PURPOSE-LED
PUBLISHING™

PAPER • OPEN ACCESS

RSSI-based Localization Zoning using K-Mean Clustering

View the article online for updates and enhancements.

RSSI-based Localization Zoning using K-Mean Clustering

Kenny Fong Peng Wye[1,2]**, Ericson Kanagaraj**[2,3]**, Syed Muhammad Mamduh Syed Zakaria**[1,2]**, Latifah Munirah Kamarudin**[2,3]**, Ammar Zakaria**[2,4]**, Kamarulzaman Kamarudin**[2,4] **and Norhawati Ahmad**[1]

[1] School of Microelectronic Engineering, Universiti Malaysia Perlis (UniMAP), Kampus Pauh Putra, Arau, Perlis, 02600, Malaysia
[2] Centre of Excellence for Advanced Sensor Technology (CEASTech), Universiti Malaysia Perlis (UniMAP), Kompleks Pengajian Jejawi II, Arau, Perlis, 02600, Malaysia
[3] School of Computer and Communications Engineering, Universiti Malaysia Perlis (UniMAP), Kampus Pauh Putra, Arau, Perlis, 02600, Malaysia
[4] School of Mechatronic Engineering, Universiti Malaysia Perlis (UniMAP), Kampus Pauh Putra, Arau, Perlis, 02600, Malaysia
[5] School of Manufacturing Engineering, Universiti Malaysia Perlis (UniMAP), Kampus Pauh Putra, Arau, Perlis, 02600, Malaysia

kennyfong@ieee.org

Abstract. This document discusses the novel approach to localize human location based on the current zone via k-mean clustering. A pilot experimental analysis of k-mean clustering to group similar RSSI pattern is compared to user defined zones. The dataset collected has demonstrated the difficulties of deploying trilateration and fingerprinting methods in dynamic conditions, as it introduces fluctuating RSSI measurements. The k-mean clustering was proposed to validate antenna placement to divide the testbed into zones, and to create a baseline accuracy which can be compared with other algorithms in the future. It was found that, for Zones 2, 3 and 4, the k-mean clustering of antennas agree with the planned antenna placement grouping. However, for Zone 1, there are differences, which may be attributed to a large metal obstacle occupying the zone. The k-mean clustering also recorded a peak accuracy of 79% with k=4, which agrees with the number of planned zones.

1. Introduction

The maturity of the Global Positioning System (GPS) has allowed the implementation of various remote time-critical applications. However, GPS is limited in the sense that it requires Line-of-Sight (LOS) between a minimum of 4 satellites and the receiver to achieve accurate localization. This issue is exacerbated further indoors where it is impossible to receive LOS signals from any satellite. The dependence on LOS transmission to the satellite renders indoor localization using GPS inaccurate and unreliable.

Therefore, researchers try to find alternatives for indoor localization where Non-Line-of-Sight (NLOS) conditions are prevalent. Many research works reported use devices such as RFID, ZigBee, WiFi, MEMS and others for RF based localization. The measurements made to locate trackers can be generalized into Time of Arrival (ToA), Angle of Arrival (AoA), Time Difference of Arrival (TDoA), and Received Signal Strength Indicator (RSSI) [1].

Lately, Received Signal Strength (RSS) has been the focus of researchers due as the RSSI functions are built in the devices and require no additional hardware to support the localization technique [2]. Although RSSI location technique does not require additional hardware, the RSS is prone to multipath effect and interferences in the complex indoor environments [3].

Another solution is implementing the RSSI localization measurement that uses fingerprinting. The fingerprinting technique can achieve a high level of accuracy in localization. However, it requires a high amount of time and labour intensive for reference point database construction for each deployment environment. The fingerprinting technique has two stages; offline and online stage. The offline stage is the collection of each unique RSSI value at the reference points to form a database for the online stage. In the online stage, location can be predicted based on the database and matched the new unknown location RSSI value with the predefined RSSI database previously [4].

However in most practical applications, high accuracy localization is not required. Most applications need area-based localization. In these sorts of applications, the actual position of the tracked object is not required, but rather to ensure the objects are within a certain area. Rather than having accurate position of the object, users are more interested to know with absolute certainty that the tracked object is within a certain area. For example, a shopkeeper would want to know if certain products are available in certain zones. Likewise, in a workshop, supervisors only need to know whether a worker is in their work zone rather the accurate position of the worker due to mobile nature of the workshop job. High accuracy position information does not deliver any added value to the monitoring supervisors. Hence, what is needed is a method to locate tracked objects in zones with high certainty and reliability.

2. Literature Review & Related Work

Previously, k-mean clustering algorithm was used to improve the accuracy of the position function in NLOS environment [5]. The position function is established with TOA based localization in LOS environment and eliminating the unwanted NLOS measurement. The author suggested that NLOS error due to the complexity of environment can cause large path loss in a signal propagation model is the main reason of the error reported. To improve the accuracy of the localization technique, k-mean clustering can be used to identify and eliminate unwanted NLOS error value to improve the localization model.

RF signal strength has also been reported to be cost-advantageous compared to other techniques such as AOA, TOA, TDOA [6]. The author uses k-mean clustering algorithm to improve the accuracy of the indoor location sensing similar to fingerprinting method. It was pointed out that if a RSSI value that has not been fingerprinted, location cannot be identified. The RSSI value would be added into the specific cluster which has the lowest Euclidean distance value between the RSSI and the centroid to estimate locations which were not fingerprinted.

In another work, Berz et al. proposed an indoor positioning system with RFID technology to estimate the item-level location of stationary objects. The Support Vector Regression machine learning is integrated into a fingerprinting indoor positioning system. To further improve the accuracy of the model, k-mean approach is applied similar to Subaashini [6]. Again, the RSSI values collected from the same tag suffers from interferences, obstacles and multipath effects that would affect the localization of the tag. Shukri et al. demonstrated the fluctuations of signal strength in RSSI based localization using the IRIS motes [7].

Razavi et al. introduced a solution of time-consuming data collection for fingerprinting RSS technique with k-mean clustering algorithm [8]. The proposed solutions achieved a similar accuracy with Nearest-Neighbor (NN) fingerprinting. The k-mean clustering algorithm is used for floor estimation in indoor mobile localization. Bai et al. suggested that fingerprinting techniques in indoor localization system suffers from heavy computation [4]. The heavy computation could be reduced with k-mean clustering and the average searching cost becomes 1/K. Since fingerprinting requires different access point (AP) to form the fingerprinting database, it could result in too many different dimensions for every anchor's fingerprint. Author suggested that Principal Components Analysis should be performed in those fingerprints and limits the dimension's property when clustering [4].

Previously work has been done on indoor propagation and detection of occupants via RSSI. The nature of indoor propagation has been studied in depth and has demonstrated that RSSI fluctuations are affected by number of occupants and obstacles [9]. A number of methods have been proposed to detect and track occupants based on RSSI variations [7]. Various algorithms have also been proposed for control of transmission to ensure reliability in transmission indoors and outdoors environment [10,11]. While the proposed methods have been able to perform to similar accuracies to methods from other researchers, further probing needs to be done to explore other modalities of RF-based localization. As part of

continuing work, an active RFID system was deployed at a warehouse in an industrial environment. The work presented in this manuscript uses this testbed to explore new approaches to tracking and localization.

3. Experimental Setup

The testbed used in this research is a large workshop environment. The workshop is divided into different zones, based on different activities done in the workshop. The placement of the directional Antenna 1-12 in the workshop is shown in Fig. 1; placed as such to cover the areas of interest labelled Zone 1-4. The antennas are placed 30m above ground, directed horizontally across the zones.
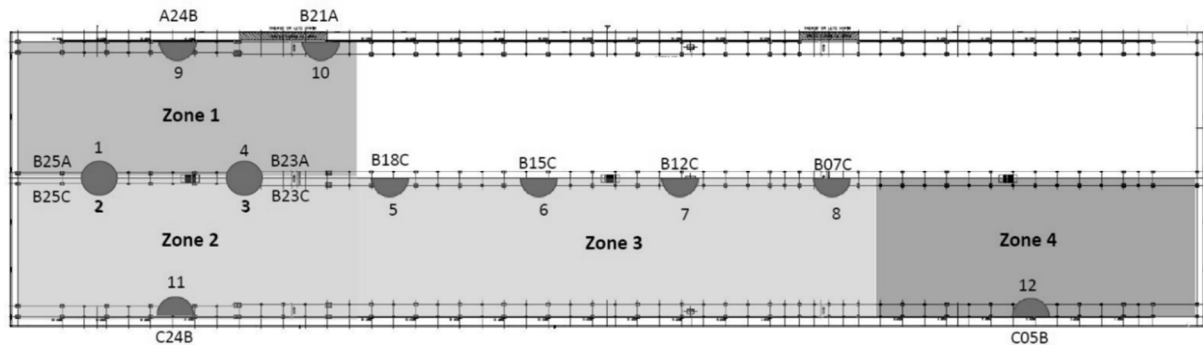

Figure 1. Testbed for the experiment environment

3.1. RSSI Mapping

RSSI mapping was done to ensure the coverage of the antenna in all the zones. The measurement position for Antenna 8 is shown in Fig. 2. From this experiment, the coverage and the strength of the signal received by the antenna from different parts of the area can be observed. The selection of measurement point was affected by the ongoing work in the workshop. Hence some parts of the area were not measured.
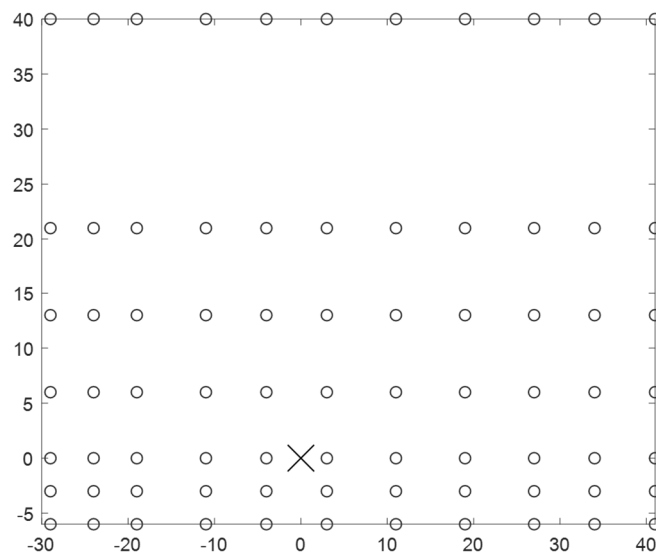

Figure 2. Measurement points for Antenna 8 RSSI mapping

3.2. K-Mean Clustering

k-mean clustering is an unsupervised machine learning that cluster data into k exclusive clusters. k-mean operates on the observations from the dataset supplied to create clusters. Each of the clusters is created where similar data is partitioned together and other non-similar data are grouped as another cluster. Each of the clusters is defined by the similarity and the centroid of the cluster. The centroid for each cluster is the point at which the sum of distances from all members in that cluster is minimized.

Based on Fig. 1, there is a total of 4 zones in which different activities are conducted. Hence, the placement of the antennas to locate objects in these zones needs to be verified to be adequate, and, the baseline reliability and accuracy of such zone-based localization needs to be established. Hence, k-mean clustering is a suitable method to demonstrate both of these objectives.

The RSSI data stored in the database is divided into user-defined number of clusters in the k-mean clustering. To determine the optimal number of clusters, k-mean model was built with initial cluster from 0 to 13 and observing the elbow pattern. Then, the accuracy of the clusters in the elbow region was studied and compared.

The RSSI of 4 transmitters was collected while it is moved within and between the different zones. The actual time and zones of the transmitter were recorded. This data is fed into the k-mean clustering method to evaluate whether the RSSI data can be used to distinguish transmission from the 4 different zones. By clustering the recorded RSSI, the natural coverage of zones by the antenna can be determined to be suitable to its intended purpose. The grouping of the k-mean as reflected by the centroids derived for each antenna would demonstrate whether the zones are suitably covered by its corresponding antenna.

3.3. Min-Max Normalization

Before the data is fed into the k-mean clustering, the data is normalized using Min-Max method. The Min-Max Normalization used is as shown in Equation 1.

$$RSSI' = 1 - \frac{RSSI - RSSI_{min}}{RSSI_{max} - RSSI_{min}}(newMax - newMin) + newMin \tag{1}$$

The normalized RSSI value is represented by RSSI' as shown in Equation 1. The maximum and minimum value of RSSI from 12 antennas is used for the normalization. The new maximum value after the normalization is 1 and the new minimum value is 0. In this Min-Max normalization, the stronger signal strength should be more approaching to 1 and vice versa.

4. Results and Discussions

The objective of this paper is to propose a novel approach for zoning-based localization with unsupervised learning. The unsupervised learning model used in this paper is k-mean clustering. As mentioned, fingerprinting required time consuming and labour-intensive data collection to build the database. Based on the measurement shown in Fig. 2, a contour graph is plotted in Fig. 3.
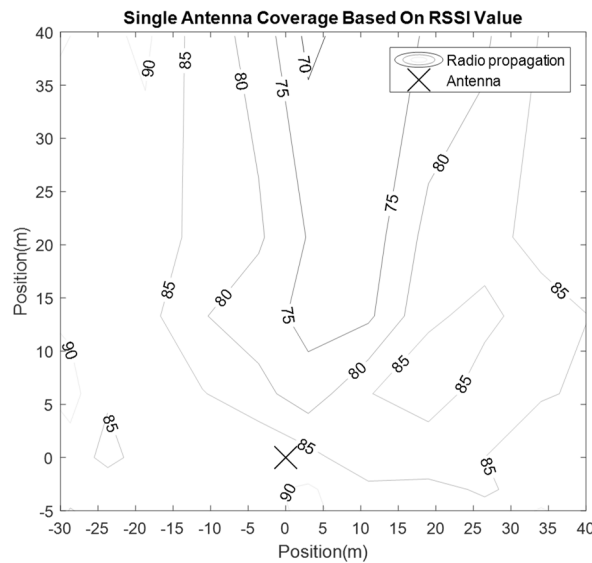
Figure 3. Measurement for Antenna 8 RSSI

The collected data shows some unexpected peaks and troughs in the distribution, due to the fluctuation observed during the measurement. Since there were many metal obstacles in the environment, some measured areas, for example (20, 10), drops in RSSI was recorded. The results demonstrate that in dynamic environments, especially with metal obstacles, RSSI based localization techniques such as trilateration or fingerprinting may not be feasible.

Then, data was collected while transmitters were moved in and between the different zones. The collected data was fed into k-mean model as training data. The elbow graph for k cluster is plotted in Fig. 4(a). Based on Fig. 4(a), the elbow pattern can be observed when k=3 until k=6. The accuracy of the model is further analysed and compared to the reference positioning as plotted in Fig. 4(b).
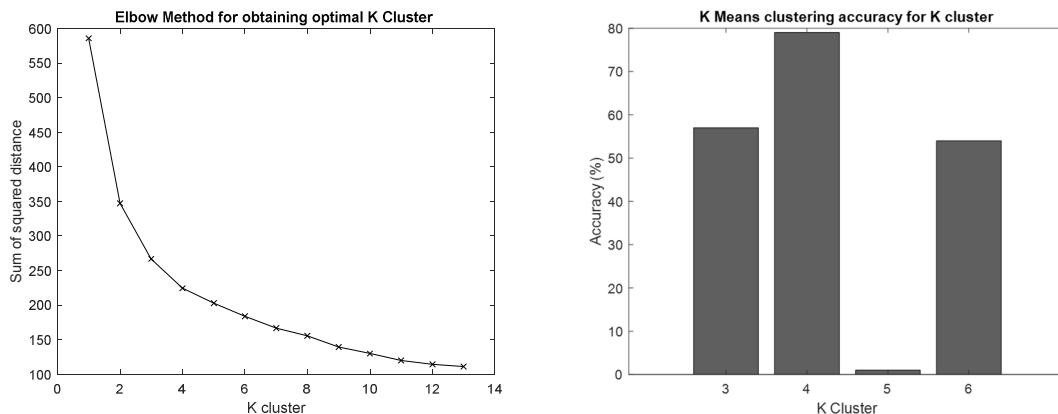

Figure 4. (a) Elbow graph for k-mean cluster. (b) Accuracy of k cluster.

Based on Fig. 4(b), the highest accuracy is observed when k=4 which is similar to the zoning division based on Fig. 1. The accuracy for k=4 is 79% which fulfils the antenna objective of zoning into 4 divisions.

The model when k = 4 is further examined and tabulated in Table 1. Here, the main features; which are the antenna reading, from the dataset that was taken to group the data can be observed in the centroids. As demonstrated in Equation (1), a higher value for centroid means that the RSSI reading for the corresponding antenna in the corresponding zone is expected to be higher.

Table 1. Tabulated centroid for Zoning division in 12 antennas.

Z	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12
1	0.0573	0.5587	0.0749	0.0148	0.0109	0.0118	0.0109	0.0109	0.0305	0.0336	0.0263	0.0109
2	0.0372	0.4990	0.1802	0.0151	0.0111	0.0129	0.0109	0.0109	0.0550	0.0451	0.4532	0.0109
3	0.0184	0.0129	0.0396	0.0139	0.0469	0.1086	0.0800	0.1294	0.0261	0.0468	0.0450	0.0159
4	0.0109	0.0109	0.0109	0.0109	0.0109	0.0109	0.0120	0.0262	0.0109	0.0109	0.0109	0.4059

The centroid of the clustering when k = =4 is summarized in Table 1. The unsupervised zoning grouping can be observed by finding the higher value centroids for each of the zones. This means that transmitters are considered to be in the corresponding zone if the RSSI readings of the corresponding antenna are higher. This way, the unsupervised zoning localization can be compared with the testbed zoning plan to determine whether the antenna placement is suitable for zone based localization.

Based on the centroids in Table 1, the zonal grouping is summarized in Table 2.

Table 2. Zones Based Antenna Grouping

Zones	No. of Antennas	Planned Zone Grouping (Antenna)	Unsupervised Zone Grouping (Antenna)
1	4	1,4,9,10	1,2,3,10
2	3	2,3,11	2,3,11
3	4	5,6,7,8	5,6,7,8
4	1	12	12

As shown in Table 2, the unsupervised k-mean zone grouping agrees with the planned zone grouping for zones 2, 3 and 4. However, there are some differences while the grouping of antenna for zone 1. This is because both zone 1 and zone 2 are near to each other and the antennas in both zones overlap with each other hence causing different features being taken into consideration. Furthermore, it was noted that there was a large metal object almost entirely occupying Zone 1; which may have affected the measurements made in the area.

The findings in this work demonstrated that zone based localization has the potential to be used reliably as localization technique in obstacle laden indoor environments. Furthermore, it was observed that fluctuations in the recorded data may render trilateration methods unreliable in real world environments without extensive computation strategy.

5. Conclusions

The work presented in this document has managed to demonstrate the feasibility of zone based localization in obstacle present environment. Furthermore, the dataset collected has demonstrated the difficulties of deploying trilateration and fingerprinting methods in dynamic conditions, as it introduces fluctuating RSSI measurements. The k-mean clustering was proposed to validate antenna placement to divide the testbed into zones, and to create a baseline accuracy which can be compared with other algorithms in the future. It was found that, for Zones 2, 3 and 4, the k-mean clustering of antennas agree with the planned antenna placement grouping. However, for Zone 1, there are differences, which may be attributed to a large metal obstacle occupying the zone. The k-mean clustering also recorded a peak accuracy of 79% with k=4, which agrees with the number of planned zones.

References

[1]　Xiao H, Zhang H, Wang Z and Gulliver T A 2017 An RSSI based DV-hop algorithm for wireless sensor networks *2017 IEEE Pacific Rim Conf. Commun. Comput. Signal Process. PACRIM 2017 - Proc.* **2017-Janua** 1–6

[2]　Shukri S and Kamarudin L M 2017 Device free localization technology for human detection and counting with RF sensor networks: A review *J. Netw. Comput. Appl.* **97** 157–74

[3]   Azmi N, Kamarudin L M, Mahmuddin M, Zakaria A, Shakaff A Y M, Khatun S, Kamarudin K and Morshed M N 2011 Interference issues and mitigation method in WSN 2.4GHz ISM band: A survey *2014 2nd Int. Conf. Electron. Des. ICED 2014* 403–8

[4]   Bai S and Wu T 2013 Analysis of K-Means algorithm on fingerprint based indoor localization system *2013 5th IEEE Int. Symp. Microwave, Antenna, Propag. EMC Technol. Wirel. Commun. MAPE 2013* 44–8

[5]   Cheng L, Wu X and Wang Y 2017 A non-line of sight localization method based on k-means clustering algorithm *Proc. 2017 IEEE 7th Int. Conf. Electron. Inf. Emerg. Commun. ICEIEC 2017* 465–8

[6]   Subaashini K, Dhivya G and Pitchiah R 2012 Zigbee RF signal strength for indoor location sensing-Experiments and results … *Technol. (ICACT), 2012 …* **1** 12–7

[7]   Shukri S, Kamarudin L M, Cheik G C, Gunasagaran R, Zakaria A, Kamarudin K, Zakaria S M M S, Harun A and Azemi S N 2017 Analysis of RSSI-based DFL for human detection in indoor environment using IRIS mote *2016 3rd Int. Conf. Electron. Des. ICED 2016* 216–21

[8]   Razavi A, Valkama M and Lohan E S 2015 K-means fingerprint clustering for low-complexity floor estimation in indoor mobile localization *2015 IEEE Globecom Work. GC Wkshps 2015 - Proc.* 1–7

[9]   Turner J S C, Kamarudin L M, Ndzi D L, Harun A, Zakaria A, Shakaff A Y M, Saad A R M and Mamduh S M 2011 Modelling indoor propagation for WSN deployment in smart building *2014 2nd Int. Conf. Electron. Des. ICED 2014* 398–402

[10]   Kamarudin L M, Ahmad R B, Ndzi D, Zakaria A, Ong B L, Kamarudin K, Harun A and Mamduh S M 2012 Modeling and simulation of WSNs for agriculture applications using dynamic transmit power control algorithm *Proc. - 3rd Int. Conf. Intell. Syst. Model. Simulation, ISMS 2012* 616–21

[11]   Mamduh S M, Shakaff A Y M, Saad S M, Kamarudin K, Kamarudin L M, Zakaria A, Kamarudin H, Ezanuddin A M M, Sa'Ad F S A, Nooriman W M and Abdullah A H 2012 Odour and hazardous gas monitoring system for swiftlet farming using wireless sensor network (WSN) *Chem. Eng. Trans.* **30** 331–6