Performance analysis of adaptive K for weighted K-nearest neighbor based indoor positioning

Siyang Liu, Raul de Lacerda, Jocelyn Fiorina

▶ **To cite this version:**

Siyang Liu, Raul de Lacerda, Jocelyn Fiorina. Performance analysis of adaptive K for weighted K-nearest neighbor based indoor positioning. 2022 IEEE 95th Vehicular Technology Conference (VTC2022-Spring), Jun 2022, Helsinki, Finland. 10.1109/vtc2022-spring54318.2022.9860699 . hal-03697741

HAL Id: hal-03697741
https://hal.science/hal-03697741

Submitted on 17 Jun 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Performance analysis of adaptive K for weighted K-nearest neighbor based indoor positioning

Siyang Liu
Laboratoire des Signaux et Systèmes
Université Paris-Saclay, CNRS, CentraleSupélec
Gif-sur-Yvette, France
siyang.liu@centralesupelec.fr

Raul de Lacerda
Laboratoire des Signaux et Systèmes
Université Paris-Saclay, CNRS, CentraleSupélec
Gif-sur-Yvette, France
raul.delacerda@centralesupelec.fr

Jocelyn Fiorina
Laboratoire des Signaux et Systèmes
Université Paris-Saclay, CNRS, CentraleSupélec
Gif-sur-Yvette, France
jocelyn.fiorina@centralesupelec.fr

***Abstract*—With the large deployment of WiFi networks, indoor localization using WiFi fingerprinting with received signal strength has been widely studied. One of the common localization methods is weighted K-nearest neighbor method (WKNN), which localizes the user to the weighted center of the K best matching points. The performance of this method is affected by the choice of parameter K. Once tuned, this parameter is usually applied to all test samples. In this paper, we study how far localization performance can be improved if this parameter is adapted for different test points. We show with two public access datasets that adapting parameter K for different test points can potentially improve localization performance by over 45% compared to the baseline of only choosing the closest neighbor. Additionally, we analyze the dataset to obtain some stochastic thresholds for dataset filtering and K selection.**

***Index Terms*—WiFi Fingerprinting, RSS, WKNN**

I. Introduction

With the rapid development of IoT, the need for indoor localization has also increased [1]. Many techniques have been applied for indoor localization such as Time of Arrival (TOA), Received Signal Strength (RSS), Angle of Arrival (AOA), Channel State Information (CSI), etc [2]. Among these techniques, RSS is an accessible information which can be obtained passively without needing additional hardware or information exchange.

The existing localization methods can be roughly divided into two categories: model-based and fingerprinting based methods [3]. Model based methods estimate distance between user and anchors with propagation models and then perform triangulation. Fingerprinting methods explore the mapping between positions and unique features such as RSS from different WiFi access points or the CSI and localize the user by matching its measurements to the dataset.

Various algorithms based on machine learning techniques have been proposed for fingerprinting. Deep learning methods train neural networks with RSS measurements and position labels. Benefiting from the non-linear activation function and adapting ability, these methods have shown excellent results in region classification with sufficient training data and computation power [4]–[8]. There are other algorithms that provide good localization performance with less data and resource requirement, such as Nearest Neighbor (NN) [9], [10], Support Vector Machine (SVM) [11].

This work is financially supported by China Scholarship Council from the Ministry of Education of P.R. China.

K Nearest Neighbor (KNN) algorithm as well as its variant WKNN localize the user by finding K best matching neighbor fingerprints and estimate the user position as the body center of these neighbor points. The matching process can be done with similarity functions [9], [10] or other data representation methods [4], [5], [12]. By tuning parameter K, which is the number of nearest neighbors considered, performance of KNN can be improved comparing to only taking the nearest neighbor into account [13].

Most papers in the literature consider a fix value of K for all test samples. Localization performance can be further improved if the choice of K is individualized to different test samples. To adapt K value, an iterative method can be applied, increasing K from 1 to a preset maximum value K_{max} and checking if the newly included point meets certain requirements. In [14], [15], a new neighbor is only included when the distance between the user's RSS vector and fingerprint of that neighbor point is below a certain threshold. Additional to the RSS distance threshold, a limit can be apply on the physical distance [16], [17]. In [18] a wireless propagation model was estimated with neighbor fingerprints and fingerprints are selected to minimize the error of estimation. In [19], a clustering based selection method is proposed for positioning with visible light communication.

These above methods show that an adaptive WKNN can provide better performance than classic WKNN. However, they do not show how far the performance can be improved by adapting K. In this paper, we obtain the limit of adaptive WKNN by solving a NP-complete problem and we show that there is large potential on performance improvement. Additionally, we propose to analysis the dataset during offline phase and obtain some stochastic parameters to help choosing K.

The remainder of the paper is organized as follows: Section II presents system model and Section III presents the method to obtain the performance limit of adapting WKNN. Section IV analyzes the dataset for K selection. Section V presents numerical results and Section VI concludes the paper.

II. System model

The fingerprinting process including two phases is shown in Fig. 1. During the offline phase, a fingerprint dataset is built by measuring the received power from different Access Points (AP) at N_r predefined Reference Points (RP). Assuming there are M fixed APs in the dataset, a RSS sample i is constituted by the Received Signal Strength Indicator (RSSI) from all M APs $RSS_i = [r_i^1, r_i^2, ..., r_i^M]$. A fingerprint F_i consists the RSS vector and the corresponding position label.

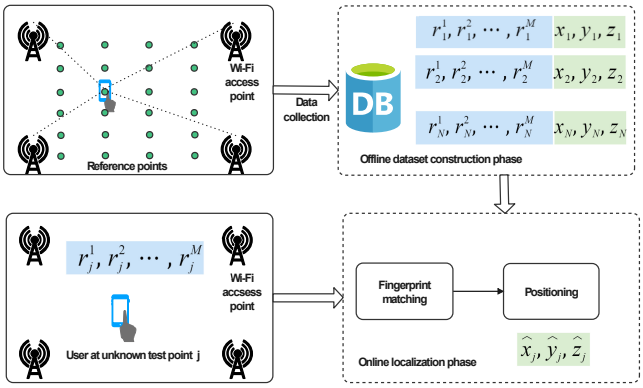

Figure 1. RSS Fingerprinting

During the online phase, a test sample searches for matching fingerprints in the dataset with a similarity function which quantifies how close two samples are in the feature space. Commonly used similarity measures include Euclidean distance, Cosine similarity, Manhattan distance, Pearson Correlation Coefficient,etc [20]. Using cosine similarity function to demonstrate, similarity value between RSS sample i and j can be obtained as:

$$s_{i,j} = \frac{\sum_{m=1}^{M} r_i^m r_j^m}{\sqrt{\sum_{m=1}^{M}(r_i^m)^2}\sqrt{\sum_{m=1}^{M}(r_j^m)^2}}. \quad (1)$$

To simplified the calculation of similarity values, positive representation is obtained for RSS vector with thresholding as:

$$P(r_i^m) = \begin{cases} r_i^m - min, & if \quad AP_m \quad is \quad present \quad in \quad F_i \\ & and \quad r_i^m > \tau \\ 0, & otherwise, \end{cases} \quad (2)$$

where min is the lowest RSSI value in the dataset and τ is the threshold value which removes the RSSI values that are too low .

For test sample j, K fingerprints with highest similarity are selected whose coordinates are denoted as (x_k, y_k, z_k) $(k = 1, ..., K)$. WKNN algorithm estimates test sample position as the weighted body center of these selected points.

The weight function allows better matching fingerprints to contribute more to the position estimation. In this paper, the weight function is chosen as:

$$w_k = s_{i,k}^2 \quad (k = 1, ..., K). \quad (3)$$

III. Performance limit of WKNN with adaptive K

It is shown in the literature that adapting K for WKNN can provide better localization performance than using a fix K. In [16], [17], [19], performance improvement is achieved by 1.43%, 1.70% and 8.96% on their own RSS fingerprinting datasets, respectively. However, the full potential improvement we can achieve with adaptive WKNN is yet unknown.

In this section, we obtain the performance limit of adaptive WKNN by optimizing localization performance for each test sample. This process is shown in Algorithm 1, in which N_r is the number of RPs in the dataset. Given the RSS vector of a user and the range of $[1, K_{max}]$, the value of K that provides minimum localization error would be the optimal K for this test sample. The mean error of all samples given their own optimal K would be the performance limit we can achieve with adaptive WKNN.

Algorithm 1 Pseudo code for obtaining performance limit of adaptive WKNN

Require: F_j: fingerprint measured at test point j; **F**: fingerprints from the dataset; K_{max}: the range of K for WKNN optimization;

Ensure: optimal K value K_{opt}^j and performance limit err_{opt}^j;

set $K_{opt}^j = 1$, compute initial localization error err_{opt}^j with $K = 1$;
for $i = 1$; $i \leq N_r$; $i++$ **do**
    compute similarity value $s_{j,i}$ between user fingerprint and all RPs in the dataset;
end for
sort RPs by decreasing order of similarity value;
for $k = 1$; $k \leq K_{max}$; $k++$ **do**
    compute user estimated position $(\hat{x}_j^k, \hat{y}_j^k, \hat{z}_j^k)$ and localization error err_j^k
    if $err_j^k < err_{opt}^j$ **then**
        $err_{opt}^j = err_j^k$, $K_{opt}^j = k$
    end if
end for

Using Algorithm 1, performance limit on two public access WiFi RSSI datasets UJIIndoorLoc [21] and Alcala Tutorial 2017 Dataset [22] are obtained. To show the ratio of improvement, a baseline is obtained by giving a uniform value of K to all test samples which corresponds to the classic WKNN.

We first show the performance limit using cosine similarity function for WKNN. As shown in Fig.2 and Fig.3, on both datasets, localization error obtained by taking the optimal K for each test sample is significantly reduced as the limit K_{max} increases. Comparing the result with optimal K and uniform K, localization error is reduced by over 3 meters for $K_{max} > 15$ in the UJIIndoorLoc dataset and for $K_{max} > 10$ in the Alcala dataset, which shows great potential of performance improvement if K is properly selected for each test sample.

For comparison, localization performance is also obtained with a random K selection. Given a predefined range of $[1, K_{max}]$, we can randomly choose K for a given test sample

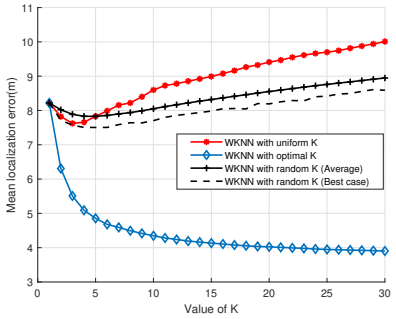

Figure 2. Performance limit with adaptive WKNN:UJIIndoorLoc dataset

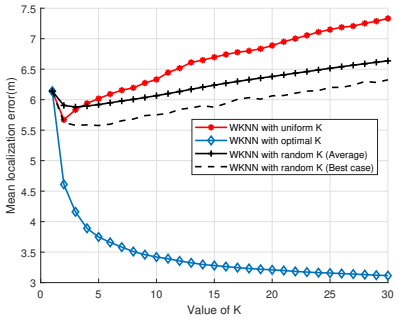

Figure 3. Performance limit with adaptive WKNN:Alcala dataset

with equal probability for each value of K in the range. In Fig.2 and Fig.3, black curves correspond to the baseline of random K selection presenting the mean and the minimum localization error among 3000 independent trials. For a larger K_{max}, random K performs better than fix K on average.

We also obtain the performance limit with Algorithm 1 using other common similarity functions [20] shown in Fig.4. On both datasets, Cosine similarity and Pearson's correlation coefficient provide better performance than the other two functions. In this paper, we will keep using Cosine similarity function because it is relatively easier to calculate.

The performance limit introduced in this section is obtained by solving the NP complete problem. We identify the optimal K with the help of the test point position label. This is not realistic since the user position is generally unknown. However, this limit shows the huge potential improvement of adapting parameter K and it leads to the open question of how we can approach this limit by selecting an individual K for each test sample.

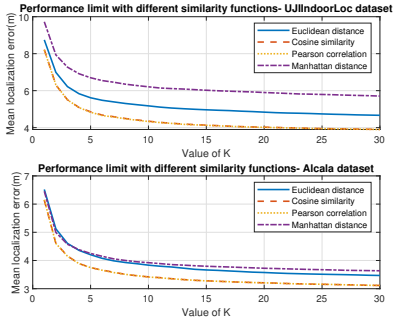

Figure 4. Performance limit with different similarity functions

IV. Adaptive K based on dataset filtering

In this section, we discuss the problem of K selection and analyze training set to provide stochastic thresholds that can be used for K selection. By setting a range of K and applying these thresholds, some fingerprints that are not well matching the test sample can be excluded and hence adapting the number of neighbors used for positioning.

From the results in Fig.2 and Fig.3, we can see that with an uniform K, to a certain extend, increasing K compensates the data incompleteness of the discrete dataset as well as mitigating the error introduced by imperfect RSS. However, when K becomes relatively large, the algorithm might select far away fingerprints which can introduce large error because the spatial relation of RSS vectors are not perfect. Therefore, we analyze the training set and study the relation between mathematical and physical space to get some insights for the localization process.

In the section II, similarity function is introduced to search for matching fingerprints in the dataset. Fingerprints with higher similarity is expected to be closer to the test sample. However, RSS is easily affected by the environment and it can fluctuate quite a lot even in the same position [23]. Even though it is hard to determine the exact distance between two points given their RSS similarity value, we can obtain some stochastic thresholds from analyzing the training set during the offline phase.

On both UJIIndoorLoc and Alcala datasets, we obtain the relation between similarity value and physical distance. The empirical mean of cosine similarity over different physical distance as well as the 95% confidence interval are shown in Fig.5 and Fig.6. As shown in both figures, in general the mean value of similarity decreases as horizontal physical distance increases which is in line with the assumption of similarity based localization methods.

Even with the large variance, we can still see that fingerprints taken from close distance have relatively high similarity values. For instance, 95% fingerprints within 10 meters radius have similarity value above 0.5 and 0.25 in UJIIndoorLoc and Alcala dataset, respectively. This information can be used to determined whether or not a neighbor fingerprint is a good match. If the similarity value between the test sample and a fingerprint is below this threshold, this fingerprint is likely to be far away from the test point and we can choose to not take it into account for localization. This similarity threshold can be noted as τ_s.

With the above similarity threshold, we can filter out fingerprints that are likely to be far away. However, given the large variance of similarity values, we still can not be sure that all fingerprints with higher similarity than τ_s are good matching points. For instance, in Fig.5, fingerprints with

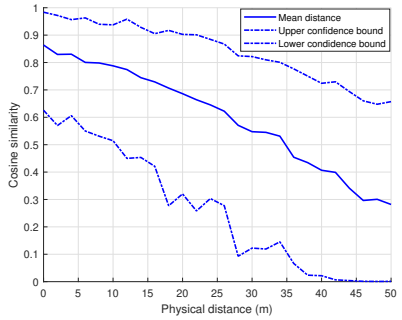

Figure 5. Mean cosine similarity over physical distance: UJIIndoorLoc dataset

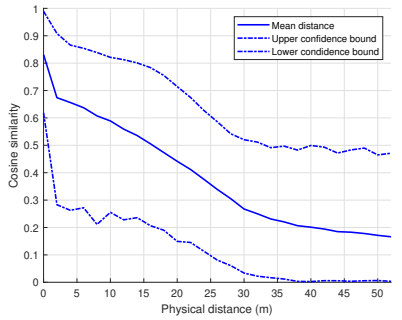

Figure 6. Mean cosine similarity over physical distance: Alcala dataset

similarity values of 0.6 can be taken from physical distance of 1 meter or much further. Therefore, we introduce another threshold τ_p. Using the highest similarity value in the dataset as reference, we try to make sure that neighbor fingerprints used for localization are matched to the test sample at a similar level. This threshold can be tuned with experiments.

The proposed adaptive WKNN with dataset filtering is detailed as follow:

Algorithm 2 Adaptive WKNN with dataset filtering

Require: F_j: fingerprint of test point j; **F**: fingerprints from the dataset; K'_{max}: the predefined range of K for adaptive WKNN;

Ensure: Estimated position of test point j

1: **for** $i = 1$; $i \leq N_r$; $i++$ **do**
2: Compute similarity value $s_{j,i}$ between user fingerprint and all RPs in the dataset;
3: **end for**
4: Sort similarity values of different RPs in a descending order and note the K'_{max} highest values as $s'_{j,k}$, $k = 1, 2, ...K'_{max}$;
5: **for** $k = 1$; $k \leq K'_{max}$; $k++$ **do**
6: **if** $(s'_{j,k} < \tau_s)$ or $(s'_{j,k} < s'_{j,1}\tau_p)$ **then**
7: break;
8: **end if**
9: **end for**
10: Estimate test point position with WKNN (K=k).

V. Simulation Results

To evaluate the performance limit of the proposed K selection method, the upper limit K_{max} need to be chosen. As shown in Fig.2 and Fig.3, for the classic WKNN, $K = 3$ and $K = 2$ provide the best result for the corresponding dataset, so we would choose these two as the limit of iteration and try to improve performance by filtering and reducing K for certain test samples. Similarity threshold τ_s can be obtained by dataset analysis in the previous section. From the lower confidence bound of 10 meter distance, τ_s for UJIIndoorLoc and Alcala dataset is set to be 0.5 and 0.25, respectively.

The last parameter τ_p can be tuned by experiment. Table I presents localization error with different threshold value. 0.95 and 0.85 are chosen as parameter τ_p for the corresponding dataset.

Table I
Localization error with different τ_p

τ_p	**0.80**	**0.85**	**0.90**	**0.95**	**0.99**
UJIIndoorLoc (m)	7.5	7.4	7.4	**7.3**	7.6
Alcala (m)	5.6	**5.5**	5.6	5.7	5.8

Performance comparison of several adaptive WKNN methods as well the performance limit obtained as section III are shown in Table. II. The performance limit drawn with the proposed method shows great potential on improving localization performance of WKNN by optimizing the K for each test sample. The proposed dataset filtering method improves localization performance by applying simple thresholds on similarity value. On the UJIIndoorLoc dataset, mean localization error is reduced by 11% and on the Alcala dataset 9.8% comparing to the classic nearest neighbor method.

Table II
Localization performance Comparison

Dataset	**Method**	***Mean error(m)***	***Improvement(%)***
UJIIndoorLoc	NN	8.2	0
UJIIndoorLoc	WKNN(K=3)	7.7	6.1
UJIIndoorLoc	Limit	3.9	**52.4**
UJIIndoorLoc	Dataset filtering	7.3	11.0
Alcala	NN	6.1	0
Alcala	WKNN(K=2)	5.7	6.6
Alcala	Limit	3.3	**45.9**
Alcala	Dataset filtering	5.5	9.8

VI. Conclusion and Perspective

In this paper, we tackle the problem of how adapting parameter K in the classic WKNN method can affect fingerprinting performance. On two public access datasets, we show that if the parameter K is optimized for each test sample instead of setting a uniform value, a huge improvement over 45% and 52% can be achieved, which is a lot more than what the existing methods in the literature offer.

We also analysis the dataset to get insights for filtering the dataset and selecting K. Even though the current improvement

is not yet close to the performance limit we obtained, as future work, we plan to improve the adaptive method by further investigating the relation among selected fingerprints.

.

REFERENCES

[1] F. Zafari, A. Gkelias and K. K. Leung, 'A survey of indoor localization systems and technologies,' *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2568–2599, 2019.

[2] H. Obeidat, W. Shuaieb, O. Obeidat and R. Abd-Alhameed, 'A review of indoor localization techniques and wireless technologies,' *Wireless Personal Communications*, pp. 1–39, 2021.

[3] M. T. Hoang, B. Yuen, X. Dong, T. Lu, R. Westendorp and K. Reddy, 'Recurrent neural networks for accurate rssi indoor localization,' *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 10 639–10 651, 2019.

[4] X. Song, X. Fan, C. Xiang, Q. Ye, L. Liu, Z. Wang, X. He, N. Yang and G. Fang, 'A novel convolutional neural network based indoor localization framework with wifi fingerprinting,' *IEEE Access*, vol. 7, pp. 110 698–110 709, 2019.

[5] P. Wu, T. Imbiriba, J. Park, S. Kim and P. Closas, 'Personalized federated learning over non-iid data for indoor localization,' *arXiv preprint arXiv:2107.04189*, 2021.

[6] N. Singh, S. Choe and R. Punmiya, 'Machine learning based indoor localization using wi-fi rssi fingerprints: An overview,' *IEEE Access*, 2021.

[7] S. Xia, Y. Liu, G. Yuan, M. Zhu and Z. Wang, 'Indoor fingerprint positioning based on wi-fi: An overview,' *ISPRS International Journal of Geo-Information*, vol. 6, no. 5, p. 135, 2017.

[8] P. Wu, T. Imbiriba, G. LaMountain, J. Vilà-Valls and P. Closas, 'Wifi fingerprinting and tracking using neural networks,' in *Proceedings of the 32nd International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2019)*, 2019, pp. 2314–2324.

[9] P. Roy, C. Chowdhury, M. Kundu, D. Ghosh and S. Bandyopadhyay, 'Novel weighted ensemble classifier for smartphone based indoor localization,' *Expert Systems with Applications*, vol. 164, p. 113 758, 2021.

[10] J. Torres-Sospedra, P. Richter, A. Moreira, G. Mendoza-Silva, E.-S. Lohan, S. Trilles, M. Matey-Sanz and J. Huerta, 'A comprehensive and reproducible comparison of clustering and optimization rules in wi-fi fingerprinting,' *IEEE Transactions on Mobile Computing*, 2020.

[11] H. M. Aydin, M. A. Ali and E. G. Soyak, 'The analysis of feature selection with machine learning for indoor positioning,' in *2021 29th Signal Processing and Communications Applications Conference (SIU)*, IEEE, 2021, pp. 1–4.

[12] X. Wang, L. Gao, S. Mao and S. Pandey, 'Csi-based fingerprinting for indoor localization: A deep learning approach,' *IEEE Transactions on Vehicular Technology*, vol. 66, no. 1, pp. 763–776, 2016.

[13] L. Siyang, R. DE LACERDA and J. FIORINA, 'Wknn indoor wi-fi localization method using k-means clustering based radio mapping,' in *2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring)*, IEEE, 2021, pp. 1–5.

[14] J. Hu, H. Liu and D. Liu, 'Toward a dynamic k in k-nearest neighbor fingerprint indoor positioning,' in *2018 IEEE International Conference on Information Reuse and Integration (IRI)*, IEEE, 2018, pp. 308–314.

[15] J. Hu, D. Liu, Z. Yan and H. Liu, 'Experimental analysis on weight k-nearest neighbor indoor fingerprint positioning,' *IEEE Internet of Things Journal*, vol. 6, no. 1, pp. 891–897, 2018.

[16] I. Lee, M. Kwak and D. Han, 'A dynamic k-nearest neighbor method for wlan-based positioning systems,' *Journal of Computer Information Systems*, vol. 56, no. 4, pp. 295–300, 2016.

[17] J. Oh and J. Kim, 'Adaptive k-nearest neighbour algorithm for wifi fingerprint positioning,' *Ict Express*, vol. 4, no. 2, pp. 91–94, 2018.

[18] S. Xu, C.-C. Chen, Y. Wu, X. Wang and F. Wei, 'Adaptive residual weighted k-nearest neighbor fingerprint positioning algorithm based on visible light communication,' *Sensors*, vol. 20, no. 16, p. 4432, 2020.

[19] J. Ma, X. Li, X. Tao and J. Lu, 'Cluster filtered knn: A wlan-based indoor positioning scheme,' in *2008 International Symposium on a World of Wireless, Mobile and Multimedia Networks*, IEEE, 2008, pp. 1–8.

[20] J. Torres-Sospedra, R. Montoliu, S. Trilles, Ó. Belmonte and J. Huerta, 'Comprehensive analysis of distance and similarity measures for wi-fi fingerprinting indoor positioning systems,' *Expert Systems with Applications*, vol. 42, no. 23, pp. 9263–9278, 2015.

[21] J. Torres-Sospedra, R. Montoliu, A. Martínez-Usó, J. P. Avariento, T. J. Arnau, M. Benedito-Bordonau and J. Huerta, 'Ujiindoorloc: A new multi-building and multi-floor database for wlan fingerprint-based indoor localization problems,' in *2014 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2014, pp. 261–270.

[22] E. Sansano, R. Montoliu, O. Belmonte and J. Torres-Sospedra, *Uji indoor positioning and navigation repository*, 2016. [Online]. Available: http : / / indoorlocplatform.uji.es.

[23] X. Tian, R. Shen, D. Liu, Y. Wen and X. Wang, 'Performance analysis of rss fingerprinting based indoor localization,' *IEEE Transactions on Mobile Computing*, vol. 16, no. 10, pp. 2847–2861, 2017.