Indoor Positioning using Wi-Fi Fingerprint with Signal Clustering

ChoRong Park and Seung Hyong Rhee
Dept. of Electronic Convergence Engineering
Kwangwoon University
Seoul, Korea
{chorong56, rhee}@kw.ac.kr

***Abstract*— Indoor positioning systems have been actively studied so far, and recently, many researchers are adopting Wi-Fi signals for their systems. It is mainly because Wi-Fi networks are prevalent in the indoor environments these days; also it is more efficient than other methods as one can estimate his location by simply comparing current RSS (Received Signal Strength) with the *fingerprint* of Wi-Fi signals measured in advance at the area. One of the simple and popular approaches for matching an RSS with the Wi-Fi fingerprint, is known as K-Nearest Neighbors (KNN). While KNN uses all data stored in the fingerprint for the comparison, it shows relatively low accuracy due to the unstable nature of Wi-Fi signals: Especially in a spacious or congested place, its performance may severely degrade. In this paper, we adopt K-means clustering for an efficient and accurate classification in KNN. In our approach, in order to mitigate the effect of unstable Wi-Fi signals, the clustering is performed using every individual RSS values rather than representative mean values, which results in the elimination of extreme RSS values in the classification. In addition, locating the position by KNN utilizes the probability distribution of RSS values in each cluster. Under a wide range of K in KNN and the number of clusters, our experiments show notably enhanced performances than those of the classical KNN.**

Keywords—indoor positioning, Wi-Fi fingerprint, Received Signal Strength (RSS), K-nearest neighbors (KNN), clustering, K-means

I. INTRODUCTION

Indoor Positioning System (IPS) locates the position of moving objects in indoor spaces. While there are various positioning methods such as ultrasound, infrared and RFID tags [1] [2], many researchers are adopting Wi-Fi signals for their systems. It is mainly because Wi-Fi networks are prevalent in the indoor environments these days; also it is more efficient than other methods as one can estimate his location by simply comparing current signal strength with the fingerprint of Wi-Fi signals measured in advance at the area.

Indoor positioning with Wi-Fi utilizes RSS (Received Signal Strength) of AP (Access Point), which is measured in dBm. Then the user's location can be estimated based on the magnitude of the RSS. However, as the wireless signal is very unstable due to multipath effects, it is very hard to determine the exact location using the RSS; for example, if there are obstacles or people in the signal path, the RSS may suddenly be measured low [1]. Similar phenomena include signal reflection, refraction and multipath propagation, all of which are common in indoor spaces [1] [3]. This is one of the reasons why the accuracy of Wi-Fi based IPS is relatively low [3]. Accordingly, many researchers have investigated algorithms in order to improve the accuracy of the IPS. One of the basic methods is K-Nearest Neighbors (KNN) [2]. In addition, trilateration method using the relationship between RSS and distance, and a probabilistic interpretation of Bayes' theorem using Maximum Likelihood Estimation are also studied [1][4].

In this paper, we utilize KNN and propose a method to improve the accuracy of the IPS. While KNN uses all data stored in the fingerprint for the comparison, it shows relatively low accuracy due to the unstable nature of Wi-Fi signals [3]. Especially, in a spacious or congested place, its performance may severely degrade. Our scheme adopts K-means clustering for an efficient and accurate classification in KNN. In our approach, in order to mitigate the effect of unstable signal, the clustering is performed using every individual RSS values measured 100 times rather than representative mean values. In addition, locating the position by KNN utilizes the probability distribution of RSS values in each cluster. Under a wide range of K in KNN and the number of clusters, our experiments show notably enhanced performances than those of the classical KNN. Remaining of this paper is as follows. Section Ⅱ briefly summarizes basic approaches and Section Ⅲ describes our proposed method. Experiment results are explained in Section Ⅳ and some concluding remarks are in Section Ⅴ.

This work was supported by the National Research Foundation of Korea (NRF) grant (No. 2017R1D1A1B03028184).

II. BASIC APPROACHES

A. Location fingerprinting

Indoor positioning based on Wi-Fi fingerprint works two phases: calibration phase and positioning phase [1]. In the calibration phase, a fingerprint is generated by measuring RSS from APs at Reference Points (RPs), which are pre-defined location. In the positioning phase, the current RSS measured at user's location is compared with the fingerprint using various algorithms. Finally, the RP which has the most similar value to the current RSS is selected as the user's location.

B. K-Nearest Neighbors (KNN)

Nearest Neighbor is one of the simple ways to compare with the fingerprint and aims to find the RP that has the closest value to the current RSS through the distance formula:

$$D_i = (\sum_{k=1}^{n} |RSS_{ik} - RSS_k|^p)^{1/p}.$$

RSS_{ik} denotes the fingerprint RSS from AP_k measured at RP_i and RSS_k represents the current RSS from AP_k measured at user's location. For each RP, calculates the distance according to the p value. If p is 1, Manhattan distance, and if p is 2, Euclidean distance [3], which we adopt in this paper. While NN finds only one smallest value, KNN finds K smallest values and determines user's location as a result of majority.

C. Maximum Likelihood Estimation (MLE)

KNN determines user's location by comparing current RSS to the fingerprint without considering distribution of the signal. However, as shown in figure 1, RSS shows a rough normal distribution form [1]. Thus simply searching for a user's location with one RSS comparison causes many errors.

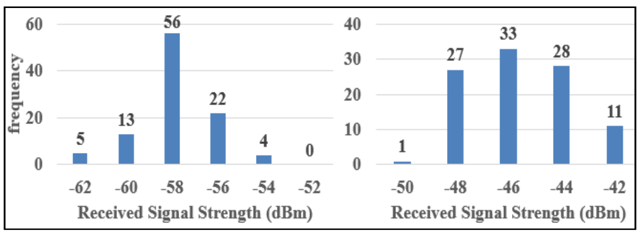


Fig. 1. The distribution of Received Signal Strength from two different APs measured at the same position, which is the 100 results measured at every 30 seconds [1].

The signal with a normal distribution form can represent the probability through Maximum Likelihood Estimation. The probability of obtaining a RSS value x from a normal distribution with mean μ and standard deviation σ is as follows. Since the signal from each AP is an independent value, the probability is equal to multiplying probability of all AP [1].

$$p(x \mid \mu, \sigma) = \frac{1}{\sigma\sqrt{2\Pi}} e^{-(x-\mu)^2/2\sigma^2} \tag{1}$$

D. Cluster analysis

Cluster analysis or clustering is the task of classifying the data with similar characteristics into a same group called a cluster. In indoor positioning, the data to be classified is fingerprint, the classification criterion is Euclidean distance, and the purpose of classification is to increase the efficiency of indoor positioning. Especially for an efficient and accurate classification in KNN, we adopt K-means for the clustering method, which will be described in the next section.

III. Proposed Method

A. K-means algorithm

K-means algorithm is popular for cluster analysis and aims to partition data into k clusters in which data belongs to the cluster with the nearest distance [2][3][4]. The process used in indoor positioning is as follows [2].

1) Determine the number of clusters (K) and set initial centroid of K clusters.

2) Calculate distance between each RP and centroid.

3) Assign each RP to the closest cluster.

4) Calculate the next centroid which is average of the RPs belonging to each cluster.

5) Repeat step 2, 3 and 4 until there is no change in the centroid of K clusters.

In step 1, the initial centroid is RSS which is determined randomly. And in step 2, RP represents the fingerprint RSS and distance is Euclidean distance. After completing the five steps, RSSs with similar characteristics are stored in each cluster. Then, during the positioning phase, a cluster that the current RSS belongs to is determined.

Our approach is different from the previous ones in that we store 100 RSS results at each RP when generating fingerprint and use every individual RSS values for clustering. Since the Wi-Fi signal shows a normal distribution form, if all of the 100 measurements are used, more diverse cases can be considered and thus, the errors in the positioning can be reduced.

B. Maximum Likelihood Estimation (MLE) using histogram

MLE using histogram is a cluster selection algorithm that determines the cluster which the current RSS belongs to. The process is as follows.

1) Perform K-means clustering.

2) Create a histogram of RSS for each AP of each cluster.

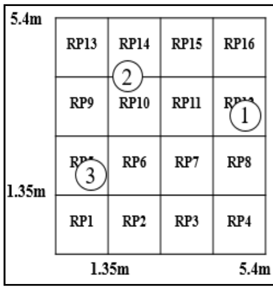
3) In each cluster, calculate the frequency corresponding to the current RSS.

4) Find the cluster which shows highest frequency.

In step 3, since the signal from each AP is an independent value, the probability is obtained by multiplying the frequency of each AP. Step 4 compares the probability of each cluster. It is based on the concept of MLE, but the difference is that the probability of obtaining the current RSS is expressed by the frequency rather than using the mean and standard deviation as in the formula (1). Since the frequency is a measure of how often a particular RSS is measured at a particular location, it can better reflect the characteristics of signal. In our experiment, we compare this method with NN, which selects the cluster with the closest distance.

IV. Experiment

We conducted experiments in a laboratory of Kwangwoon University and used Raspberry Pi3 to build APs and mobile users. As shown in the left side of figure 2, positioning map was divided into 16 cells and three APs were installed. The center of each cell was set as RP and the fingerprint stored 100 results of RSS measured at every 30 seconds. The right side of figure 2 shows a portion of the fingerprint.



RP1	AP1	AP2	AP3
1	-62	-56	-45
2	-64	-59	-46
3	-62	-56	-47
...	...	...	...
99	-64	-54	-46
100	-62	-55	-45

Fig. 2. Positioning map and part of the fingerprint stored at RP1.

Experiment procedure is as follows. The user moved in arbitrary direction and current RSS was measured 5 times at every 30 seconds. All 5 RSSs were applied to the cluster selection algorithm and the cluster was determined as a result of majority.

The first experiment was conducted to compare the accuracy of classical KNN and proposed KNN with two cluster selection algorithm. Method1 is proposed KNN with NN, which selects the cluster with closest distance and Method2 is proposed KNN with MLE using histogram which selects the cluster with highest frequency and Basic is classical KNN. In this experiment, 2-means clustering was used and K value was changed as shown in figure 3.

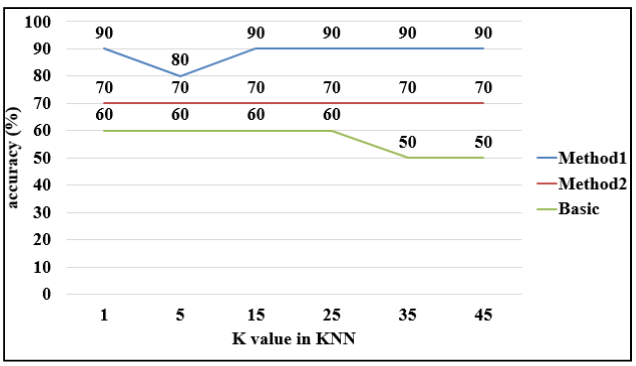


Fig. 3. Result of the first experiment. Method1 is proposed KNN with NN, Method2 is proposed KNN with MLE using histogram and Basic is classical KNN.

In figure 3, while Basic shows less than 60%, Method1 shows more than 80% and Method2 shows 70%. Proposed KNN with both two methods shows better accuracy than classical KNN in all cases. Especially, when K value is more than 35, the difference is greatest as Basic shows 50% and Method1 shows 90%. The accuracy of the proposed KNN shows a difference according to the method. Two cases in the user's movement path were classified into different clusters, and the result shows that Method1 is more accurate in all cases.

The second experiment was conducted to compare the accuracy based on the number of clusters. In this case, the K value of KNN was fixed at 25 and both cluster selection algorithms were used.

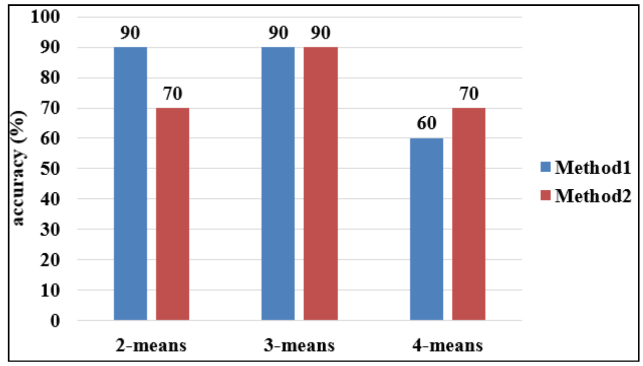


Fig. 4. Result of the second experiment. K value in KNN was fixed at 25.

In the 3-means, both methods show 90% accuracy, but in 2, 4-means, the result depends on the method. In the 2-means, two cases in the user's movement path were classified into different clusters, and the result shows that Method1 is more accurate. In the 4-means, one case was classified into different clusters, and the result of Method2 is better. It can be seen that the proposed cluster selection algorithm, which is MLE using histogram, can improve accuracy. In addition, all results regardless of the number of clusters and the cluster selection algorithms, shows better accuracy than classical KNN.

V. Conclusion

In this paper, we adopt K-means clustering for an efficient and accurate classification in KNN, and in order to mitigate the effect of unstable signal, the clustering is performed using every individual RSS values measured 100 times. In addition, locating the position by KNN utilizes the probability distribution of RSS values in each cluster. In all experiments, our proposed KNN shows better accuracy than the classical KNN, regardless of the number of clusters and the cluster selection algorithms. Also, we showed that the MLE using histogram can improve the accuracy.

We have shown that a clustering algorithm that efficiently utilizes all information of RSS is a suitable method for Wi-Fi based Indoor Positioning Systems. The problem of how to perform more accurate and correct clustering will be investigated further.

References

[1] Noah Pritt, “Indoor Location with Wi-Fi Fingerprinting,” IEEE Applied Imagery Pattern Recognition Workshop, Oct 2013.

[2] Bulut Altintas and Tacha Serif, “Improving RSS-Based Indoor Positioning Algorithm via K-Means Clustering,” 17th European Wireless, April 2011.

[3] Yazhou Zhong, Fei Wu, and Juan Zhang, and Bo Dong, “WiFi indoor localization based on K-means,” Int. Conf. on Audio, Language and Image Processing, pp. 663-667, July 2016.

[4] A. Razavi, M. Valkama, and E. –S. Lohan, “K-Means Fingerprint Clustering for Low-Complexity Floor Estimation in Indoor Mobile Localization,” IEEE Globecom Workshops, Dec 2015.