# Research on Fingerprinting Indoor Positioning Algorithm via Whitening RSS and Fuzzy C-Means Clustering

Jianli Zhao, Bin Zhang, Xin Wang and Shuo Sun

*Shandong University of Science and Technology, Qingdao, China*
*jlzhao@sdust.edu.cn, zhangbin9203@gmail.com, sdustwx@gmail.com,*
*sdustsshuo@163.com*

## *Abstract*

*Due to pervasive penetration of the Wireless Local Area Networks (WLAN) and portable smartphones, the indoor positioning technology based on WLAN has developed rapidly. Especially the Wi-Fi fingerprint-based positioning has attracted continuous attention because of its low cost, high applicability and accuracy. The current fingerprint-based positioning algorithm still has a tremendous challenge both in accuracy and efficiency. To improve and optimize the issues, a novel indoor positioning algorithm based on Whitening RSS and Fuzzy C-Means (FCM) clustering approach is proposed in this paper. Firstly, in offline phase, the preprocessing of gathered RSS data samples is conducted by whitening methods, removing the correlation of the signals, to build the fingerprint database. Then the computation complexity is reduced by dividing the whole database into various sub-databases through FCM clustering algorithm. Meanwhile, the membership degrees obtained from FCM are added to the position calculation as a weight parameter to improve accuracy. Simulation results indicate that the proposed algorithm extraordinary achieves high performance in reducing the computation complexity and enhancing the positioning accuracy.*

***Keywords****: Indoor Positioning, FCM, Whitening method, accuracy, computation complexity*

## 1. Introduction

As the proliferation of wireless LANs infrastructures established in urban buildings, the indoor positioning technology based on WLAN has developed rapidly [1]. Compared with other traditional positioning methods like time of arrival (TOA), time difference of arrival (TDOA) and angle of arrival (AOA) [2] as well as radio propagation models such as log-distance path model and weighted path loss model [3], fingerprint-based indoor positioning technology has been gradually attracting much attention because it takes full advantage of existing equipment to achieve high performance under the circumstances of complex indoor environment demanding neither line-of-sight nor angle measurement [4].

The computation complexity would extremely increase, however, as the growth of the scale of the target area such as supermarket, airport, gallery *etc.* Furthermore, lots of RSS (received signal strength) values containing redundancy, especially, public APs (access points) are deployed densely in which these RSS values are influenced by original information from different Aps [5]. Therefore, reducing the computation complexity without lowering the positioning precision attracts more attention.

The paper is organized as follows. In Section 2 we discuss the related work. Section 3 presents the proposed WR-FCM model in detail. In Section 4, we discuss our experimental evaluation. Finally, we conclude our work highlight a few of future work in Section 5.

## 2. Related Work

Lots of scholars have conducted the thorough study on the question regarding how to decrease computational complexity in the search matching process while assuring indoor positioning accuracy, and have owned some good results. Existing literature shows that using the clustering algorithm can effectively narrow the range of space provided for search matching in large scale areas, greatly cut down the required amount of calculation, which categorizes users into various sub-regions beforehand, utilizing the corresponding sub fingerprint database for matching, to improve the positioning accuracy. In addition, a few de-noising pretreatment technologies in raw RSS data processing, which can effectively remove the correlation between signals from different APs, are used to reduce its influence on the precision of positioning results.

Youssef *et al.* [6] proposed a novel probabilistic indoor positioning method, the Joint Clustering technique, which divides the reference points into various clusters on the basis of that each cluster share the same set of received APs. Although it can be applied to WLAN-based indoor positioning systems to reduce their computational cost and enhance their accuracy, this method is suitable for the situation to a small target area containing a little APs. Another positioning system using clustering technique was designed by Ma *et al.* in [7].In this work, one algorithm named Cluster Filtered KNN, CFK, was proposed combining the average-linkage agglomerative Hierarchical Clustering with KNN to improve the KNN algorithm which partition the neighbors into different candidate points sets. This method is limited by the high computation complexity and susceptible to the similarity between signals in spite of the results presenting it outperforms KNN. In [8], Altintas *et al.* improved the KNN algorithm through intensifying the neighbors' selection by applying the K-Means clustering method in fingerprinting indoor positioning technology. Similarly, the K-Means clustering is used for floor estimation in indoor mobile localization by Razavi *et al.* in [9] and it significantly improves the complexity and speed of floor detection while achieving a floor estimation precision. Sun et al [10] proposed a hybrid algorithm, KNN-FCM, which used the method of fuzzy C-means algorithm to divide reference points to multiple clusters, respectively, according to the received signal strength from APs and the physical location coordinates of the reference points, then calculated the user's location via KNN. Besides, in the work of [11], Zhou *et al.* introduced a fingerprint algorithm based on FCM. The simulation results of [10] and [11] demonstrate utilizing FCM availably reduces the computing time and increases the accuracy. In order to improve positioning precision, Fang *et al.* [12] [13] [14] employed principal component analysis (PCA) technique to remove the correlation between the RSS collected from various APs to extract a few new variables, which are independent and able to represent the majority of the raw data, for computing the users' location in the meantime it decreases the computational complexity. With the purpose of not changing the characteristic nature as well as removing the redundancy of raw signals, Chen *et al.* [15] made use of ZCA whitening technology in the process of dealing with relevant issues between the RSS, obtaining a favorable effect.

ZCA whitening technology and fuzzy C-means algorithm are effective in fingerprint-based position system but no previous research work has combined the two methods to apply in indoor positioning. Therefore, in this paper, one indoor positioning algorithm, named WR-FCM, based on whitening RSS and fuzzy C-Means clustering is proposed.

## 3. Description of the Algorithm Model

### 3.1. Fingerprinting-Based Positioning Technology

Fingerprint-based positioning technology is usually divided into two phases: an off-line phase as well as an on-line phase [16]. During the off-line phase, a site survey concentrated on collecting the detected RSS samples from different APs at many

reference points that known locations previously. All samples forming the fingerprints and being stored at a database, hence, the fingerprint database is built up for online query. In the on-line phase, a user sends a vector containing current real-time RSS data at his or her position that can be seen as a fingerprint to the server and the location of the user is estimated by matching the measured fingerprint with the fingerprint database. The K-nearest neighbors (KNN) is one of popular fingerprint-based algorithms [17] and generally leveraged for indoor positioning in WLAN which employs K most similar reference points while calculating the estimated result.

## 3.2. Whitening Technique

Whitening technique, a form of preprocessing of the data, whose goal is to remove redundancy between the data and retain the essence attribute of raw input data, more formally, our intentions are that the training input of learning algorithm through whitening has the following properties: all dimensions are less correlated with each other and all variances of each dimension are equal.

The commonly-used whitening technique is ZCA whitening method. ZCA Whitening, known as zero-phase filter, leverages a whitening matrix $\mathbf{P}$ to transform the sample matrix $\mathbf{X}$ as $\hat{\mathbf{X}} = \mathbf{PX}$ so that the features of transformed matrix $\hat{\mathbf{X}}$ are uncorrelated each other and all variances of each dimension are exactly 1, satisfying the two conditions of whitening[18][19].

If the data matrix $\mathbf{X}$ meets the condition that $E\left[\mathbf{XX}^T\right] = \mathbf{\Sigma} = \mathbf{VDV}^T$, the whitening matrix $\mathbf{P}$ is given by:

$$\mathbf{P} = \mathbf{VD}^{-1/2}\mathbf{V}^T \tag{1}$$

Where $\mathbf{V}$ is one orthogonal matrix and $\mathbf{D}$ is one diagonal matrix.

## 3.3. Fuzzy C-Means Clustering Algorithm

FCM clustering algorithm is one of simplest objective function-based clustering algorithm to solve the clustering questions through adopting fuzzy mechanism for data, in relative terms, the fundamental theory of FCM clustering is extremely mature. The form objective function of the FCM clustering algorithm is usually described as follows:

$$J_m(\mathbf{U},\mathbf{P}) = \sum_{k=1}^{n}\sum_{i=1}^{c}\mu_{ik}^{m}(d_{ik})^2, m \in [1,\infty)$$
$$(d_{ik})^2 = \|x_k - p_i\|_{\mathbf{A}} = (x_k - p_i)^T\mathbf{A}(x_k - p_i) \tag{2}$$

Where $J_m(U,P)$ denotes the weighted square sum of distances between the data samples and the centers of clusters, $m$ is the weighted index number whose value is any real number greater than 1, $n$ and $c$ separately represents the number of data samples and clusters, $\mu_{ik}$ is the degree of membership of $x_k$ for the cluster $i$, $\mathbf{U} = \left[\mu_{ik}\right]_{c \times n}$ denotes membership matrix, constrained to that $\sum_{i=1}^{c}\mu_{ik} = 1$ for each $x_k$ data sample, $p_i$ is the $i$-th center of the cluster, and $d_{ik}$ indicates the Euclidean distance between the data sample $x_k$ and the center $p_i$ with $\mathbf{A}$ denoting a symmetric positive definite matrix [20].

FCM algorithm is based on minimizing $J_m(U,P)$ for dividing the data samples and its steps can be summarized in Table 1.

### 3.4. WR-FCM Algorithm

Firstly, in off-line phase, the preprocessing of gathering RSS data samples is conducted by ZCA whitening methods, removing the correlation of the signals, to build the fingerprint database. Then divide the whole database into $c$ sub-databases space for matching through FCM clustering algorithm and obtain the matrix $\mathbf{U}$ and $\mathbf{P}$ whose elements are represented separately the degree of membership of each fingerprint in $c$ sub-databases and the centers of $c$ clusters. In on-line phase, first of all, categorize the user into the only one corresponding cluster, which is chosen as the delegate for positioning while others eliminated, according to calculating the Euclidean distances between the detected real-time data sample and clustering centers. Subsequently, the estimated coordinates of the user is acquired by leveraging weighted KNN algorithm in the selected subspace, where the value of membership of this fingerprint in the sub-database are taken as a weight parameter. The process of WR -FCM is presented in Table 2.

### Table 1. FCM Clustering Algorithm

**Step1:** Initialize $\mathbf{U}^{(0)}$ and $\mathbf{P}^{(0)}$ randomly, form iterative counter $t = 0$, input the number of clusters $c$ and positive threshold $\varepsilon$ .

**Step2:** Calculate the membership matrix $\mathbf{U}^{(t)}$ of the degree of membership $\mu_{ik}$ by

$$\mu_{ik} = \left\{ \Sigma_{j=1}^{c} \left( \frac{d_{ik}}{d_{jk}} \right)^{\frac{2}{m-1}} \right\}^{-1} , \forall i, k$$

(3)

**Step3:** Calculate $\mathbf{P}^{(t)}$ of the center of cluster $p_i$ by

$$p_i = \frac{\Sigma_{k=1}^{n} (\mu_{ik})^m \cdot x_k}{\Sigma_{k=1}^{n} (\mu_{ik})^m}, i = 1, 2, \cdots c$$

(4)

**Step4:** If $\left\| \mathbf{P}^{(t+1)} - \mathbf{P}^{(t)} \right\| > \varepsilon$ , then set up $t = t + 1$ and turn to step 2. Otherwise, stop the algorithm and export $\mathbf{U}$ and $\mathbf{P}$ .

### Table 2. WR-FCM Algorithm

**Step1:** Scan the RSS samples $\mathbf{X}$ which is the $d \times N$ matrix and its columns are $x_t$ .

**Step2:** Apply ZCA whitening on the whole dataset $\mathbf{X}$, subtracting out the mean and dividing by the standard deviation, transforming $\mathbf{X}$ to $\tilde{\mathbf{X}}$ and build the fingerprint database.

**Step3:** Apply Fuzzy C-Means clustering algorithm on the fingerprint database, classifying the database to $c$ clusters and obtain the membership matrix $\mathbf{U}$ and the center matrix $\mathbf{P}$ .

**Step4:** Choose one sub-database as the delegate, named $\Omega$ , which has most similar RSS characters with the target.

**Step5:** Select $k$ reference points from $\Omega$ as the neighbors by the Euclidean distance $d_{ik}$ between the real-time detected fingerprint and the fingerprints from $\Omega$ .

**Step6:** Calculate the estimated location coordinates $(x, y)$ as follows:

$$(x,y) = \frac{1}{K} \sum_{t=1}^{K} \left( \frac{\mu_{it}(x_t, y_t)}{\sum_{t=1}^{k} \mu_{it}} \right) \qquad (5)$$

Where, $\mu_{it}$ is the degree of membership of sample $x_t$ in the $i$-th cluster, namely chosen delegate, and $(x_t, y_t)$ is the coordinates of samples in $\Omega$.

## 4. Experimental Evaluation

In this section we present a series of experiments to evaluate the performance of WR-FCM algorithm. We first describe the datasets used in the experiments and the setup. Then, we compare the positioning performance of WR-FCM in terms of positioning error cumulative probability with other four approaches including indoor positioning algorithm based on KNN, K-Means and PCA.

### 4.1. Datasets

Our experiment was conducted at the third floor of the No.13 building in Shandong University of Science and Technology where previously it was established comprehensive wireless networks infrastructure such as the access points of CMCC and China-Unicom. Therefore, 26 available APs were selected from the deployed facilities and totally 32 points over four corridors were chosen as reference points while the distance of adjacent reference points is nearly 1.5m. The experimental environment, the corridor area on the 3rd floor, is as shown in Figure 1 which marked with blue.
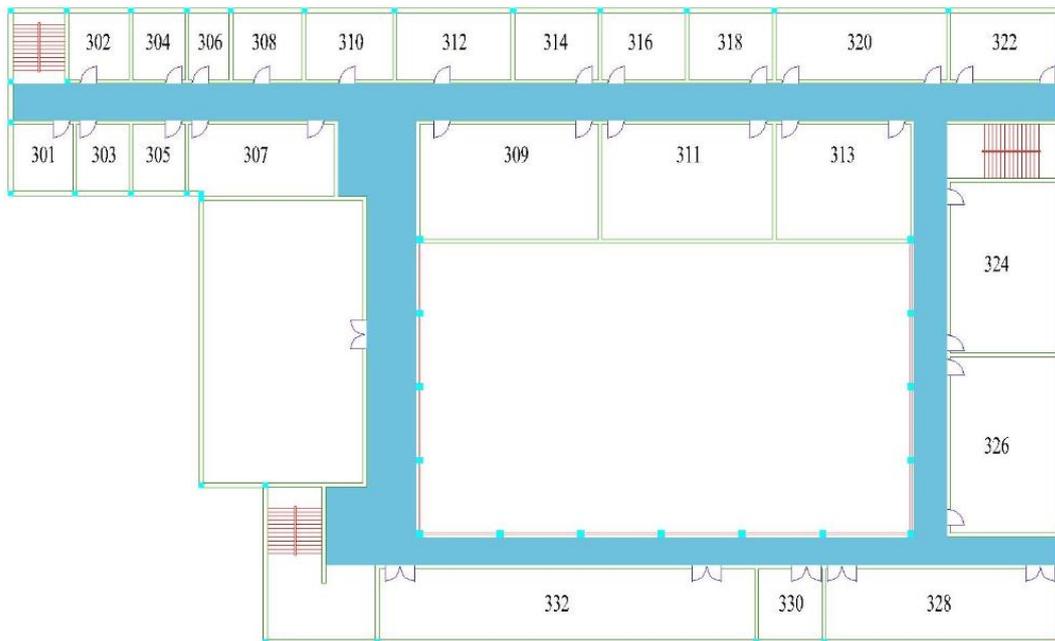


**Figure 1. Experimental Environment**

We record 50 RSS samples at each point so that 1600 samples were stored as the training dataset in off-line phase. Moreover, we chose eight points as test points where 400 samples were collected as the test dataset during on-line phase. All the samples above were obtained by HUAWEI Honor 6, which is a kind of public mobile phone. Figure2 recorded the real work of collecting data samples.

**Figure 2. The Real Scene of Collecting Samples**

The performance is measured by cumulative percentage of positioning error distance whose equation as follows:

$$\text{Dis}_{\text{error}} = \sqrt{\left((\tilde{x},\tilde{y})-(x,y)\right)^2}$$

(6)

### 4.2 Results

We compared our proposed model WR-FCM with the indoor positioning algorithm based on KNN, K-Means [8] and PCA [12]. We set the parameter $m$ of FCM clustering in equation (3) as 2, and the number of clusters was given by 3 because of its well performance in clustering time and precision. In addition, the number of neighbors $k$ is determined to be 4 empirically. The Figure 2 illustrates the contradistinction of WR-FCM with the other four approaches through cumulative probability distribution of error distance.
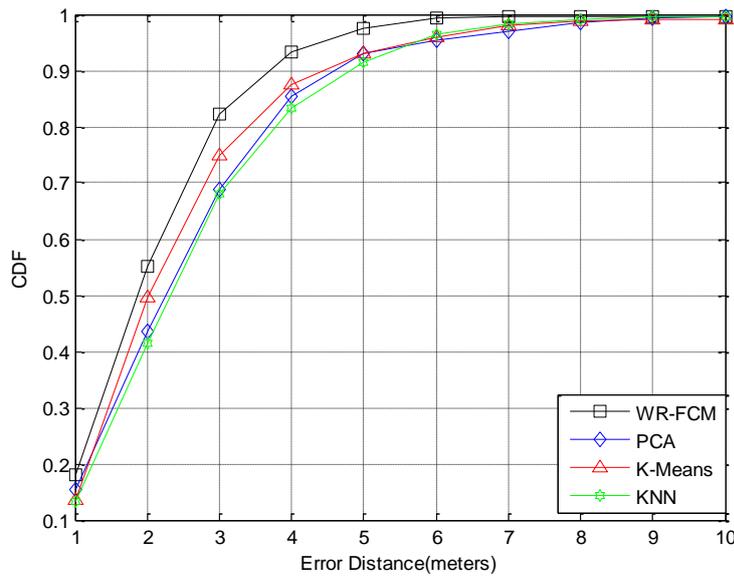
**Figure 3. Cumulative Probability Distribution of Error Distance**

The figure clearly reports our proposed approach outperforms other approaches. The errors within 2 meters of KNN, K-Means and PCA are 41.4, 49.6, 43.5percent while our approach obtains 55.1%. Furthermore, the mean error is decreased by 42.6 percent as compared to KNN algorithm with 1.53 meters versus 2.67 meters.

Due to the computation complexity is in direct proportion to the size of search space, generally, for positioning algorithm with matching type, the computation complexity is decreased effectively by reducing the search space through clustering. In our approach, FCM clustering was applied to reduce the search space so that the computation complexity of our approach was decreased from $O(N^2)$ to $O(M^2)$ where $(M<<N)$ qualitatively.

Table 3 reports the comparison of WR-FCM with the other four approaches on the computing time in quantitative analysis.

**Table 3. Comparison of Computing Time**

| Algorithm | Time(seconds) |
|-----------|---------------|
| KNN | 0.43 |
| WR-FCM | **0.188** |

As shown in Figure 3 and Table 3, comprehensively, WR-FCM algorithm not only dramatically increases the positioning accuracy, but also reduce the computing time in on-line phase effectively.

## 5. Conclusion and Future Work

In this paper, we proposed a novel indoor positioning approach, WR-FCM algorithm, based on ZCA whitening and Fuzzy C-Means algorithm. The whitening technique greatly removes the redundancy between RSS detected from different APs and FCM clustering partitions whole search region to various sub space thus effectively reduces the

computation time-consuming and improves positioning accuracy during on-line phase. The experimental results indicate that our proposed model outperforms other usual methods.

Actually, some drawbacks also exist in this research, for instance, the site survey in our experiment is labor-intensive and time-consuming and the heterogeneous issues leading to loss of precision. Existing literatures [21][22][23] have applied Crowdsourcing to indoor positioning overcoming the above-mentioned problems, a few better achievements obtained. Which, therefore, puts forward the research direction and emphasis in the future work.

## Acknowledgements

## References

[1]  S. He, S.-H. G. Chan, "Wi-Fi fingerprint-based indoor positioning: recent advances and comparisons", IEEE Communications Surveys and Tutorials, vol. 18, no. 1, **(2016)**, pp. 466-490.

[2]  Y. Q. Xie, Y. Wang, P. C. Zhu and X. H. You, "Grid-search-based hybrid TOA/AOA location techniques for NLOS environments", IEEE Communications Letters, vol. 13, no. 4, **(2009)**, pp. 254-256.

[3]  H. Zou, H. Wang, L. Xie and Q. S. Jia, "An RFID indoor positioning system by using weighted path loss and extreme learning machine", Proceedings of IEEE 1st International Conference on Cyber-Physical Systems, Networks, and Applications (CPSNA), **(2013)**; Taipei, China.

[4]  H. B. Liu, G. Yu, J. Yang, S. Simon, Y. Wang, Y. Y. Chen and F. Ye, "Push the limit of WiFi based localization for smartphones", Proceedings of the 18th annual international conference on Mobile computing and network, ACM MobiCom, **(2012)**; Istanbul, Turkey.

[5]  Y. Tsuda, Q. Kong and T. Maekawa, "Detecting and correcting WiFi positioning errors", Proceedings of the international joint conference on Pervasive and ubiquitous computing, ACM UbiComp, **(2013)**; Zurich, Switzerland.

[6]  M. A. Youssef, A. Agrawala and A. U. Shankar, "WLAN location determination via clustering and probability distributions", Proceedings of the First IEEE International Conference on Pervasive Computing and Communications, **(2003)**.

[7]  J. Ma, X. S. Li, X. Tao and J. Lu, "Cluster filtered KNN: A WLAN-based indoor positioning scheme", Proceedings of International Symposium on a World of Wireless, Mobile and Multimedia Networks, **(2008)**; Newport Beach, CA.

[8]  B. Altintas and T. Serif, "Improving RSS-Based Indoor Positioning Algorithm via K-Means Clustering", Proceedings of the 11th European Wireless Conference-Sustainable Wireless Technologies, **(2011)**; Vienna, Austria.

[9]  A. Razavi, M. Valkama and E. S. Lohan, "K-Means Fingerprint Clustering for Low-Complexity Floor Estimation in Indoor Mobile Localization", IEEE Globecom Workshop on Localization and Tracking: Indoors, Outdoors and Emerging Networks, **(2015)**.

[10] Y. L. Sun, Y. B. Xu, L. Ma and Z. A. Deng, "KNN-FCM Hybrid Algorithm for Indoor Location in WLAN", Proceedings of the 2nd International Conference on Power Electronics and Intelligent Transportation System, Shenzhen, **(2009)**.

[11] H. Zhou and N. N. Van, "Indoor Fingerprint Localization Based on Fuzzy C-means Clustering", Proceedings of the sixth International Conference on Measuring Technology and Mechatronics Automation, **(2014)**.

[12] S. H. Fang, T. N. Lin and P. C. Lin, "Location fingerprinting in a decorrelated space", IEEE Transaction on Knowledge and Data Engineering, vol. 20, no. 5, **(2008)**, pp. 685-691.

[13] S. H. Fang, P. C. Lin and T. N. Lin, "Indoor Location by a novel probabilistic approach", Proceedings of IEEE 8th Workshop on Signal Processing Advances in Wireless Communications, **(2007)**.

[14] S. H. Fang, T. N. Lin and P. C. Lin, "Principal Component Localization in Indoor WLAN Environments", IEEE Transaction on Mobile and Computing, vol. 11, no. 1, **(2012)**, pp. 100-110.

[15] L. N. Chen, B. H. Li, K. Zhao, C. Rizos and Z. Q. Zheng, "An improved algorithm to generate a Wi-Fi fingerprint database for Indoor Positioning", Sensors, vol. 13, no. 8, **(2013)**, pp. 11085-11096.

[16] A. K. M. Hossain and W. Soh, "A survey of calibration-free indoor positioning systems", Journal of Computer Communications, vol. 66, **(2015)**, pp. 1-13.

[17] P. Bahl and V. N. Padmanabhan, "RADAR: An in-building RF-based user location and tracking system", Proceedings of Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies, Tel.Aviv, **(2000)**.

[18] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images", Citeseer, **(2009)**, pp. 48-50.

[19] A. Coates and A. Y. Ng, "Selecting receptive fields in deep networks", Proceedings of Neural Information Processing Systems, **(2011)**.

[20] M. Sayadi, L. Tlig and F. Fnaiech, "A New Texture Segmentation Method Based on the Fuzzy C-Means Algorithm and statistical Features", Applied Mathematical Sciences, vol. 1, no. 60, **(2007)**, pp. 2999-3007.

[21] A. Rai, K. K. Chintalapudi and V. N. Padmanabhan, "Zee: zero-effort crowdsourcing for indoor localization", Proceedings of the 18th annual international conference on Mobile computing and networking, ACM MobiCom, **(2012)**; Istanbul, Turkey.

[22] Wu, Z. Yang and Y. Liu, "Smartphones Based Crowdsourcing for Indoor Localization", IEEE Transaction on Mobile Computing, vol. 14, no. 2, **(2015)**, pp. 444-457.

[23] K. Chang and D. Han, "Crowdsourcing-based radio map update automation for Wi-Fi positioning systems", Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Crowdsourced and Volunteered Geographic Information, ACM GeoCrowd, **(2014)**.