Improving RSS-Based Indoor Positioning Algorithm via K-Means Clustering

Conference Paper · January 2011

2 authors:

Bulut Altintas
Yeditepe University

Tacha Serif
Yeditepe University

Improving RSS-Based Indoor Positioning Algorithm via K-Means Clustering

Bulut Altintas
Department of Computer Engineering,
Yeditepe University,
Istanbul, Turkey
baltintas@cse.yeditepe.edu.tr

Tacha Serif
Department of Computer Engineering,
Yeditepe University,
Istanbul, Turkey
tserif@cse.yeditepe.edu.tr

***Abstract*—Advances in mobile technologies and devices has changed the way users interact with devices and other users. These new interaction methods and services are offered by the help of intelligent sensing capabilities, using context, location and motion sensors. However, indoor location sensing is mostly achieved by utilizing radio signal (Wi-Fi, Bluetooth, GSM etc.) and nearest neighbor identification. The most common algorithm adopted for Received Signal Strength (RSS)-based location sensing is K Nearest Neighbor (KNN), which calculates K nearest neighboring points to mobile users (MUs). Accordingly, in this paper, we aim to improve the KNN algorithm by enhancing the neighboring point selection by applying k-means clustering approach. In the proposed method, k-means clustering algorithm groups nearest neighbors according to their distance to mobile user. Then the closest group to the mobile user is used to calculate the MU's location. The evaluation results indicate that the performance of clustered KNN is closely tied to the number of clusters, number of neighbors to be clustered and the initiation of the center points in k-mean algorithm.**

Keywords-component; Received signal strength, k-Means, clustering, location estimation, personal digital assistant (PDA), wireless, indoor positioning

I. INTRODUCTION

The penetration of portable devices such as notebooks, netbooks, iPads, PDAs, and smart phones in our lives, has caused mobile and ad-hoc services to become increasingly popular. Due to the mobile nature of these devices, context and location sensing capabilities not only increase their usability but also entrench intelligence.

WLAN-based location sensing, and especially fingerprinting technique, has been seen as a viable solution for indoor location identification where global positioning systems fall short. K Nearest Neighbors (KNN) is the most common algorithm utilized for fingerprinting, which employs K nearest-neighbors to predict user location. The algorithm can be improved by modifying and refining the selection process of K nearest neighbors utilizing clustering and dynamic selection methods. Clustered KNN algorithms and dynamic KNN algorithms have been developed in the past and could be found in the literature. However, to the best of our knowledge, no previous work has utilized k-means clustering algorithm to improve the KNN algorithm performance.

Accordingly, the structure of this paper is as follows: Section II introduces the keystones and main methods of indoor positioning techniques and reviews of some existing improved KNN algorithms, while Section III presents the ideas behind the location sensing and provides a more detailed overview of the algorithms used. Section IV expands on the tests undertaken and on their findings. Finally, conclusions are drawn and opportunities for future work identified in the last section.

II. RELATED WORK

There are several methods that could be used to sense (or calculate) location. Today, the most popular location detection method is using the Global Positioning System (GPS); however this technology is limited to outdoor environments, since GPS receivers requires a clear view of the sky to calculate its location [1]. To overcome this limitation, other techniques and technologies such as ultrasound [2], infrared [3], RFID tags [4] and radio signals [5] have been developed and evaluated for indoor location sensing. Nevertheless, among all these technologies, due to its penetration and availability, IEEE 802.11b/g/n (Wi-Fi) radio signals and access points (APs) have been one of the most popular [6].

Various Wi-Fi-based approaches that use radio frequency signal to measure mobile user's distance to APs have been implemented and tested. Namely, these approaches are Time-of-Arrival (TOA) [7], Time-Difference-of-Arrival (TDOA) [8], Angle-of-Arrival (AOA) [9] and Received-Signal-Strength (RSS) [10].

Implementing TOA, TDOA and AOA is cumbersome and costly tasks, since they all require hardware and/or software modifications on APs [11]. On the other hand, RSS-based location detection can be implemented without any modification on the AP end; furthermore, collecting RSS data is a straightforward task to perform. Currently one of the most esteemed solutions for RSS-based positioning is the fingerprinting technique.

This approach utilizes deterministic or probabilistic method to determine the user position. Deterministic method calculates the distances to the reference points in signal space. However, probabilistic method makes a comparison between the probabilistic distribution of signal strengths at reference points and real-time signal strength readings of MU. Accordingly, most suitable reference points are chosen by K-Nearest Neighbor (KNN) algorithm – a basic algorithm for positioning which is widely used in fingerprinting technique – and then user position is obtained. The performance of the KNN can be improved extensively by deploying auxiliary selection method or clustering algorithms.

One of positioning systems that utilizes clustering algorithm was developed by Ma et al. [12]. In this work, which was named Cluster Filtered KNN (CFK), KNN algorithm was improved by utilizing clustering technique to partition the neighbors into multiple clusters and then one these clusters was chosen as a delegate. The results showed that KNN improved with average-linkage agglomerative Hierarchical Clustering (HC) outperforms KNN. Similarly, Sun et al. [13] developed KNN-FCM hybrid algorithm for WLAN based indoor positioning systems. In order to improve the KNN performance, they applied fuzzy c-means (FCM) clustering algorithm. K-nearest neighbors determined by KNN algorithm are classified into numerous clusters through FCM and one of them is chosen to calculate user position. Their simulation results indicate that KNN-FCM hybrid algorithm generally has better results than KNN when the distance error is less than 2 meters. From a different perspective, Roshanaei and Maleki [14] improved KNN algorithm by selecting the best number of nearest neighbor as a K value dynamically. Dynamic-KNN (D-KNN) algorithm, which combines AOA and KNN algorithm as a hybrid method, utilizes the adaptive antenna system to determine the user location area by intersection of several obtained AOAs. Thus, selection of the K neighbors within the intersecting AOAs results with the best K values. Their simulation results show that D-KNN outperforms KNN.

III. Indoor Positioning System Design

The proposed system utilizes fingerprinting technique and k-means clustering algorithm which are detailed in this section. Because of its simplicity and accuracy of results, deterministic approach of fingerprinting technique was selected [15]. A proposed initialization algorithm for k-means clustering is developed and also explained in the following part.

A. Fingerprinting Technique

This technique is composed of two phases: Training (Offline) phase and Tracking (Online) phase. During the training phase (Fig. 1.a), signal strengths from APs are collected at pre-identified locations, which are called reference points (RPs). The objective of this operation is building the fingerprint database which will be used in the tracking phase. Because mobile user's location is determined based on the surrounding RPs, they should be distributed in the target area evenly and homogenously.

In the tracking phase (Fig 1.b), MU's surrounding AP RSSs are compared with the RPs dataset collected in the training phase to identify the best matching RPs. The tracking phase could use deterministic and probabilistic algorithms to match real-time RSS readings with RPs signal data.

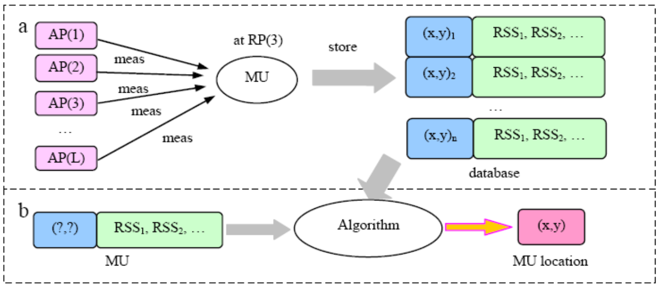

Figure 1. Two phases of fingerprinting (a) training phase and (b) tracking phase

Deterministic approach utilizes "K Nearest Neighbor (KNN)" or its variation "K Weighted Nearest Neighbor (KWNN)". These two algorithms use Euclidean distance (Eq. 1 where q = 2) in signal space to identify K nearest RPs by comparing MU's real-time RSS readings.

The signal distance between the real-time RSS readings vector $[s_1, s_2 \ldots, s_n]$ and the RSS vector in the database $[S_1, S_2 \ldots, S_n]$ is computed by applying Eq. 1.

$$L_q = \left(\sum_{1}^{n} |s_i - S_i|^q \right)^{1/q} \tag{1}$$

The quantity L_q is a positive real value, where a lower value indicates a smaller difference between the two compared vectors. KNN algorithm ranks the list of RPs in ascending order by using the resulting L_q and then takes the direct or weighted average of the K-nearest neighbor coordinates. The average of the coordinates (x,y) can be used to estimate MU's location.

B. K-Means Clustering Algorithm

In order to improve the KNN algorithm performance we deploy k-means clustering algorithm. K-means is one of the simplest learning algorithms that solve the well known clustering problem.

It classifies a given data set through a certain number of clusters. The algorithm aims to find the centre point of a cluster by minimizing the distance between the cluster centre and members of the same cluster. Suppose we have a set of RPs as nearest neighbors, which are determined by KNN algorithm, x_j, j=1,,N and we would like to organize them into K clusters $C = C_1, C_2, .., C_k$. The algorithm is composed of the following steps in Table 1 [16].

TABLE I. K-Means Clustering Algorithm

1. Initialize K centroid points which represent initial group centre point (centroid).
2. Calculate distances between RPs and centroids.
3. Assign each RP to a cluster that has the closest centroid.
4. When all RPs are assigned, recalculate clusters centroids.
5. Repeat step 2, 3 and 4 until there is no change for each cluster.

After clustering operation is completed, the average distance between RPs and mobile user are calculated for each cluster. Then, the cluster with the lowest average value, cluster with the closest proximity to mobile user, is favored as the delegate cluster to determine the user's current position. The current position of user is obtained from centroid of delegate cluster.

The performance of the k-means algorithm varies based on the number of clusters, k, number of points in data set, q, and initialization of centroids. In our system, to prevent abundant calculations, we set the number of clusters to two. Accordingly, in our experiments we aim to identify the optimal number of RPs to be employed in the clusters

The k-means algorithm does not necessarily find the most optimal clustering sets. The algorithm is significantly sensitive to the selection of the centroids and the number of the clusters. Therefore, the approach taken to initialize K centroids and the number of clusters are important. One popular initialization approach is to randomly choose centroid points. On the other hand initialization can also be performed after obtaining information on the arrangement of the points by a coarse analysis of the data set.

In order to analyze the arrangement of the RP set, a dynamic initialization algorithm is developed in our system. This algorithm attempts to place centroids as far as possible from each other. To achieve this, algorithm performs coarse grouping of RPs by their IDs, which are assigned consecutively to neighboring RPs. It takes the average of the RPs ID numbers in the data set, then assign nearest RP ID to the average value as first centroid and furthest RP as second centroid.

The algorithm first specifies a set of RPs which are relative closer to each other than the remaining RPs. Following the creation of closest RP set, one centroid is selected from the members and one centroid is selected from the non-members of this set.

IV. EXPERIMENTS & EVALUATION

In the project, existing WLAN infrastructure in the school building was utilized. This WLAN included three D-Link Wireless APs, one U.S. Robotics AP and one Cisco Aironet 340 series AP. In the development process, the project was tested on three different PDAs. These were Samsung i900, HP IPAQ 6515 and HP IPAQ 5500. The system was developed on Visual Studio 2008 and .NET compact framework was used on the PDAs.

The indoor positioning system test bed was set up at our university's school building. The test bed was 500 sq. meters (approx. 5381 sq. ft.) area, which was covered by five 802.11b/g APs which are marked with blue stars and 57 RPs that are marked with a red cross (Fig. 2).

All tests are performed stationary at pre-identified points in test bed area. These tests are aimed to observe of performance variation with various q values – the number of RPs within a cluster. Tests are performed with q value of 5, 7 and 9. For each q values, a number of measurements were carried out at 150 different pre-identified test locations by utilizing the fingerprinting algorithm and the findings were compared with actual coordinates within the building.

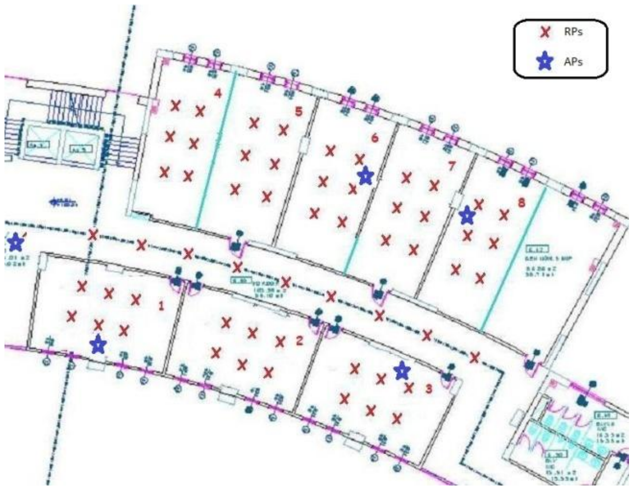

Figure 2. Floor plan of the test bed and RPs

In order to compare the KNN algorithm with clustered KNN algorithm, user position is determined by using both KNN algorithm with K = 4 and Clustered KNN algorithm with same Signal Strengths that are obtained from mobile users current point.

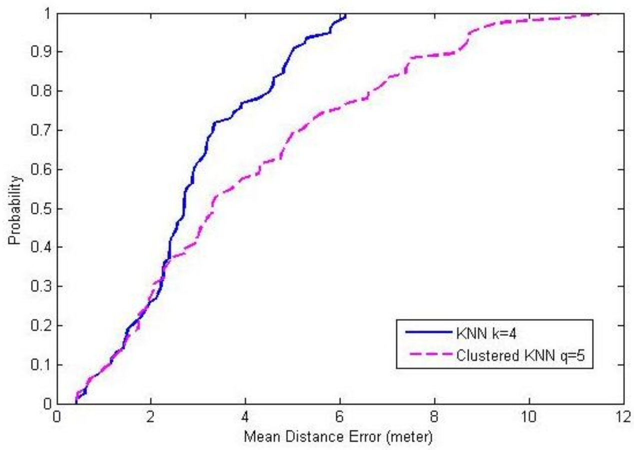

Figure 3. Comparison between Clustered KNN with $q = 5$ and KNN

As shown in Fig. 3, KNN algorithm performs better than Clustered KNN algorithm when q is equal to 5, since delegate cluster consists of two or three RPs. As a result, this scenario pans out with a less accurate prediction of user's location. This is caused by severe fluctuations in the received signal strength for mobile user even at fixed location. In such case, some distant RPs can be determined as nearest neighbors even if they are not. Nevertheless, miscalculated neighbors can be compensated by averaging three or four nearest neighbors. Thus, K is mostly selected as three or four in KNN fingerprinting algorithms. Hence, clustered KNN algorithm with two RPs cannot compensate miscalculated neighbors. This can be clearly seen in Figure 3, where it shows a correlation between clustered KNN algorithm and KNN algorithm up to 2 meters distance error. This is the case because distance error under 2 meters means that delegate cluster rarely includes miscalculated neighbors.

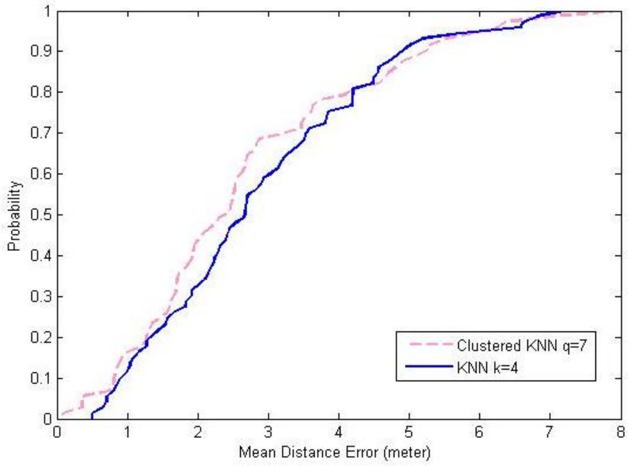

Figure 4. Comparision between Clustered KNN with $q = 7$ and KNN

Clustered KNN algorithm with q equal to 7 outperforms KNN algorithm as shown in Fig. 4. The reasons for this is the size of the delegate cluster, which mostly it contains at least three or four RPs. As a result miscalculated neighbors can be eliminated by averaging these RPs. Moreover, the clustering algorithm attempts to coarse grouping by utilizing initialization algorithm described in previous section. This initialization algorithm eliminates the miscalculated nearest neighbors. Consequently, filtering miscalculated RPs within the nearest neighbors set in two stages help the system to predict user position more accurately than KNN algorithm.

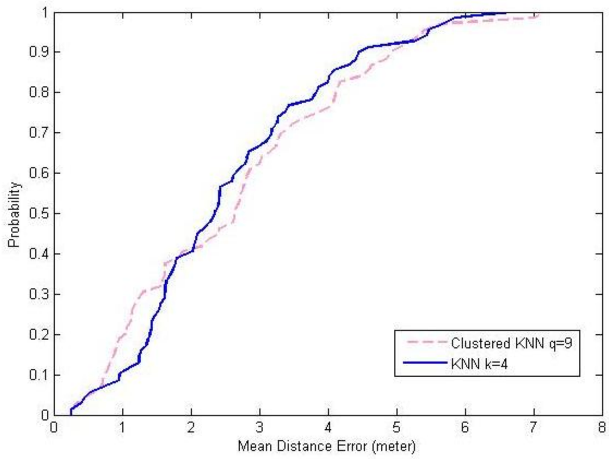

Figure 5. Comparison between Clustered KNN with $q = 9$ and KNN

Our experiments show that when q equal to 9, clustered KNN algorithm performs worse than KNN algorithm. The number of RPs in delegate cluster is around five and six. By setting q to nine we enlarge the selected group size thus increasing the probability of miscalculated nearest neighbors. As a result, the possibility of eliminating miscalculated neighbors caused by signal strengths fluctuations decreases. Therefore, more than one wrongly selected neighbor in delegate cluster cannot be compensated; consequently, user position cannot be predicted as accurate as KNN algorithm. However, Figure 5 also indicates that clustered KNN algorithm outperforms KNN when distance errors are less than two meters. We believe this is due to fewer numbers of miscalculated neighbors.

TABLE II. AVERAGE DESTANCE ERRORS & STANDARD DEVIATION OF TEST RESULTS

	Clustered KNN		KNN	
	Mean	Std. Dev.	Mean	Std. Dev.
q=5	4,11	2,7	2,9	1,45
q=7	2,7	1,73	2,89	1,6
q=9	2,68	1,66	2,60	1,44

From a different perspective, distance error mean of clustered KNN algorithm with varying q values can be compared with its relative KNN algorithm results. To conduct these tests, a single signal strength is measured and distances to RPs is calculated with both clustered KNN algorithm and KNN algorithm. For example average distance error of clustered KNN algorithm with q is equal 7 is approximately 2,7 meters and with the same received signal strengths average distance error of KNN algorithm is approximately 2.9 meters (Table 2).

Mean distance errors and standard deviations of test results show that standard deviation of clustered KNN algorithm is larger than of KNN algorithm. This is the case even if mean distance error of clustered KNN is less than KNN algorithm. We believe that this largely depends on the initialization algorithm employed select the foremost centroids. Depending on the accuracy of the initially selected centroids, k-means clustering algorithm either produce more acceptable results or more inaccurate results by selecting a delegate cluster which is far away from the mobile user's actual location.

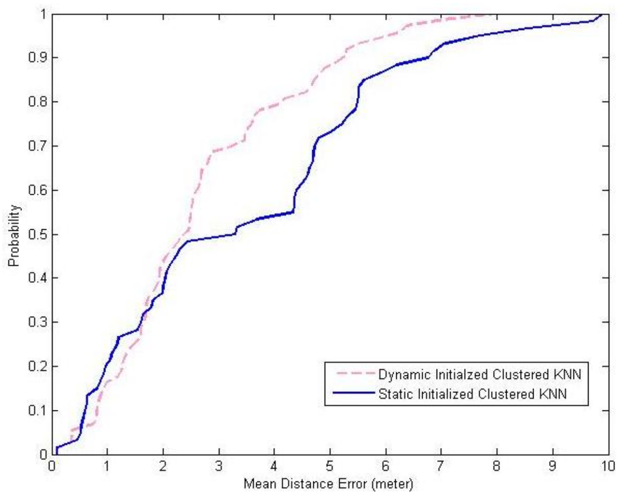

Figure 6. Comparision between static initialization and dynamic initialization

In this study, finally, we compare the static and dynamic initialization process of centroids and investigate how this affects the k-means clustering algorithm. Since our typical initialization process is conducted dynamically, we already had the accuracy data set ready for this group. However, to collect the accuracy rate for static initialization, two centroids where identified based on their distance to the mobile user. One centroid was chosen to be the closest neighbor to the mobile user and the second was the most distant. Subsequently, when the k-means algorithm was employed, the results indicated that dynamically initialized centroids outperformed statically initialized centroids as shown in Fig. 6. This confirms that coarse grouping of adjacent RPs method improves and refines the selection of accurate centroids.

V. CONCLUSION

In this paper, we evaluated a WLAN positioning approach that utilized clustered KNN to estimate user location in indoor environments. The clustered KNN algorithm improves the performance of KNN algorithm by eliminating miscalculated nearest neighbors. The developed algorithm utilized k-means clustering with a coarse initialization method to select the best nearest neighbors.

The results obtained throughout our evaluation indicate that the clustered KNN algorithm outperforms KNN algorithm when the number of nearest neighbors to be clustered is selected accurately. However, the standard deviation of clustered KNN algorithm is greater than KNN algorithm due to the coarse initialization algorithm applied prior to k-means clustering.

Accordingly, as future work, the proposed system could be improved by utilizing probabilistic approach of KNN and implementing filtering methods for more accurate location estimation. Furthermore, a calibration free radio map generation could be developed to decrease the workload on the offline phase of fingerprinting technique.

REFERENCES

[1] G. Sun, J. Chen, W. Guo, and K.J.R. Liu, "Signal processing techniques in network-aided positioning: a survey of state-of the-art positioning design." IEEE Signal Processing Magazine, vol. 22, no. 4, pp 12-23, 2005.

[2] M. Hazas and J. Ward, "A novel broadband ultrasonic location system," Proc. 4th Int. Conf. Ubiquitous Computing, pp. 264-280, 2002.

[3] R. Want, A. Hopper, V. Falcao and J. Gibbons, "The active badge location system," ACM Transactions on Information Systems, vol. 10, no 1, pp. 91-102, 1992.

[4] L. M. Ni, Y. Liu, Y.C. Lau and A. P. Patil, "Landmarc: indoor location sensing using active rfid," First IEEE International Conference on Pervasive Computing and Communications, pp. 407-415, 2003

[5] V. Honkavirta, T. Perala, S. Ali and R: Piche, "A comparative survey of WLAN location fingerprinting methods," Proceedings of the 6th Workshop on Positioning, Navigation and Communication, 2009.

[6] P. Bahl and V. N. Padmanabhan, "RADAR: an in-building RF-based user location and tracking system". NFOCOM 2000. 19th Annual Joint Conference of the IEEE Computer and Communications Societies, vol. 2, pp. 775-784, 2000.

[7] M. Llombart, M. Ciurana and F. Barcelo-Arroyo, " On the scability of a novel WLAN positioning system based on time of arrival measurements," Proceedings of the 6th Workshop on Positioning, Navigation and Communication, 2008

[8] C. Yang, Y: Huang and X. Zhu, "Hybrid TDOA/AOA method for indoor positioning systems," Location Technologies, 2007

[9] R. Yamasaki, A. Ogino, T. Tamaki, T. Uta, N. Matsuzawa and T. Kato, "TDOA location system for IEEE 802.11 b WLAN," IEEE Wireless Communications and Networking Conference, vol. 4, pp. 2338-2343, 2005.

[10] S. Mazuelas, A. Bahillo, R.M. Lorenzo, P. Fernandez, F. A. Lago, E. Gracia, J. Blas and E. J. Abril, "Robust indoor positioning provided by real-time RSSI values in unmodified WLAN networks," IEEE Journal of Selected Topics in Signal Processing, vol. 3, no. 5, 2009.

[11] S.A. Golden and S. S. Bateman, "Sensor measurements for Wi-Fi location with emphasis on time-of-arrival ranging," IEEE Transc. Mobile Computing, vol. 6, no. 10, pp. 1185-1198, 2007.

[12] J. Ma, X. Li, X. Tao, and J. Lu, "Cluster filtered KNN: A WLAN-based indoor positioning scheme," World of Wireless, Mobile and Multimedia Networks WoWMoM, pp. 1-8, 2008.

[13] Y. Sun, Y. Xu, L. Ma and Z. Deng, "KNN-FCM Hybrid Algorithm for Indoor Location in WLAN," Power Electronics and Intelligent Transportation System PEITS, vol. 2, pp. 251-254, 2009.

[14] M. Roshanaei and M. Maleki, "Dynamic-KNN: A novel locating method in WLAN based on Angle of Arrival," Industrial Electronics & Applications ISIEA, vol. 2, pp. 722-726, 2009.

[15] B. Li, J. Salter, A. G. Dempster and C. Rizos, "Indoor positioning techniques based on Wireless LAN," Auswireless Conference, pp. 1-7, 2006.

[16] R. Xu, and D. Wunsch, "Survey of clustering algorithms. IEEE Transactions on Neural Networks," vol. 16, pp. 645–678, 2005.