International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 6, Issue 9, September 2016)

Indoor Wi-Fi Positioning System Based On K-means Cluster Analysis

Ali H. Saeed[1], Dia M. Ali[2]
[1,2]*University of Mosul & Mosul/ Iraq*

***Abstract*—Indoor Positioning Systems (IPS) based on the scene analysis algorithm are widely adapted because of the diverse set of features possessed by this algorithm which makes it attractive to any system designer, however scene analysis algorithm is not without any disadvantages, this algorithm suffer from large databases and high dimensionality which causes high computational complexity, to reduce the impact of this disadvantage K-means cluster analysis is used, where 70% of the size of the recorded database from the testbed was reduced and this significantly reduced the computational complexity which allowed us to use more sophisticated estimation procedures. The KNN algorithm was tested in three different ways to figure out an estimate of the position, the best results was achieved using (WKNN) with value of K=3.**

***Keywords*— Cluster analysis, Indoor Positioning Systems, K-means, K-Nearest Neighbour, Scene Analysis, WLAN.**

I. INTRODUCTION

The world is becoming more and more addictive to technology, and with each technological breakthrough has a set of applications that utilizes the full aspects of that technology. One of the applications that is getting a huge deal of interest is position detection and this application is particularly important because the environments (indoor and outdoor) around us are expanding and growing more complex every day.

The importance of any technology is always based on the applications derived from that technology, and how much is the influence of that application on the people's well-heeled life style. From this perspective position detection systems have gained the interest of researchers and developers, due its wide range of applications in the our life, where these systems are considered the corner stone of Location Based Services (LBS)[1], security , employee tracking and turn by turn navigation.

For the outdoor environments, there are two main Global Navigation Satellite System (GNSS) used for positioning and navigation, they are the Global Positioning System (GPS) lunched by the united States of America and the Global Náya Navigatsionnaya Sputnikovaya Sistema (GLONASS) lunched by the Russia (there is also GALILEO system sponsored by the European union, but this system is still under development)[2], these systems are satellite based and therefore suffer from performance degradation in Indoor environments (also in urban cities and near large buildings) due to the absence of the Line of Sight (LOS) between satellite and the device on ground, which causes multipath and weak signal reception issues.

The factors that affect the propagation of RF signals in indoor environments are quite diverse, but In general, there are three main factors that influence the location accuracy in the indoor environment which are relative humidity, people blocking, and opening or closing of doors [4]

There are basically two types of deployments that govern the design of any position detection system, client-based and infrastructure-based. In a client-based deployment, a tag (the device that needs to be located) receives the radio signal from multiple beacons (devices with known locations) to calculate the location by itself, while in an infrastructure based deployment, a group of beacons collect the signal measurements from an tags and send them to a central server for position estimation[5].

The design of any indoor positioning system must follow one of these approaches:

1- The first approach deals with enhancing the conventional positioning systems to overcome the challenges posed by indoor environments and estimate the location with a reasonable accuracy.
2- The second approach requires building an entirely new system specifically for detecting the location, these systems are often very reliable and can be accurate up to 15 cm. However, adding a new system with one single purpose to any environment may not be a straightforward process due to many constrains, such as the cost which is the decisive factor.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 6, Issue 9, September 2016)

3- The third approach is based on exploiting a pre-exist communication system (which wasn't intended for finding the location) and utilize the information provided by that system to estimate the position. The researches tend somehow to this approach because of the already deployed infrastructure of the system which reduces the cost, nevertheless, this approach still requires intelligent algorithms and extensive preprocessing to compensate for the low accuracy of the measured metrics[3]

The main goal of any localization system can be formulated as that of finding or estimating the position (in (2D) or (3D) space) of a point of interest within a coordinate system constructed using some known references. In any typical positioning system, the location of the point of interest is estimated by considering the displacement between that point and a Reference Point (RP) belongs to the system used for positioning. To accomplish this task a direction and/ distance estimations is required at the point of interest, the process requires taking measurements at the site and applying these measurements to algorithms to conclude the position. What makes positioning systems so helpful and efficient is that they are based on wireless networks, and with the rapid proliferation of this type of networks we can expect more enhanced positioning systems with higher accuracy and better performance [2].

This study presents a novel procedures to design a client-based deployment of a Wi-Fi based position detection system with reasonable accuracy, low computational cost (processing speed and storage capacity), low cost, and simplified planning procedures. The system utilizes the presence of a set of Access Points (APs) distributed in the study area (neither the number of APs nor their distribution were intended for position detection systems). (PS) principle proposed in this study is a client-based deployment based on the Received Signal Strength (RSS)-fingerprints technique for position estimation with adaptation of cluster analysis as a space reduction technique for the RSS-fingerprint dataset, , In this case the APs represent the beacons, while the tag is represented by a Wi-Fi adapter connected to a laptop computer.

The ground stone of the design in this paper is based on Received signal strength (RSS) scene analysis (Fingerprints), the reason behind choosing such a measuring principle is that Radio Propagation Models (RPM) is unreliable, because it is very difficult to model an indoor radio propagation with acceptable accuracy (in terms of positioning systems) due to sever multipath of the indoor environment which is caused by the environment layout, reflecting surfaces, and moving objects [11].

Fingerprinting consists of two phases, offline phase and online phase, the offline phase produce a dataset that contains coordinates of points in the study area and their respective RSS, however these datasets are often quite large and create a burden on the device, so Cluster Analysis is used as a space reduction tool for the dataset, the Clustering procedure can significantly reduce the computational complexity (processing time and storage capacity) and the margin of errors that associates large RSS-fingerprint dataset [7], in the online phase the KNN tool is tested in two ways, the first as a classification and estimation tool, the second only as a classification tool and for estimation of the position, the geometric principles of polygon centroid is used.

The main challenge to the techniques based on location fingerprinting is that the received signal strength could be affected by diffraction, reflection, and scattering in the propagation indoor environments [3].

II. Theoretical Background

This section represents a brief overview about the algorithms and procedures that will be attended in this paper.

A. Received Signal Strength (RSS) Scene Analysis (Fingerprints)

The RSS scene analysis is one of the most popular algorithms used in Indoor Positioning Systems (ISP) based on WiFi signals, the RSS scene analysis algorithm collects the RF features (fingerprints) of the scene, then the position of the object is estimated by a matching process between the online measurements and the closest *a priori* location fingerprints. RSS is commonly used with scene analysis algorithms as a location dependent characteristic of the signal to infer the position. Scene analysis consists generally of two stages: offline stage and online stage [3]. During the offline stage, the environment in which the positioning system will be installed in is surveyed at specific locations called Reference Points (RP). the environment survey is basically a procedure of recording the coordinates of the RP along with the corresponding characteristics of the signals from nearby APs/base stations effecting that location, the result of that procedure is called the Radio Map, which is a dataset that contains the RSS from the APs/base stations along with the coordinates of the RP from which these readings were collected, as indicated in eq (1) and eq (2), the radio map of the RSS part is denoted by Φ as eq. 1. Demonstrates:

$$\Phi = \begin{bmatrix} \phi_{1,1} & \phi_{1,2} & \cdots & \phi_{1,M} \\ \phi_{2,1} & \phi_{2,2} & \cdots & \phi_{2,M} \\ \vdots & \vdots & \cdots & \vdots \\ \phi_{N,1} & \phi_{N,2} & \cdots & \phi_{N,M} \end{bmatrix} (1)$$

where $\phi_{i,j}$ is considered the actual RSS from AP_i at RP_j, M is the number of available APs, and N is the number of RPs, therefore each row of Φ represents the RSS of one RP.

$$\phi_j = [\phi_{j,1}, \phi_{j,2}, \phi_{j,3}, \dots \phi_{j,M}], j = 1, 2, \dots . N \ (2)$$

The radio map can be denoted as $(P^j_{x,y}, \phi_j), j = 1, 2, \dots . N, \phi_j \in \mathbb{R}^M$ where the element $P^j_{x,y}$ represents the coordinates of the RP

In the online stage, location is estimated by matching the currently observed RSS from the Test point (TP) with pre_collectd data to figure out the position. The main challenge of using RSS scene analysis algorithm for positioning systems is that the signal strength vulnerable to the effects of diffraction, reflection, and scattering due to propagation in the indoor environments. There are several algorithms used for scene analysis based positioning to conclude the position in the online stage:

- Nearest neighbors techniques and variations thereof [17].
- Bayesian statistical matching [18].
- Maximum likelihood estimation [19].
- Correlation discriminate kernel selection [20].
- Neural networks [21].

The RSSs can be viewed as sensor data that refer to indoor locations; however, the characteristics of RSS of WLANs such as IEEE 802.11 itself have been studied properly. Although there is an extensive knowledge available regarding radio frequency (RF) phenomena and properties of the received signal in indoor environments (such as the distance-dependent property through path loss models and the fluctuation of signal because of the multipath effect), such knowledge is aimed toward communications capability and receiver design, making it limited for understanding positioning applications. A system designer needs to understand the underlying mechanism of RSS to efficiently model, design and analyze an indoor positioning system [8].

Up till now, there are three main categories of localization technologies, RSS scene analysis, radio Propagation Modeling (RPM), and Time or angle measurement based localization. the scene analysis based positioning systems are the most efficient way for WiFi systems because of the complexity of radio propagation there is no accurate and general model to fully describe the behavior of radio waves in indoor environments, and the precise requirement of synchronization needed for the Time and Angle measurements makes the system highly complicated and difficult to design with no better results than the other measuring metrics [14].

Recommended font sizes are shown in Table 1.

B. Cluster Analysis

Cluster analysis is basically an unsupervised pattern discovery procedure that refers to the grouping of records, observations, or cases into classes of similar objects [12]. A *cluster* is a collection of records that are similar to one another and dissimilar to records in other clusters. Clustering differs from classification in that there is no target variable for clustering. The clustering task does not try to classify, estimate, or predict the value of a target variable. Instead, clustering algorithms seek to segment the entire data set into relatively homogeneous subgroups or clusters, where the similarity of the records within the cluster is maximized, and the similarity to records outside this cluster is minimized [6]. Cluster analysis builds a typology using the information derived from the dataset.

Clustering is often used as a preliminary analysis tool in a data mining studies, with the resulting clusters being used as further inputs into different techniques for complete analysis. Due to the enormous size of many present-day databases, it is often helpful to use cluster analysis first, to reduce the search space for the algorithms. Cluster analysis encounters many issues that need to be addressed first before performing the clustering For example, how to measure similarity, how to standardize or normalize numerical variables, and how many clusters do we expect to uncover [13].

Cluster analysis is divided into two basic categories, hierarchical clustering, and non-hierarchical clustering. This paper deals with non- hierarchical cluster analysis and specifically Kmeans cluster analysis. K means is considered one of the most common tools to perform non-hierarchical cluster analysis.

C. K-Nearest Neighbor Algorithm (KNN)

K-NN is a location fingerprinting-based positioning algorithms that uses pattern recognition technique to identify the position of the user in the online stage of the RSS scene analysis algorithm, its most often application is for classification, although it can also be used for estimation and prediction which makes it useful for positioning systems. *k*-Nearest neighbor is an example of *instance-based learning*, in which the training data set is stored, so that when a reading with unknown variables are recorded, these variables may be found simply by comparing it to the most similar records in the training set, the similarity is measured by calculating the distance between the TP and each reading in the dataset, the formula of the distance is given by:

$$D_i = \left(\sum_{j=1}^{M} \left\| \phi_{test,j} - \phi_{i,j} \right\|^p\right)^{\frac{1}{p}}, i = 1, 2, \dots, N, j = 1, 2, \dots, M \quad (3)$$

Where $\phi_{test,j}$ is the RSS from AP_j recorded at the TP, D_i is the Manhattan distance and Euclidean distance when p=1 and 2, respectively. Then the K RP with the smallest values of D are selected to calculate the position of the TP.

KNN algorithms use one of two methods to estimate the unknown parameter; the first method is the locally weighted averaging [6].

$$\hat{P} = \frac{\sum_i^n W_i P_i}{\sum_i^n W_i} \quad (4)$$

Where W_i is the weight of the ith reference point, P_i is the position of the ith reference point, n is the number of nearest neighbors set by K, $\hat{P}$ is the estimated position, this method considers the distance between the TP and the K selected RPs and assigns higher weights to the RPs closer to the TP and lower weights to the RPs farther to the TP, this will give a higher impact for the closer RPs on the final result than the farther RPs.

Another method is based on simply calculating the average of the variables to conclude the unknown variable without considering how far the points are from the target variable, this method is less sophisticated than the locally weighted average method, but it is less accurate too.

$$\hat{P} = \frac{\sum_i^n P_i}{n} \quad (5)$$

However this paper proposes another method to determine the unknown variable, this method is based on simple geometry which uses the principle of Centroid of Polygon, as will be explained in the next paragraph.

D. Figures and Tables

Figures and tables must be centered in the column. Large figures and tables may span across both columns. Any table or figure that takes up more than 1 column width must be positioned either at the top or at the bottom of the page.

E. Figure Captions

The centroid (Center of gravity) is the geometric centre of area of a body, shape or section [9], centroid can be thought of as a single point at which the weight could be held and be in balance in all directions [10].

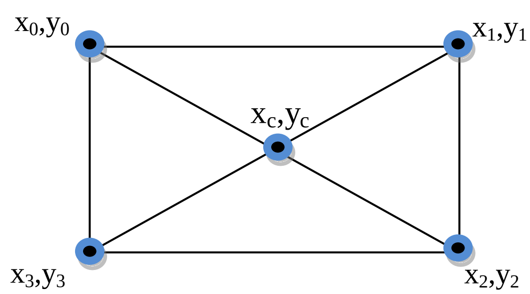


Fig. 1. Location of the Centroid

$$A = \frac{1}{2}\sum_{i=0}^{N-1}(x_i y_{i+1} - x_{i+1} y_i) \dots . eq \quad (6)$$

$$X_c = \frac{1}{6A}\sum_{i=0}^{N-1}(x_i + x_{i+1})(x_i y_{i+1} - x_{i+1} y_i) \dots . eq \quad (7)$$

$$Y_c = \frac{1}{6A}\sum_{i=0}^{N-1}(y_i + y_{i+1})(x_i y_{i+1} - x_{i+1} y_i) \dots . eq \quad (8)$$

Where A is the area oh the polygon, X_c is the X coordinate of the centroid, Y_c is the Y coordinate of the centroid

This paper suggests a procedure to use the idea of centroid of polygon with the K-NN tool to estimate the position, the procedure states that the coordinate output of the K-NN tool can form a polygon and the position of interest is at the centroid of that polygon.

IJETAE
EXPLORING RESEARCH AND INNOVATIONS

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 6, Issue 9, September 2016)

III. RESEARCH MYTHOLOGY

This section explains the steps that led to building the system, starts by describing the testbed of the research, and then the data collection process is described along with the devices used for data collection.

A. The Research Testbed

The test bed of the research was chosen in the first floor of the building of communication engineering within the campus of the university of Mosul, the dimensions of the building are (45 * 27) m as shown in Fig. 2, 4 Access Points (AP) were placed in the floor to form the signal sources that will be used as signal sources for the positioning system, as shown in Fig. 2, AP4 is directly above AP3. In order to obtain higher resolution for RSS, the transmitted power of AP4 was reduced from high to medium (naturally this will cause scalability issue).

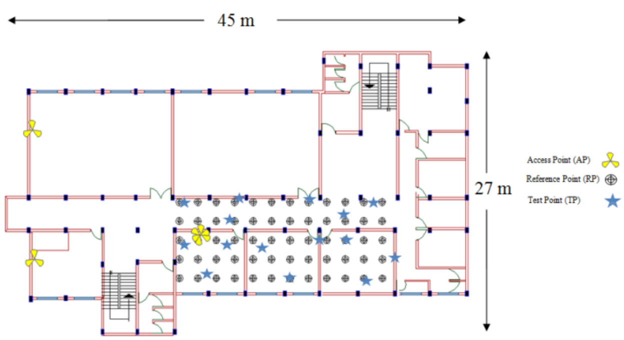


Fig. 2. Structure of the Testbed

B. Data Collection

To collect the data an (*hp*) laptop computer running Microsoft Windows 7 with (D-Link) external dual band Network Interface Card (NIC) connected with it. The reason of choosing the D-Link NIC is that this NIC has relatively stable RSS variations of long periods of time [15] as shown in Fig. 3 and Fig. 4, and since collecting the RSS fingerprint database was performed over several days then D-Link NIC was the reasonable choice over other NICs.

Although the D-link NIC that is used to build this system is different from the D-link NIC used by [15], however [15] also proved that NICs from the same vendor can give the same RSS fingerprints for the same environment, which means that the NIC used for this system has the same relative stability in RSS variations to the NIC tested by [15].

Two software were used to collect the data, the first is "NetSurveyor" which is program for recording the parameters of the real time signal of the Wi-Fi network (such as RSS, BSSID, channel...etc), this program has the ability to export the recoded data to an xml sheet which facilitates the processing step, the second program was a virtual GPS program that provided the coordinates of the RP. As shown in Fig. 2, (the part of the study area that wasn't covered by the scan was left for future developments on the system).

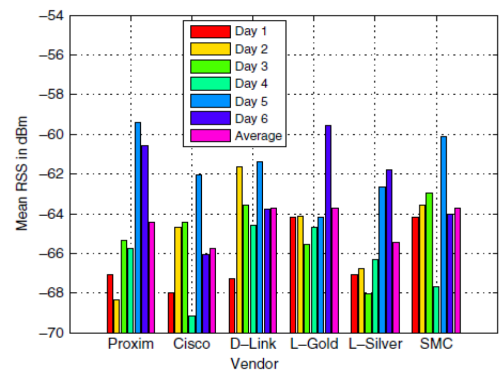


Fig. 3. Comparing mean RSS of different vendors. [15]

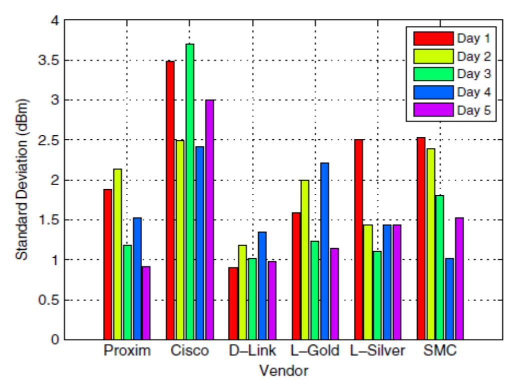


Fig. 4. Comparing standard deviation of different vendors. [15]

IV. BUILDING THE POSITIONING SYSTEM

This section discusses the procedures that led to the building of the system and then the actual steps of building the system, the first part of this section describes the data preparation, and the second part demonstrates the procedure of building the system.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 6, Issue 9, September 2016)

A. Data Preparation

The data required for building the system are the coordinates of the test location and its corresponding RSS in dBm (recorded from four different APs), a Dataset was created and contained the RSSs and the coordinates (each row contained 4 RSSs readings from the 4 different APs and X-Y coordinates), the data was collected in uniform fashion where a 12 readings (12 row) was collected from each physical location, the readings came from four groups of 3 readings each, where each group was recorded in a different direction (North, East, South, West) of the physical location to include all the variations in the RSS that occur during the propagation of the signal, and then added to them another row which contained the average of the 12 reading and the same coordinates. The (Microsoft Office Excel) program was used to contain the data set in XLSX format which facilitates extracting the data for further processing.

A 3-D interpolation based on a Delaunay triangulation[i] was used on the scattered data set to define the signal strength for the areas between the RP. As shown in figures (5, 6, 7, and 8) respectively, where each figure represent the RSS distribution pattern of each AP in the testbed.

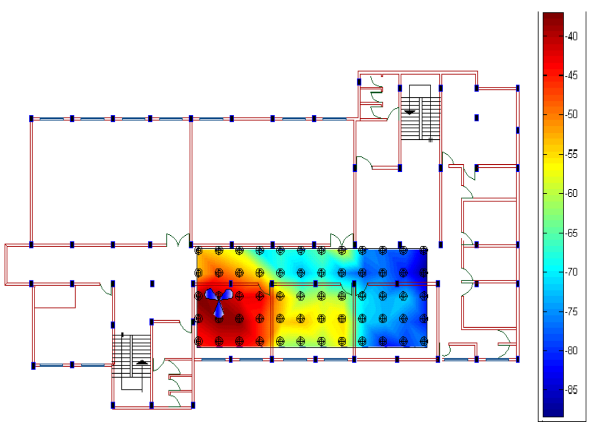


Fig. 5. RSS Distribution of AP1

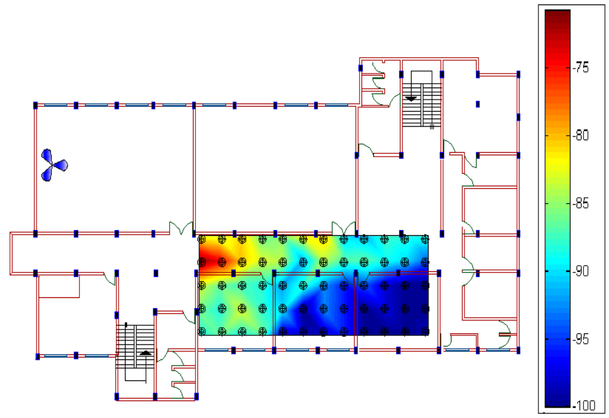


Fig. 6. RSS Distribution of AP2

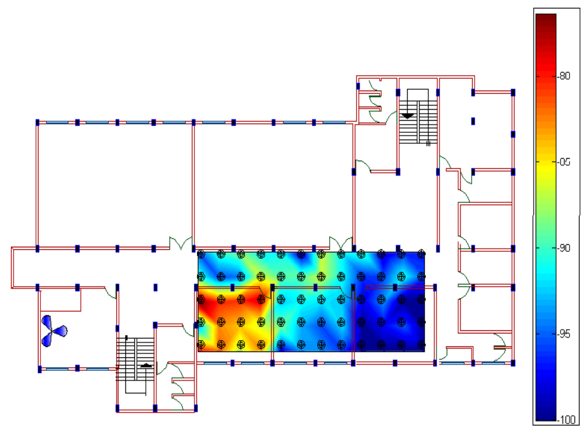


Fig. 7. RSS Distribution of AP3

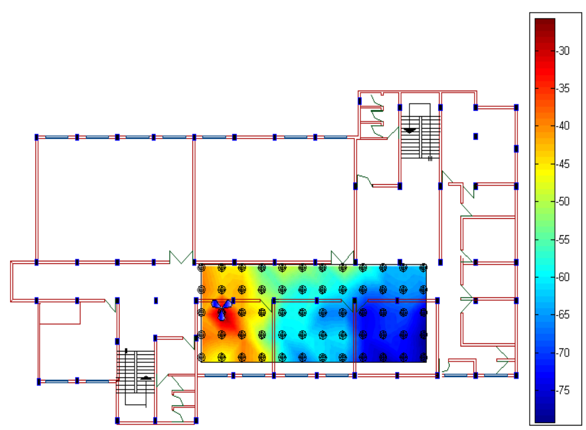


Fig. 8. RSS Distribution of AP4

B. Constructing the System

The complete construction of the system must go throw two phases, the Initial Phase and the Deterministic Phase. The initial phase deals with is mainly reserved for the clustering process (specifically K-means clustering), where the entire dataset is subjected to the K-means function, the K-means function perform an iteration procedure after the random initial seeds are chosen to conclude the most appropriate centroids, in this study the number of iterations is set to 300 to ensure that the K-mean function reaches the correct centroids, furthermore, the whole clustering procedure is repeated several times because after all the initial seeds are chosen randomly (which will result eventually different centroids almost every time the clustering is repeated) and the right configuration of centroids is chosen for the next phase. The proposed system was tested with three different numbers of clusters, 10, 50, and 90. While the deterministic phase deals with calculating the position based on the readings from the Test Point(TP) through the K-NN tool, K was tested over 4 values (2,3,4 and 5), and the position have been calculated in three different methods:

1) The first one consists of simply averaging the output of KNN to conclude the position through eq (5).
2) The second one used the estimation principle of WKNN through eq (4).
3) The third one is more elaborate which used the principles of geometry (specifically, the centroid of polygon) to calculate the position through eqs (7 & 8).

V. BUILDING THE POSITIONING SYSTEM

This section discuses the proposed algorithm and provides the performance analysis for each procedure to estimate the position based on the collected data. Four different values of K will be evaluated (2, 3, 4, 5) for each procedure, and since the clustering process was repeated several times only the best results for each value of K with number of clusters= 90 cluster will be presented, along with a comparison between the performance of the system when the number of clusters is 10, 50, and 90. The results of each procedure will be displayed in terms of Cumulative Distributions Function (CDF), Probability Density Function (PDF) and the Scalability measure.

A. The K-Nearest Neighbor (KNN) Procedure

The procedure of using the KNN tool for estimating the unknown variables is quite uncomplicated, it simply states that in order to find the missing variable from the K nearest neighbors, the system only needs to calculate the average of the neighbors of the reading of that variable, as shown in eq (5).

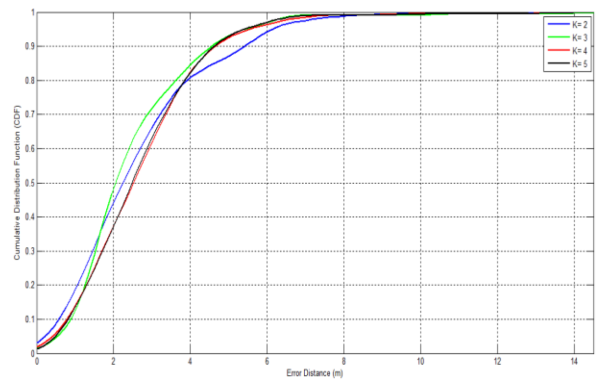


Fig. 9. CDF of KNN Estimation Procedure

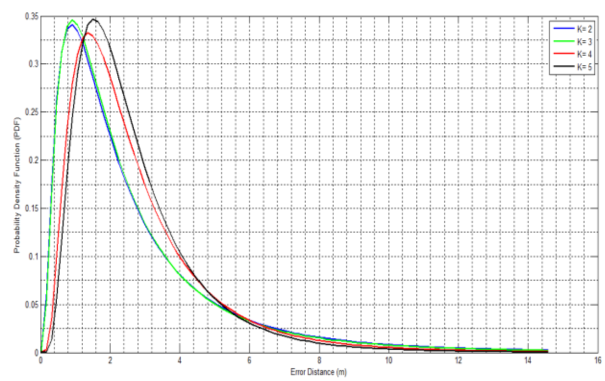
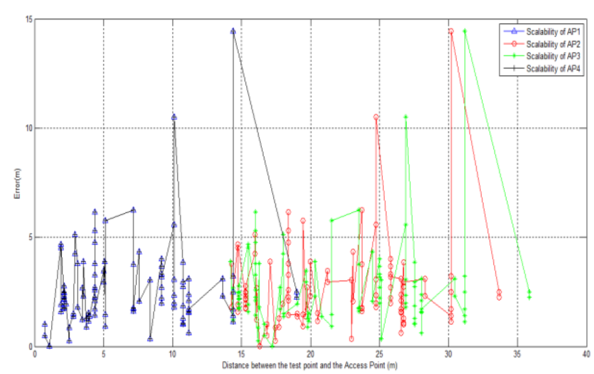


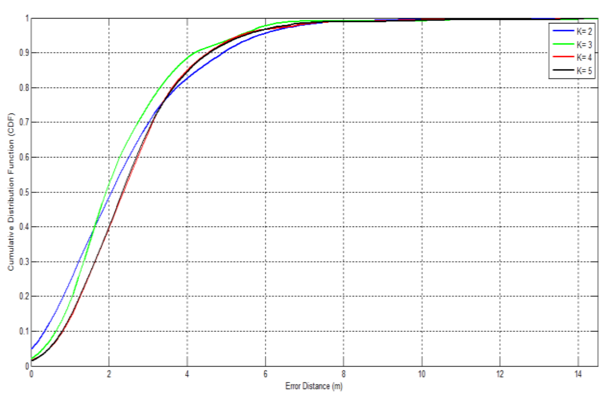
Fig. 10. PDF of KNN Estimation Procedure

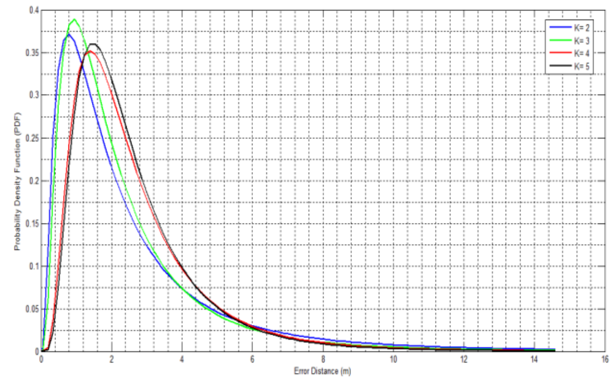


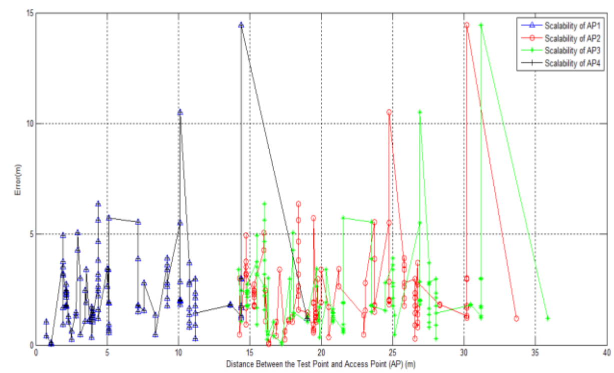
As the results indicate, the best performance came from K=3 in terms of the Cumulative Distribution Functions (CDF) and Probability Distribution Function (PDF), as shown in Fig. 11, and Fig. 12, respectively. Fig. 17 demonstrates the Scalability effect over the system, as shown in the Fig. 17, the systems performance has a reasonable immunity regarding the distance between the TP and the Access Point (AP), however, in extreme cases where the RSS reaches the its minimum levels, we can see that the performance of the system suffers from relatively high error measurements.

B. The Weighted K-Nearest Neighbor (WKNN) Procedure

This procedure is similar in principle to the concept of conventional KNN estimation procedure, where WKNN also performs averaging calculations to estimate the unknown variable, however the WKNN procedure has one major difference from the KNN, the WKNN takes into account the distance between the reading and the neighbors, where every neighbor is assigned a weight to it which corresponds to its distance from the reading, and the further the neighbor the smaller the weight assigned to it, which means that neighbor will have an effect proportional to its weight (distance from the TP) on the calculations of the unknown variable as eq (4) describes.



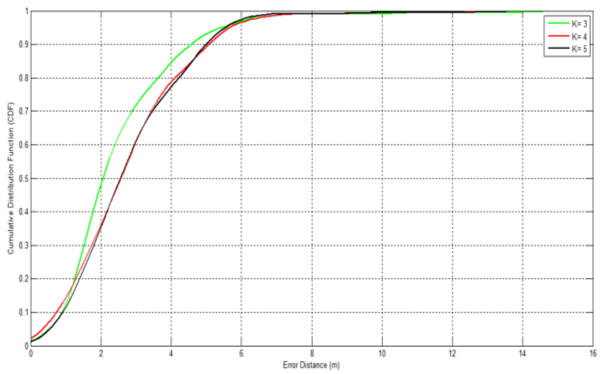


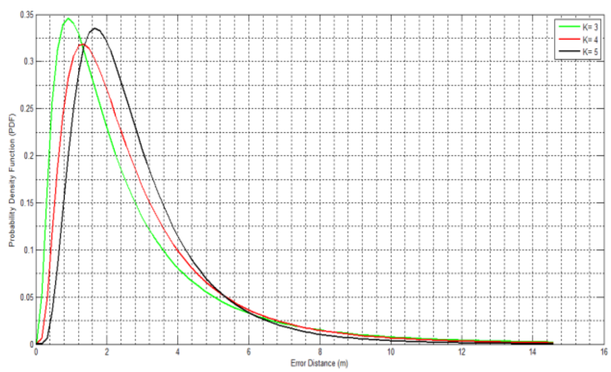


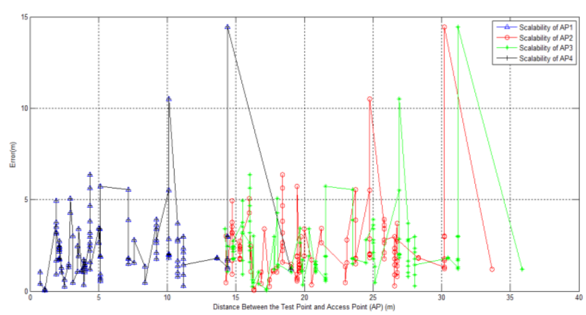
C. The Centroid of Polygon (COP)

The Centroid of Polygon procedure is rather different from the typical KNN estimation procedure, despite the fact that it uses the KNN tool, but only as a classification tool, and for the estimation, this procedure suggests to use a principle steamed from pure geometry to calculate the position as stated in equations (7 and 8), this principle is the Centroid of Polygon, COP is mainly used with positioning algorithms based on proximity [2]. COP was tested with KNN over 3 values of K only (3, 4, and 5), the situation of K= 2 was excluded because the coordinate output of the KNN tool will form a straight line and the centroid will be always in the middle of that line which means simple averaging of that coordinates will produce the centroid, and this is what the regular KNN procedure is doing already.

IJETAE
EXPLORING RESEARCH AND INNOVATIONS
International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 6, Issue 9, September 2016)







D. Comparison between the Three Procedures

In order to provide more obvious vision of the performance of each procedure a comparison between these procedure is in order, five different performance metrics of differentiation were chosen, Accuracy which was calculated as the mean distance error, Precision which was measured in terms of CDF and PDF (the CDF Term was measured at 4m while the PDF term was measured at 0.9m), scalability measure, and complexity.

TABLE I
COMPARING THE PERFORMANCE OF THE THREE PROCEDURES

	KNN	WKNN	COP
Accuracy	2.5 m	2.2 m	2.5 m
Precision (CDF/PDF)	85% / 0.34	89% / 0.39	85% / 0.34
Scalability	Good	Very Good	Good
Complexity	Low	Moderate	High

VI. CONCLUSIONS

The obtained results showed several indications, the most important ones were:

- The best results was achieved using the WKNN estimation procedure with K=3, and the number of clusters was set to 90 clusters, the precision of the system performance reached 89% for 4m error distance based on the CDF plot, and 0.39 PDF for 0.9m error distance.
- Among all the values of K that was tested for the KNN tool, K= 3 achieved the optimum performance for the system in terms of the CDF and PDF.
- The COP estimation procedure provided reasonable performance accuracy due to the semi random distribution of the clusters centroids, however, it is believed that this procedure can perform better and obtain more reliable results with more uniform distribution of the Reference Points (RP).
- Using cluster analysis played an important role in reducing the impact of computational complexity which reduced the costs of CPU processing and the memory size required to store the database,
- The precision level of clustering is governed by the number of clusters assigned in the design phase, where increasing the number of clusters has indeed increased the level of precision, nevertheless increasing the number of clusters beyond a certain threshold (the threshold is determined by the size of the dataset) could lead to backward results and decrease the system performance, the reason of this situation is that some elements of the dataset don't end up in the right clusters.

REFERENCES

The heading of the References section must not be numbered. All reference items must be in 8 pt font. Please use Regular and Italic styles to distinguish different fields as shown in the References section. Number the reference items consecutively in square brackets (e.g. [1]).

[1] Jochen Schiller, Agnès Voisard, Location-Based Services, Publisher: Elsevier, publishing date:2004, ISBN: 1-55860-929-6.

[2] David M. Rodríguez, Frantz Bouchereau, César V. Rosales, Rogerio E. Caldera, position location techniques and applications, Publisher: Elsevier Inc. , publishing date: 2009, ISBN: 13-978-0-12-374353-4.

[3] Hui Liu, Houshang Darabi, Pat Banerjee, Jing Liu , Survey of Wireless Indoor Positioning Techniques and Systems, IEEE Transactions On Systems, Man, And Cybernetics Part C: Applications And Reviews, Vol. 37, NO. 6, November 2007.

[4] Y.-C. Chen, J.-R. Chiang, H.-H. Chu, P. Huang, A.W. Tsui, Sensor-assisted Wi-Fi indoor location system for adapting to environmental dynamics, in: Proceedings of the ACM/IEEE MSWiM 2005, October 2005.

[5] Yih-Shyh Chiou , Chin-Liang Wang, Sheng-Cheng Yeh , Ming-Yang Su, Design of an adaptive positioning system based on WiFi radio signals, Published by Elsevier, Computer Communications 32 (2009) 1245–1254

[6] Daniel T. Larose, DISCOVERING KNOWLEDGE IN DATA, An Introduction to Data Mining, Publisher: John Wiley & Sons, Inc, , publishing date: 2005, ISBN 0-471-66657-2 (cloth).

[7] Liu, X.; Zhang, S.; Zhao, Q.; Lin, X., A Real-Time Algorithm for Fingerprint Localization Based on Clustering and Spatial Diversity. In Proceedings of International Congress on Ultra-Modern Telecommunications and Control Systems and Workshops, Moscow, Russia, 18–20 October 2010; pp. 74–81.

[8] Kamol Kaemarungsi, and Prashant Krishnamurthy, Analysis of WLAN's received signal strength indication for indoor location fingerprinting, Elsevier journal of Pervasive and Mobile Computing 8 (2012) 292–316, available on line on 24 September 2011.

[9] Philip Garrison, Basic Structures for Engineers and Architects, published by: Blackwell Publishing Ltd, publishing date: 2005 , ISBN 1-4051-2053-3.

[10] Barry Onouye, Kevin Kane, Statics and Strength of Materials for Architecture and Building Construction, published by: Pearson Education, Inc, publishing date:2012, ISBN 978-0-13-507925-6.

[11] K. Pahlavan, X. Li, and J. Makela, Indoor geolocation science and technology, IEEE Commun. Mag., vol. 40, no. 2, pp. 112–118, Feb. 2002.

[12] Krzysztof J. Cios, Witold Pedrycz, Roman W. Swiniarski, Lukasz A. Kurgan, Data Mining, A Knowledge Discovery Approach, publisher: Springer Science +Business Media, LLC, publishing date:2007, ISBN-13: 978-0-387-33333-5.

[13] Jiawei Han, Micheline Kamber, Jian Pei, Data Mining: Concepts and Techniques, published by: Morgan Kaufmann, an imprint of Elsevier, publishing date: 2012, ISBN: 978-0-12-381479-1.

[14] Mu Zhou, Zengshan Tian, Kunjie Xu, Xiang Yu, Haibo Wu, Theoretical entropy assessment of fingerprint-based Wi-Fi localization accuracy, Elsevier journal , Expert Systems with Applications, 40 (2013) 6136–6149.

[15] Kamol Kaemarungsi, Prashant Krishnamurthy, Analysis of WLAN's received signal strength indication for indoor location fingerprinting, Elsevier journal, Pervasive and Mobile Computing 8 (2012) 292–316.

[16] Mikhail J. Atallah, Marina Blanton , Algorithms and Theory of Computation Handbook, Second Edition, Volume 2: Special Topics and Techniques, CRC Press, 20/11/2009, ISBN: 1584888210, 9781584888215.

[17] F. Evennou and F. Marx. Advanced integration of wifi and inertial navigation systemsfor indoor mobile positioning. EURASIP Journal on Applied Signal Processing, v 2006,n 17, 2006, 11 pp., 2006.

[18] Lionel Reyero and Gilles Delisle. A pervasive indoor-outdoor positioning system. Journal of Networks, 3(8):70 - 83, 2008. GPS; GSM; Positioning; Ubiquitous; WLAN.

[19] Chunwang Gao, Zhen Yu, Yawen Wei, Steve Russell, and Yong Guan. A statistical indoor localization method for supporting location-based access control. Mobile Networks and Applications, 14(2):253 - 263, 2009.

[20] Azadeh Kushki, Konstantinos N. Plataniotis, and Anastasios N. Venetsanopoulos, Kernel-based positioning in wireless local area networks. IEEE Transactions on Mobile Computing, v 6, n 6, June, 2007, p 689-705, 2007.

[21] Shih-Hau Fang and Tsung-Nan Lin. Indoor location system based on discriminant-adaptive neural network in ieee 802.11 environments. IEEE Transactions on Neural Networks,19(11):1973 - 1978, 2008.

[i] The Delaunay triangulation of a data set is a triangulation technique such that the unique circle that surrounds each triangle contains no other points in the set. The convex hull of a data set of points is the smallest convex set which contains all the points of the original set[16].