE) of four approaches.

### B. K-Nearest Points Searching Algorithm

In most cases, the Euclidean Distance (ED) is used for searching the nearest points which will be included in a Weighted K-Nearest Neighbor (WKNN) approach to obtain the location of TP. This method chooses the subset of RPs with shortest ED to the TP. Here, another algorithm, called RSSI ranking - Kendall Tau Correlation Coefficient (RR-KTCC) proposed by Zixiang Ma et al. [2] which was used to address the issue about equipment difference in BLE positioning system, will be employed to search for the nearest points.

$\{AP_1, AP_2, \dots, AP_k\}$  is the received list of APs at one RP or TP. This AP list is sorted in descending order by comparing the mean of RSSI sample values from each AP. The AP lists collected at each RP and TP all comply with this rule. Then, we divide the AP set into subsets that only containing two APs of which positions are ordered, for example,  $\{AP_1, AP_2\}$  indicates the mean of RSSI sample values from  $AP_1$  is bigger than the mean of  $AP_2$ .

As for two AP sets, set  $A$  and set  $A'$ , if the subset (e.g.  $\{AP_1, AP_2\}$ ) from set  $A$  and subset (e.g.  $\{AP_1, AP_2\}$ ) from set  $A'$  are the same, we can say these two subsets are concordant. Or they are discordant.

Therefore, the KTCC displayed as  $\tau$  between two AP sets can be shown as follows:

$$\tau = \frac{\text{num of concordant subsets} - \text{num of discordant subsets}}{K*(K-1)/2} \quad (14)$$

where  $K$  is the size of AP set.  $\tau$  is used to measure the ordinal association between two rankings. And the range value is  $[-1, 1]$ . 1 means a perfect agreement between two rankings; -1 means a perfect disagreement.  $\tau$  can be regarded as a distance measure since the greater the value, the higher agreement between the two rankings, which means closer distance in signal distance space. We chose the top- $K$  RPs with the greatest  $\tau$  relative to TP.

### C. Performance Evaluation Index

In this paper, the WKNN approach is employed to get the location of TP. The principle of this method can be shown as follows:

$$(x, y) = \sum_{i=1}^K \frac{1/D_i}{\sum_{i=1}^K 1/D_i} (x_i, y_i) \quad (15)$$

Where  $(x, y)$  infers the coordinate of TP, and  $(x_i, y_i)$  denotes coordinate of  $i$ -th RP.  $K$  is the total number of RPs which are the nearest RPs to TP. In this paper,  $K$  is 4 as advised by [18-21] when KNN is used. In general, if  $K$  equals 3 or 4 will yield good results.  $D_i$  represents the Euclidean distance between TP and  $i$ -th RP which can be described by

$$D_i = \sqrt{\sum_{j=1}^n (\bar{f}_j^i - \bar{f}_j)^2} \quad (16)$$

Here  $\bar{f}_j^i$  is the mean value of RSSI samples from  $j$ -th AP collected at  $i$ -th RP.  $\bar{f}_j$  is the mean value of RSSI samples from  $j$ -th AP at the TP. And  $n$  is the number of APs selected. Therefore, the positioning error can be denoted as

$$PE = \sqrt{(x - x_0)(y - y_0)} \quad (17)$$

where  $(x_0, y_0)$  is the known coordinate of the TP.

Apart from the positioning error (PE), mean positioning error (MPE), max error (MaxE), the standard deviation of error (STDE), root mean square of error (RMSE) and cumulative distribution function (CDF) are also used to assess the performance of the proposed method [22].

### D. Indoor Positioning Performance

Three approaches are employed to act as baseline algorithms to make a comparison with our proposed method, IOD. Fig. 6 shows the overall mean positioning error (MPE) of four approaches. IOD in the red color shows a best performance compared to the other methods when the number of APs is 7 or less. Moreover, a MPE trough of 2.46m for the IOD appears when using 6 APs. It is worth noting that a 2.87m level of accuracy is achieved when just one or two APs are used. The location accuracy remains below 3m no matter how many APs are used. Starting at 8 APs, the MPE lines of all methods tend to be stable, and the gap between them is reduced.

Table I lists the details of MPE of four methods when the number of APs are 6,7,8,9 and 10 respectively. When the

number of APs equals 6, the MPE of IOD is 2.46m. Only when 8 or 10 APs are used, the MPE of MI can surpass this accuracy level, with 2.43m and 2.44m respectively. However, the accuracy only increases by 0.03m using 8 APs and 0.02m using 10 APs. There seems to be little gain in accuracy at the cost of adding 2 or 4 APs' with their increase computation and energy cost, especially for devices with limited resources. Thus, using a six-AP subset can help improve computation efficiency while preserving an acceptable location accuracy.

TABLE I. THE MPE OF FOUR METHODS WHEN THE NUMBER OF AP IS 6,7,8,9 AND 10 RESPECTIVELY

Number of APs	MPE			
	IOD	IG	MI	SD
6	2.46	2.82	3.21	2.77
7	2.49	2.77	3.18	2.71
8	2.48	2.66	2.43	2.74
9	2.55	2.68	2.49	2.64
10	2.57	2.71	2.44	2.58

Fig. 7 demonstrates the cumulative density function (CDF) of four approaches when 6 APs are being used. Moreover, the detailed numerical results of CDF are listed in Table II.

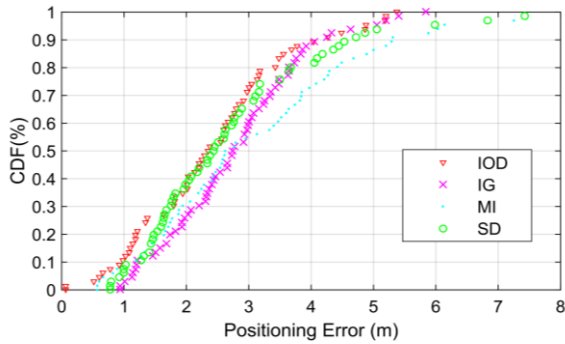


Fig. 7. CDF of four approaches.

TABLE II. THE DETAILS OF CDF

Accuracy	Methods			
	IOD	IG	MI	SD
within 1m	10.61%	1.52%	7.58%	7.58%
within 2m	34.85%	25.76%	30.30%	36.36%
within 3m	72.73%	59.09%	54.55%	65.15%

Red triangle line represents the cumulative probability of IOD method. Within a 1-m accuracy, the IOD outperforms the other three methods. The cumulative probability of IOD is over 9% higher than IG's and over 3% higher than the cumulative probability of MI and SD. Within a 2m accuracy, IOD performs a little weaker than SD but still better than IG and MI. Besides, the cumulative probability of accuracy within 3m, IOD is 13.64%, 18.18%, and 7.58% better than IG, MI, and SD.

TABLE III. THE COMPREHENSIVE EVALUATION OF PERFORMANCE OF FOUR APPROACHES

Indexes	Methods			
	IOD	IG	MI	SD
MPE	2.46	2.82	3.21	2.77
MaxE	5.37	5.84	11.17	8.32
STDE	1.23	1.11	1.95	1.57
RMSE	2.74	3.03	3.75	3.18

Four error analysis indexes are reported, including MPE, MaxE, STDE, and RMSE, as shown in Table III. For MPE, MaxE, and RMSE, IOD shows a better performance than any other methods. The STDE of IOD is 0.12 higher than STDE of IG but over 0.7 and 0.3 lower than the STDE of MI and SD. From the comparison of these evaluation indexes, IOD is superior to the other baseline algorithms.

### E. Boundary Analysis of Accuracy

Positioning accuracy can be affected by many factors in a fingerprint-based WiFi indoor positioning system, e.g., AP selection strategies, AP topology construction, nearest RPs searching algorithm. This paper mainly focuses on the AP selection method. This experiment was conducted in an office space using all APs which can be detected by a mobile phone. So it is hard to consider the topology of APs, e.g., height, distribution, etc., which will be tackled in future work. The nearest RPs searching algorithm is implemented by using the Euclidean distance (ED). The major factor affecting accuracy in this experiment setup is matching a TP to the nearest RPs.

To test the limits of location accuracy, we divided our whole scene into sub-grids as shown in Fig. 5. Within each sub-grid, the four RPs in black color is used for determining the TPs' location in yellow color. Then each TP use four physically close RPs to skip the process of nearest RPs searching.

Fig. 8 shows the overall MPE using a different number of APs of four methods. Red color line with squares expresses the MPE of IOD. When the number of APs is more than 3, IOD outperforms the other methods. After the number of APs surpass 9, the accuracy of IOD is stable and the least value is 1.37m which is the upper bound of accuracy for our experiment.

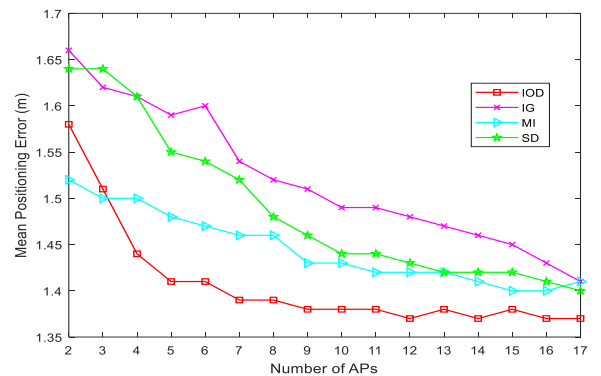


Fig. 8. The MPE of sub-grids of four approaches.

Besides, as shown in the Fig. 8., IOD shows the best performance compared to other methods and it means, to some extent, IOD has a better scalability and signal stability (less fluctuations).

#### F. Performance Comparison of ED and RR-KTCC for Nearest Points Searching

Two nearest points searching algorithms were employed to search for the nearest RPs for positioning in the large-scale scene as shown in Fig. 9. The red color line is the MPE using ED and green one belongs to RR-KTCC. The MPE of RR-KTCC has a lower speed than IOD to be stable as the number of APs increases. When the number of AP is less than 14. ED outperforms RR-KTCC. When 14 or 15 APs being used, RR-KTCC has better performance than ED. While the number of APs is more than 15, ED goes back to the leading position, and the gap between them tends to be stable.

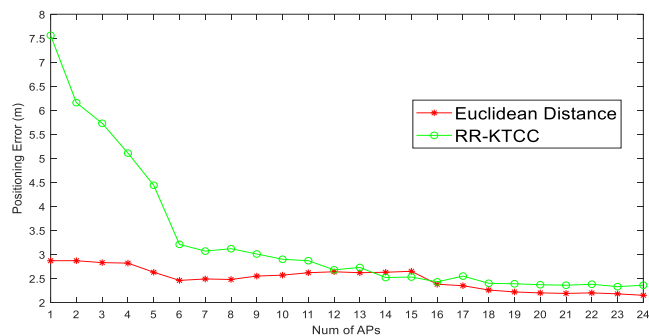


Fig.9. The MPE of ED and RR-KTCC for Nearest Points Searching.

One reason for this phenomenon is that KTCC is used to get the discrepancy between two sets. One evident characteristic of KTCC approach is that the size of AP set requires much big enough. Alternatively, when used for indicating the distance, the distance error will be much higher. What needs to be said is that RR-KTCC is designed for addressing the bias caused by heterogeneous devices. Therefore, using different devices to testify the validation of this algorithm is our next work.

#### V. CONCLUSIONS AND FURTHER WORK

This paper presents an RSSI Interval Overlap Degree (IOD) algorithm for AP selection used as part of a WiFi fingerprint location determination. A comparison between IOD and ED and RR-KTCC was made to determine the position. Our experiments demonstrate that: the proposed AP selection strategy, IOD, outperforms some other AP selection approaches like IG, MI, and SD when less APs are used. IOD demonstrates not only a clearly better accuracy but a higher RF signal stability, with less fluctuations. At the same time, this experiment results show that the nearest points searching approach based on ED is better than RR-KTCC when the number of AP is small and when the same mobile device is used both the offline phrase and online phrase to determine the location.

Our focus in further research includes calibration of systematic deviation of heterogeneous devices, topology condition and an extension for online AP selection using the

proposed method. Another topic is to determine the energy cost of different location determination algorithms and the effect in the ways they select APs for local area networks similar to work that has been done for wide area networks [23].

#### REFERENCES

- [1] J. H. Schiller and A. Voisard, Location-based services. Elsevier, 2004.
- [2] Z. Ma, S. Poslad et al. "A BLE RSSI Ranking based Indoor Positioning System for Generic Smartphones". 2017 16<sup>th</sup> Wireless Telecommunications Symposium (WTS'17), pp. 1–8, 2017.
- [3] D. Meng and S. Poslad. A Reflective Context-Aware System for Spatial Routing Ap-plications. Proc. 6th Int workshop on Middleware for pervasive and ad-hoc computing (MPAC'08), pp. 54-59, 2008.
- [4] B. Rao, L. Minakakis. Evolution of mobile location-based services. Communications of the ACM, vol. 46, no. 12, pp. 61-65, 2003.
- [5] C. Guo, W. Guo, G. Cao, and H. Dong, A lane-level LBS system for vehicle network with high-precision BDS/GPS positioning. Computational Intelligence and Neuroscience, vol. 2015, Article ID 531321, 13 pages, 2015.
- [6] A. Balasubramanian and A. Venkataramani. Energy consumption in mobile phones: a measurement study and implications for network applications. Proc. 9th ACM SIGCOMM conference on Internet measurement(IMC'09), pp 280-293, 2009.
- [7] Z. Ma, S. Poslad, S. Hu, and X. Zhang, "A fast path matching algorithm for indoor positioning systems using magnetic field measurements," in 2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), pp. 1–5, 2017.
- [8] T. Oshin and S. Poslad. ERSP: An Energy-efficient Real-time Smartphone Pedometer. 2013 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 2067 – 2072, 2013.
- [9] K. Yu, I. Oppermann. "Performance of UWB position estimation based on time-of-arrival measurements." in Proc. Int. Workshop Ultra Wideband Syst. (IWUWBS), Oulu, Finland, pp. 400–404, May 2004.
- [10] P. Vorst, J. Sommer, C. Hoene et al., "Indoor positioning via three different RF technologies," in Proceedings of the 4th European Workshop on RFID Systems and Technologies (RFID SysTech '08), pp. 1–10, June 2008.
- [11] W. Xue, W. Qiu, X. Hua, et al. Improved WiFi RSSI Measurement for Indoor Localization. IEEE Sensors Journal, vol. 17, no.7, pp. 2224-2230, 2017.
- [12] J. W. S. Liu, L. J. Chen, J. Su, C. C. Li and E. T.H. Chu. Building/Environment Data/Information Enabled Location Specificity and Indoor Positioning. IEEE Internet of Things Journal, vol. 4, no. 6, pp. 2116-2128, 2017.
- [13] Z. Liang, Z., I. Barakos, and S. Poslad, "Indoor Location and Orientation Determination for Wireless Personal Area Networks," 2nd Int. Workshop on Mobile Entity Localization and Tracking in GPS-less Environments, MELT 2009, pp. 91-105, 2009.
- [14] Y. Chen, Q. Yang, J. Yin, and X. Chai. Power-efficient accesspoint selection for indoor location estimation. IEEE Transactions on Knowledge and Data Engineering (TKDE) , vol. 18, no. 7, pp. 877-888, 2006.
- [15] S. Poslad, A. Ma, Z. Wang, H. Mei, Using a smart city iot to incentivise and target shifts in mobility behaviour—Is it a piece of pie?. Sensors vol. 15, pp. 13069–13096, 2015.
- [16] M. Youssef, A. Agrawala, and A. Udaya Shankar, WLAN location determination via clustering and probability distributions. IEEE Int. Conf. Pervasive Comput. Commun. pp. 143–150, Mar., 2003.
- [17] F. M. Jhuang, C. F. Hung, C. C. Tuan, et al. "An AP selection with RSS standard deviation for indoor positioning in WiFi." Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS), 2015 9th International Conference on. IEEE, pp. 403-407, 2015.
- [18] H. Zou, Y. Luo, X. Lu, H. Jiang, L. Xie. A Mutual Information Based Online Access Point Selection Strategy for WiFi Indoor Localization.

In Proceedings of the Automation Science and Engineering (CASE), Gothenburg, Sweden, 24–28 August 2015, pp. 180–185, 2015.

- [19] J. Yin, Q. Yang, and L. Ni, et al. A group-discrimination-based access point selection for WLAN fingerprinting localization. *IEEE Transactions on Vehicular Technology*, vol. 63, no.8, pp. 3967-3976, 2014.
- [20] Y. S Chen, T. Lo and W. C. Ma. Efficient localization scheme based on coverage overlapping in wireless sensor networks. *5th Int. Conf. on Communications and Networking in China (CHINACOM)*, pp. 1–5), 2010.
- [21] B. Li, J. Salter, A. Dempster, and C. Rizos, “Indoor positioning techniques based on wireless LAN,” in *1st IEEE International Conference on Wireless Broadband and Ultra Wideband Communications*, pp. 13–16, 2006.
- [22] W. Zhang, X. Hua, K. Yu, W. Qiu, X. Chang, B. Wu, X. Chen, Radius based domain clustering for WiFi indoor positioning. *Sensor Review*, vol. 37, no. 1, pp. 54-60, 2017.
- [23] T. Oshin, A. Ma and S. Poslad (2012) A Method to Evaluate the Energy-Efficiency of Wide-Area Location Determination Techniques Used by Smartphones. *15th IEEE International Conference on Computational Science and Engineering (CSE)*, pp. 326-333, 2012.