PAPER • OPEN ACCESS

Wi-Fi based indoor localization using trilateration and fingerprinting methods

To cite this article: Abd Elgwad M. El Ashry and Bassem I. Sheta 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **610** 012072

View the article online for updates and enhancements.

Recent citations

- Comprehensive Analysis on Least-Squares Lateration for Indoor Positioning Systems
Korhan Cengiz

Wi-Fi based indoor localization using trilateration and fingerprinting methods

Abd Elgwad M. El Ashry[1], Bassem I. Sheta

Military Technical College, Cairo, Egypt.

[1] Email: A.elashry@mtc.edu.eg

Abstract, Nowadays, mobile devices such as personal computers and smartphones are emerging as a major key in today's computing platforms for indoor object localization systems due to the object localization in indoor areas strongly suffers from limitation of using GNSS (Global navigation satellite system) systems due to low satellite availability and high signal attenuation. During the last decade, many researchers have developed indoor localization systems which are the process of obtaining user or device location through mobile devices using Wireless Fidelity (Wi-Fi) network signals with promising results and acceptable performance. In these Wi-Fi based localization systems, indoor positioning relies on different types of measurements including Time-Of-Arrival (TOA), Time-Difference-Of-Arrival (TDOA), Angle-Of-Arrival (AOA), and Received Signal Strength (RSS) of Wi-Fi signal.

In this paper, the techniques and algorithms that used for the RSS-based localization such as Trilateration and Fingerprinting which depend on the RSS from the access point (WI-FI). Using Received Signal Strength Ranging approach in the Trilateration method which depends on database that contains path-loss-exponent and shadowing parameter that differ according to the environment, solving the equation using different Access Points (APs) at least 3 APs, the no of the APs and there locations were varied to get the best accuracy which depends on the horizontal dilution of position (HDOP). At the Fingerprinting method depends on matching the recorded offline RSS from nearby access points (AP) to the online RSS received by the user on the move is reviewed. A comparison of location fingerprinting methods involving deterministic method (k-nearest neighbor method and weighted k-nearest neighbor method), probabilistic methods by estimation of likelihood functions with several approaches (non-parametric and parametric)are also explained. The performance parameters of this study include the two-dimensional root mean square error (2D-rms) which measures the localization accuracy. Moreover, the effect of increasing/decreasing the number of APs on the system accuracy is also discussed. The aim of this paper is to announce which method can provide better performance than the others and under what conditions.

keywords: Wi-Fi signal, Indoor navigation, Trilateration, Dilution of position, Location fingerprinting.

1. Introduction

One of the major aspects that should be taken into consideration when investigating indoor localization is the practical aspects of the method used; such as system deployment, cost and time for system training. That is why Wi-Fi is one of the best effective approaches used for indoor-Positioning since most of the buildings as malls or large office have already Wi-Fi devices that provides whole building coverage as network AP, all smart personal devices support Wi-Fi and not need to any new infrastructure or hardware changing . With all the advantages, Wi-Fi is a best technology in literature for indoor-localization [1]. Indoor environments are actually challenging, for many reasons:

Severe multipath, High attenuation, signal scattering and fast temporal changes. While, there are indoor settings facilitate Localization: Small coverage areas, Low weather influences, fixed geometric and Lower dynamics due to slower.

1.1. Wireless positioning techniques
Even though there are many wireless technologies are used for indoor-localization, the techniques that the Positioning is depend on are limited. It can be categorized into three techniques: Trilateration, fingerprint and triangulation.

1.1.1. Trilateration: Lateration measures the location of the user by measuring distances from many reference points [2]. It can be also called range-measurement techniques, the distance can be measured using received-signal-strengths (RSS) depending on the inverse relationship between the distance and the signal strength. Time difference of arrival (TDOA) and Time of arrival (TOA) that depend on the propagation time.

a.   RSS-Based Method:

The RSS-Based Trilateration technique uses parameters of known APs as signal strength and real coordinates of APs inside the buildings. The signal is received by the mobile-device that measures the distance between the Access-Points and the device [4]. A specific database must be created for each environment such as (path loss exponent and shadowing parameter).

b.   TOA (Time of Arrival)

The distance is directly proportional to the propagation-time. TOA measurements should be taken the signals from three APs at least [2]. For TOA-based techniques, the one-way propagation time is measured, and the distance between the user and APs is measured, then the intersection between the circles is the user's position. In general, direct TOA has the main problem that all the devices such as receivers and transmitters have to be synchronized in the same system.

c.   TDOA (Time difference of Arrival)

The TODA depend on the difference of the arriving time to the target (user) from two of transmitters [4]. In a unilateral system (self -positioning), the user is a receiver, the system using the same frequency, all the transmitters transmit the signal frequency that should be at least three transmitters to achieve two dimensions, the transmitters transmit at different times to avoid the interference. The clock of the transmitters must be synchronized, so the transmitters send periodic beacons at different times to guarantee no overlap of the transmissions. The receiver should know the strategy of sending the transmission pulses and compute the difference time between each other, then note the time arrival difference between the pulses that received from the transmitters so the receiver shouldn't be synchronized with the transmitters.

1.1.2. Triangulation: The basic idea is how to find the position of the target using the angles between the target and at least two ground station. The angle of arrivals (AOA)[2][3]It is one component of the radar system that use to find the direction of the target using two stations to find the 2D dimension, it doesn't need synchronization between the target and the base stations and any cooperation from the target, but more complex hardware requirement depending on the angle that measured from the base stations with knowing the stations positions. So the target position can be calculated by the intersection between the angle direction lines**.**

1.1.3. Fingerprinting Method: It consist of two phases [4] Offline phase: Construction data base called Radio-Map that contain the feature vector of each position in the interest area, (RSS, Location).
Online-phase**:** the device uses the samples of the signal strength that received from the APs to compare with radio map (Data base) in order to calculate the user's location.

2. Location Trilateration (RSS Ranging)

The main idea of the Trilateration is how to calculate the distance between the user and the Access-Point. There are many ways to measure the distance such as TOA, TDOA, and RSS. In this section, the RSS Ranging is used instead of TOA or TDOA because it doesn't need synchronization between transmitters, or between transmitters and receivers, using the attenuation of the signal strength to measure the distance between the user and the Access-Point. Using the propagation models, a database that contains the path loss exponent, shadowing parameter and solving all the equation using linear and iterative approaches with varying the no of APs starting from three APs at least to achieve the best performance with a suitable accuracy in the sense of HDOP (Horizontal Dilution of precision).

The RSS in these points decreases exponentially based on the distance between the user and Access-Point. Thus, this dependency can be considered as a function of distance. The line of position of constant distance is a circle around the AP [3]. The intersection of 3 Access-Points radiuses provides a point or an area of the receiver as shown in Fig. 1.
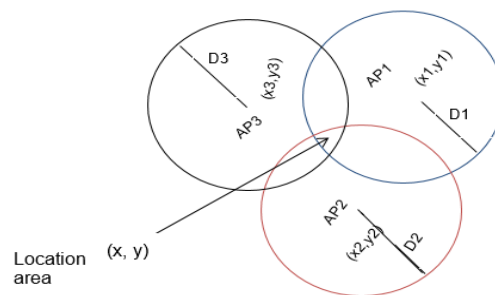

Fig. 1 The user's position of the intersection of three Circles

$$d_1{}^2 = (x_1 - x)^2 + (y_1 - y)^2 \tag{1}$$
$$d_2{}^2 = (x_2 - x)^2 + (y_2 - y)^2. \tag{2}$$
$$...$$
$$d_n{}^2 = (x_n - x)^2 + (y_n - y)^2 \tag{3}$$

where x_1 , x_2 , x_n, y_1, y_2, y_n are the coordinates of APs, d_1, d_2, d_n are the distances. The solution of these equations are points of circles intersection providing an the user position.

The Wi-Fi trilateration approach basically consists of two steps: the first one is calculating the distance by signal-propagation model for indoor WLAN signals and the second step, computing location by solving the nonlinear equations to get the user's position. So the distance between the mobile unit should be measured which is inversely proportional with the received power using the propagation models.

2.1. Propagation models

2.1.1. Simplified Path-Loss Model: The simplified model is commonly used for system design, in feasible measurement, there is a direct relationship between the path-loss and distance. It may

be different when it is in different environments. The average large-scale path loss for an arbitrary T-R separation is expressed as a function of distance by using a path loss exponent, n [5][6]:

$$\overline{p_L(d)} \propto (\frac{d}{d_0})^n \qquad (4)$$

$$\overline{p_L(d)} = \overline{p_L(d_0)} + 10nlog(\frac{d}{d_0}) \qquad (5)$$

where n is path-loss-exponent which can change at different environments, d_o is a reference point in close to transmitting antenna.

The value of n relies on the exact propagation environment, in free-space, n =2, and when obstacles are present, n will take a higher value.[5].

2.1.2. Log-normal Shadowing: Shadowing as discussed before produced by obstacles from the Tx to Rx that   reduce signal power due to reflection-absorption-diffraction and scattering. The path loss $p_L\ d$ at a particular location is random and distributed log-normally (normal in dB) about the mean distance dependent value [5]. That is

$$p_L\ d(dB) = \overline{p_L(d)} + \chi_\sigma \qquad (7)$$

$$p_L\ d(dB) = \overline{p_L(d_0)} + 10nlog\left(\frac{d}{d_0}\right) + \chi_\sigma \qquad (8)$$

where χ_σ is shadowing parameter that can be represented by the standard deviation σ [5]

Note that:

$$p_r\ d(dBm) = p_t\ d(dBm) - p_L\ d(dB) \qquad (9)$$

The close-in reference distance d_0, the path-loss-exponent n, and the standard deviation σ,

2.1.3. Calculation of path-loss-exponent (n) and the standard deviation (σ): As discussed before, Pass Loss-exponent (n) varies at different environment, so it should be measured at each environment that has different construction from the others , to characterize the area for RSS ranging, received power $P_r(d_0)$ is 1st measured by allocating a receiver d_0 apart from the Tx. d_0 is generally located at 1meter [7]. After$P_r\ (d_0)$ is calculated, the Rx is moved to another locations to calculate Pass Loss -exponent n [6] :

$$n = \frac{P_r(d_0) - P_r(d)}{10log(\frac{d}{d_0})} \qquad (10)$$

where P_r (d) is the received power measured at a distance d to the transmitter ,which is expressed in dBm

The value of n is also calculated from the measured data, by linear regression such that the difference between the estimated and measured Pass Loss is minimized in a MSE sense over a high range of measurement locations and T-R partings.

The (MSE) is given with the next expressing:

$$MSE = \frac{1}{k}\sum_1^k (P_r(d_i) - \widehat{P_r(d_i)})^2 \qquad (11)$$

where :

$P_r(d_i)$ is the actual value of received power at some distance

$\widehat{P_r(d_i)})$ is the estimated value of received power at the same distance

k is the number of measurement samples

The value of n that will minimize this MSE (n) value (MMSE) minimum mean square error is achieved by:

- Write MSE as a function of n.

- Find the value of n that minimizes this function and it is achieved by the derivation of MSE(n) with respect to n and solving for n that the derivative equal to zero [5].

The Standard Deviation σ is the square-root-of-the-variance that given by:

$$\sigma^2 = \frac{\sum_1^k (P_r(d_i) - \overline{P_r(d_i)})}{k}$$

$$\sigma = \sqrt{\frac{\sum_1^k (P_r(d_i) - \overline{P_r(d_i)})}{k}} \tag{12}$$

2.1.4. Measuring the user position (Iterative approaches): To get the best solution to the user coordinates (x, y), this method is used as a solution in GPS to get a user position [8] but instead of using a Peseudoranges and the errors between it and the measured value, the distance that estimated form the relation with the received power as illustrated in equation (8):

$$d_1 = \sqrt{(x_1 - x)^2 + (y_1 - y)^2}$$
$$d_2 = \sqrt{(x_2 - x)^2 + (y_2 - y)^2}$$
$$d_3 = \sqrt{(x_3 - x)^2 + (y_3 - y)^2}$$

The equations in a general form:

$$d_i = \sqrt{(x_i - x)^2 + (y_i - y)^2} \tag{13}$$

Linearization using Taylor series expansion and discard the higher order terms the result is:

$$\delta d_i = \frac{(x_i - x)\delta x + (y_i - y)\delta y}{\sqrt{(x_i - x)^2 + (y_i - y)^2}}$$

$$\delta d_i = \frac{(x_i - x)\delta x + (y_i - y)\delta y}{d_i} \tag{14}$$

where the distance d_i and the positions of the AP x_i , y_i are known. δx, δy are the only unknowns, x and y are known values because started with initial values for x and y to find another new set of solutions. The new position x and y can be used be again as known values. This process will continue until the absolute values of δx and δy are very small. The final values of x and y are the desired solution. This method refers to as the iteration method.

With δx and δy as unknowns, the above equation becomes a set of linear equations. This procedure refers to as linearization. The above equation (14) can be written in matrix form as:

$$\begin{bmatrix} \delta d_1 \\ \delta d_2 \\ \delta d_3 \end{bmatrix} = \begin{bmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \\ z_{31} & z_{32} \end{bmatrix} \begin{bmatrix} \delta x \\ \delta y \end{bmatrix} \tag{15}$$

where

$$z_{i1} = \frac{x_i - x}{d_i} \qquad z_{i2} = \frac{y_i - y}{d_i}$$

The solution of equation (15) is:

$$\begin{bmatrix} \delta x \\ \delta y \end{bmatrix} = \begin{bmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \\ z_{31} & z_{32} \end{bmatrix}^{-1} \begin{bmatrix} \delta d_1 \\ \delta d_2 \\ \delta d_3 \end{bmatrix} \tag{16}$$

The last equation doesn't give the desired results directly; however, the needed results obtained from it. to find the desired location solution, this equation used repetitively in an iterative-way. A value is used to calculate the needed result is reached, and this value can be defined as:

$$\delta v = \sqrt{\delta x^2 + \delta y^2} \tag{17}$$

Compare *δv* with an random threshold; if *δv* is higher than the threshold, add these values *δx*, *δy* to the initial position *x* and *y*, a new set of positions that will be expressed as *x-new*, *y-new*. These values will be

used as the initial-position in the next steps. Repeat the procedures until δv is lower than the threshold. The final solution is the desired user position (x, y).

Positioning solution with (n) APs when higher than three APs are existed, a more popular approach to find the user location is to use all the APs [8]. The positioning solution can be obtained in the same way. equation (13) is rewritten but (i= 1, 2, 3, 4,...., n) not at least 3

$$d_i = \sqrt{(x_i - x)^2 + (y_i - y)^2} \tag{18}$$

Linearize the equation using least square:

$$\begin{bmatrix} \delta d_1 \\ \delta d_2 \\ \delta d_3 \\ . \\ . \\ \delta d_n \end{bmatrix} = \begin{bmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \\ z_{31} & z_{32} \\ & . \\ & . \\ z_{n1} & z_{n2} \end{bmatrix} \begin{bmatrix} \delta x \\ \delta y \end{bmatrix} \tag{19}$$

$$z_{i1} = \frac{x_i - x}{d_i} \qquad z_{i2} = \frac{y_i - y}{d_i}$$

Equation (19) can be simplified as:

$$\delta d = Z\, \delta p \tag{20}$$

where:

$$\delta d = [\delta d_1 \; \delta d_2 \ldots \ldots \delta d_n]^T$$

$$Z = \begin{bmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \\ z_{31} & z_{32} \\ & . \\ & . \\ z_{n1} & z_{n2} \end{bmatrix}$$

$$\delta p = [\delta x \; \delta y]^T \tag{21}$$

where $[\ ]^T$ denotes the matrix-transpose, since Z isn't a square-matrix, it cannot be inverted-directly, the pseudo inverse of the Z can be used to obtain the solution [9]. The solution is:

$$\delta p = [Z^T Z]^{-1} Z^T \delta d$$

δx and δy Can be found from this equation. Repeat the procedures to get the user position (x, y).

2.1.5. Position Accuracy (Dilution of precision): Using the Dilution of precision (DOP) [8] used to measure user position accuracy. All the different DOPs are a function of APs geometry.

The smallest DOP value means the best APs geometry for measuring the user's location. To minimize the HDOP (2 Dimension x and y), the volume contained by the three APs must be maximized. The horizontal dilution of precision is defined as:

$$\text{HDOP} = \frac{1}{\sigma}\sqrt{\sigma_x{}^2 + \sigma_y{}^2}$$

where σ is the measured RMS error of the distance equation (19) which has a zero mean, σ_x, σ_y, are the measured RMS errors of the user position in the xy directions.

If the system is over determined (APs > 3), the Least Squares (LS) method is commonly used to find a solution that best fits the measurements:

$$\delta p = [Z^T Z]^{-1} Z^T \delta d \tag{22}$$

The correction vector δp is then used to update the user position solution estimate, and another iteration of the algorithm starts. The operations is stopped when the magnitude of the correction vector $\|\delta p\|$ becomes negligible.

The variance matrix of the position error, assuming zero mean of both position and measurement errors E { δp }=0 and E{δd}=0, can be written as [10]:

$$\text{Var}[\delta p] = [Z^T Z]^{-1} Z^T \text{Var}\,[\delta d]\, Z [Z^T Z]^{-1}$$

Assuming the errors of pseudorange measurements are mutually uncorrelated and have equal variance σ_d^2 the measurement error variance matrix can be expressed in the form.

$$\text{Var } [\delta d] = \sigma_d^2 \, I$$

The variance matrix of position error can be reduced to:

$$\text{Var } [\delta p] = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{yx} & \sigma_y^2 \end{bmatrix} = \sigma_d^2 \, [Z^T Z]^{-1} \tag{23}$$

where $H = [Z^T Z]^{-1}$ part is dimensionless and depends only on the transmitter geometry .

Position accuracy using Distance Root-Mean-Square (DRMS) can be determined as a product of the measurement error , and a dimensionless multiplier dependent only on the transmitter-receiver geometry. Hence, we can extract the HDOP from the following equation:

$$\alpha_{DRMS} = \sqrt{\sigma_x^2 + \sigma_y^2} = \sigma_\rho \sqrt{H_{1.1} + H_{2.2}}$$

where HDOP $= \sqrt{H_{1.1} + H_{2.2}}$

3. Location fingerprinting

One of the best useable techniques in indoor positioning is Fingerprinting that doesn't require line-of-sight with APs, does not need knowledge of the transmitters positions , does not need synchronization between transmitters, or between transmitters and receivers and is not affected by multipath, loss through obstacles and it consists of two phases Offline phase and online phase.

Offline phase where the system collected the RSS from the APs at selected positions in the area of interest, each measurement labeled with the correct location, (RSS, Location), the result called radio map as shown in Fig. 2**.** that obtains the construction of the Radio map. Online phase, the system uses the signal strength samples received from the Access-Points to compare with radio map (Data base) to calculate the target position as shown in Fig. 2**.**
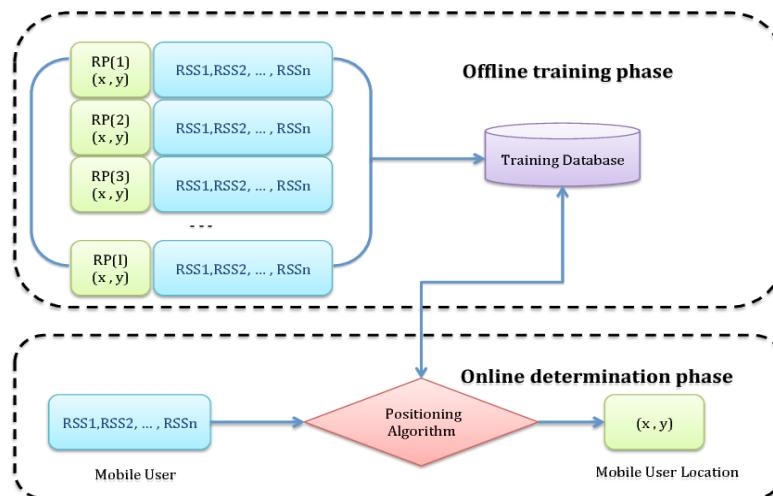

Fig. 2 Fingerprinting algorithm [11]

3.1. Preparation of RSS Map

Construction of the Radio-Map is very critical part of the process and it is constructed in the offline phase, which contains position and signal strength feature relate to the location. The motive for that is to reduce the memory size to store the radio-map and the computational time of the position estimation can be achieved. Different approaches to modify the raw data that can represent the RSS values by single value such as (mean, mode, median…etc.)

3.1.1. Raw data: In the offline phase, RSS values are measured at stationary locations for a period of time and stored in the radio map. The i_{th} element in the radio map has the form:

$$M_i = (\text{position of CP}, \{a_{ij} | \text{j} \in N_i \}), \qquad i = 1 \ldots M \qquad (24)$$

where a_{ij} is the list of RSS values measured from Access-Point AP_j at cell no i . The set of APs in range at the i_{th} calibration point is the set N_i , Thus the number of APs heard is the size of the list N_i, which is denoted as $|N_i|$. The number of samples measured from AP_j is the length of the list a_{ij}, denoted as $|a_{ij}|$.

3.1.2. Mean: The best common choice as a preprocessing method is to store only the mean of a_{ij} in the radio map [12]. The i_{th} element in the radio map has the form

$$M_i = (\text{position of CP}, \{|\,\bar{a}_{ij}\,| \text{ j} \in N_i \}, \}), \qquad i = 1, \ldots, M \quad (3.2)$$

where the mean of a_{ij} is

$$\bar{a}_{ij} = \frac{1}{|\,a_{ij}|} \sum_1^{|\,a_{ij}|} a^t{}_{ij} \qquad (25)$$

where $|\,a_{ij}| = $ length (a_{ij}) and $a^t{}_{ij}$ is the t_{th} element of the list a_{ij}.

3.1.3. Median: The median value is the middle number in sorted order and is the mean of the middle two numbers in sorted order [12].

$$M_i = (\text{position of CP}, \{|a| \text{ j} \in N_i \}, \}), \qquad i = 1, \ldots, M \qquad (26)$$

3.1.4. Mode: The mode value is most recurrent value of the measured data at each calibration point [12].

$$M_i = (\text{position of CP}, \{|m| \text{ j} \in N_i \}, \}), \qquad i = 1, \ldots, M \qquad (27)$$

where the value of (m) is the mode value.

3.1.5. Mean and variance: To give any information about the variation of the data. Fingerprints can be extended to store also the variance of the RSS samples [13]. Thus the i_{th} element in the radio-map has the form

$$M_i = (\text{position of CP}, \{ \bar{a}_{ij}, \widehat{\sigma^2}{}_{ij} \text{ j} \in N_i \}, \}), \qquad i = 1, \ldots, M \qquad (28)$$

where the variance of $\bar{a}_{ij}$ is

$$\widehat{\sigma^2}{}_{ij} = \frac{1}{|\,a_{ij}|} \sum_1^{|\,a_{ij}|} (a^t{}_{ij} - \bar{a}_{ij})^2 \qquad (29)$$

It can be used with the Gaussian distribution.

3.1.6. Histogram

$$M_i = (\text{position of CP}, \{ H_{a_{ij}} | \text{j} \in N_i \}, \}), \qquad i = 1, \ldots, M, \qquad (30)$$

where the histograms are defined as the set of pairs,

$$H_{a_{ij}} = \{(b_t, h_t)\} \qquad (31)$$

where b_t the histogram bin and the corresponding histogram height is h_t. Binrange or width is an interval b, which is normally equal in each histogram. The number of bins is

$$n = \frac{\max(a_{ij}) - \min(a_{ij})}{b}$$

3.2. Measuring of location online phase
The objective in the online phase estimates the position from the received signal (y) from several APs. The most common choice is to use one sec as a time step. The formulation of the measurement can be used the similar structure with the radio-map formulation, thus the measurement y has the form:

$$y = \{y_j | j \in N_y\} \in R \tag{32}$$

where y_j is the list of RSS values measured from AP and the set of APs in range is N_y.

3.3. WI-FI fingerprinting Methods

Radio-map based techniques can be categorized into two broad categories: deterministic techniques and probabilistic techniques [15,16, 17, 18].

3.3.1. Deterministic method: Deterministic method doesn't exploit all the Received signal strength samples at the two phases (offline phase and online phase), it can represent the RSS samples by a scalar value such as (Mean, Median and Mode) in Radio map that was modified or preprocessed before to reduce the size and minimize the time of the operation.

   a.   K-Nearest neighbor algorithm (KNN)

In the KNN approach, the vector $\bar{y}$ is used as a measurement and compared to the radio map. Compute the distance and sort it up, Sort neighbors and select RPs whose D_i is lower. The Euclidean norm is widely used, but the Manhattan norm is also common. fingerprints, that is :

$$\|x\|_p = \left(\sum_{i=1}^n |x|^p\right)^{\frac{1}{p}} \tag{33}$$

Substituting $x_i = \bar{y}_i - \bar{a}_{ij}$, where $\bar{y}_i$ is the average of RSS values measured from AP, in equation (33) gives:

$$\|X\|_p = \left(\sum_{i=1}^n |\bar{y}_i - \bar{a}_{ij}|^p\right)^{1/p} \tag{34}$$

In this section the MU's location estimator $\hat{x}$ is the average of the coordinates of the K-nearest neighbors

$$\hat{x} = \frac{1}{K} \sum_{i=1}^k p_i \tag{35}$$

where p_i is the position of the nearest neighbors. The parameter K that indicate the number of nearest calibration points was varied (K=2,3 and 4) to get best performances .

   b.   Weighted K-nearest neighbor

In this approach the location of the MU as a weighted average of the fingerprint locations is calculated, that is

$$\hat{x} = \frac{1}{\sum_{i=1}^k w_j} \sum_{i=1}^k w_i p_i^k , \qquad p_i^k \in L_K^n \tag{36}$$

where all weights [17]

$$w_i = d(\bar{y}, \bar{a}_i)^{-1}, \tag{37}$$

3.4. Probabilistic location estimation

Probabilistic approach [14] exploits all the samples of measurements collected during the offline phase and online phase. However, the distribution of random RSS samples is used at each cell so that the radio map should be consisted of more information such as (Mean, Variance or Histogram) depending on the RSS distribution. In this approach, the static location estimation is considered. The idea in the probabilistic methods in location fingerprinting is to compute likelihood functions between the RSS in the two phases and sorted it descending.

Using Baye's theorem, can be rewritten as:

$$\text{argmax}_i [P(i/y)] = \frac{\text{argmax}_i P(y/i).P(i)}{P(y)} \tag{38}$$

where y is RSS values measured from AP AP and i represents the calibration point.

Since P(y) is constant for all i [18], we can rewrite equation (38):

$$\text{argmax}_i[P(i/y)] = \text{argmax}_x[P(y/i).P(i)] \tag{39}$$

P(i) can be assumed that all the locations are equally likely and the term P(i) can be factored out from the maximization process. So Baye's theorem can be rewritten:

$$\text{argmax}_x[P(i/y)] = \text{argmax}_i[P(y/i)] \tag{40}$$

The term [P(y/i)] is called the likelihood function and there are several approaches for computing the likelihood function with (parametric and non-parametric approaches).

where

p(y|i) $= pv_i(y - \bar{a}_\iota)$ and $v_i = y - \bar{a}_\iota$.

We assume that the components of the random vector v_i are independent. Thus,

$$P(y/i) = \prod_j Pv_{ij}(y_j - \bar{a}_{ij}) \tag{41}$$

where j is the number of APAPs and y is RSS values measured from AP AP and i represent the calibration point.

3.4.1. Gaussian method: The Gaussian approximation has been used discussed in the literature [23, 24]. The Gaussian approximation can provide a good fit with some of the histograms.

The likelihood function can be computed as:

$$\boldsymbol{Pv_{ij}}\big(\boldsymbol{y_j} - \overline{\boldsymbol{a}}_{ij}\big) = \frac{1}{\sqrt{2\pi\sigma_{ij}{}^2}}\boldsymbol{e}^{\left(-\frac{\left(y_j - \bar{a}_{ij}\right)^2}{\sigma_{ij}{}^2}\right)} \tag{42}$$

3.4.2. Log-normal method: In this method the left-skewness of some of the histograms is exploited by log-normal distribution [21]. The left-skewness occurs due to the observation that the variations of the weaker RSS values are larger than the stronger RSS values.

The likelihood function can be computed as :

$$\boldsymbol{Pv_{ij}}\big(\boldsymbol{y_j} - \overline{\boldsymbol{a}}_{ij}\big) = \frac{1}{(y_j - \bar{a}_{ij})\sqrt{2\pi\sigma_{ij}{}^2}}\boldsymbol{e}^{\left(-\frac{\left(\log(y_j - \bar{a}_{ij}) - \bar{a}_{ij}\right)^2}{\sigma_{ij}{}^2}\right)} \tag{43}$$

3.4.3. Exponential method: In the case of skewed histograms, the exponential function provides better approximation and improves of different fingerprints close to each other.

The likelihood function can be computed as:

$$\boldsymbol{Pv_{ij}}\big(\boldsymbol{y_j} - \overline{\boldsymbol{a}}_{ij}\big) = \frac{1}{2}\boldsymbol{e}^{|(y_j - \bar{a}_{ij})|} \tag{44}$$

3.4.4. Kernel method: The kernel method is a non-parametric method, because the idea is to estimate the underlying probability density function from the sample pattern. The idea is to impose a probability mass to a ''kernel" around each of the samples a_{ij}^t.

The kernel density estimation makes it probable to interpolate the RSS data to the entire signal- strength space and fill the possible incorrect gaps in the RSS histograms [4].

The computation of the likelihood is done by using the equation

$$P(y/i) = \frac{1}{|a_{ij}|h} \sum_1^{|a_{ij}|} K\left(\frac{y-a_{ij}}{h}\right) \tag{45}$$

where K($\cdot$) denotes the kernel function, a_{ij}^t is the t_{th} element of the a_{ij} vector and h > 0 is a smoothing parameter, which determines the width of the kernel [23]. Some of these kernel functions [12] are:

a.   Gaussian kernel

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{\left(-\frac{x^2}{2}\right)} \tag{46}$$

b.   Exponential kernel

$$K(x) = \frac{1}{\sqrt{2}} e^{(-|x|)} \tag{47}$$

where $x = \frac{y-a_{ij}}{h}$

4. Experimental work and results

The experimental works were divided into two parts, the first one is used to the Trilateration technique using alternative no of WI-FI access points at least 3 APs to find the 2D position inside the long narrow corridor, these APs were located with a good geometry shape to give minimum HDOP then the path loss exponent and the shadowing parameters that compensates the multipath effects were measured. The corridor was divided to rectangular coordinates (x,y)   Started from (0,0) to (1,84) .

The second experimental work used fingerprinting. The radio map were collected in a department corridor with a dimension 8×85 $m2$ which divided into 20 grid cells and the CPs (calibration points) at the center of the cells as shown in Figure 1) . RSS values from one AP at the CPs varied through the grids due to obstacles such as human, doors…etc., all the grids has the same area. The radio map was collected at 20 CPs with a fixed orientation to obtain a unique feature vector to each cell but different RSS samples at different time1 sample/sec. Three Radio maps were constructed ,The first radio map has 60 RSS samples with 60sec measurement period, the second has 100 RSS samples with 100 sec measurement period and the second has 300 RSS samples with 300 sec measurement period. using personal laptop . With Vistumbler program that retrieves the basic information of each Aps (SSID, RSS) and RSS values are in dBm scale as shown in Fig. 3. There are 3 Aps were placed inside the corridor so that all Aps are heard at each CPs
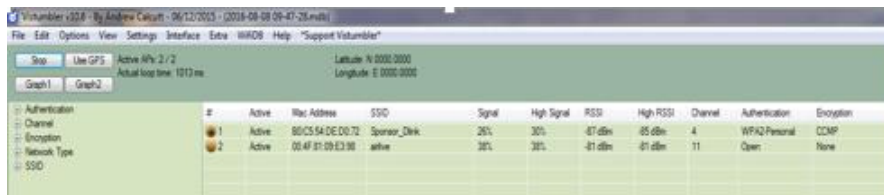

Fig. 3 windows desktop program

This application receive from all available Active Aps, Select the three APs that used in experiments and Import the data of the 3 APs to Excel sheet as shown in Fig. 4.
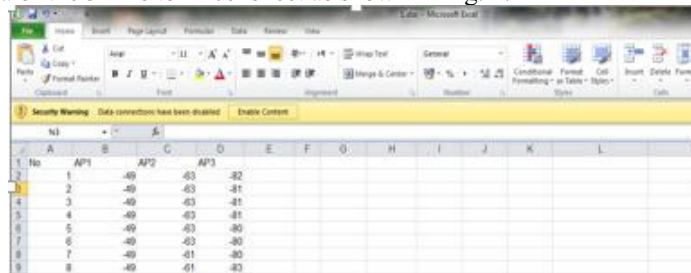

Fig. 4 simplified the data to use in the experimental

The RSS values from the same AP at the same position varied due to obstacles such as human, doors…etc., all the grids has the same area as shown in Fig. 5.
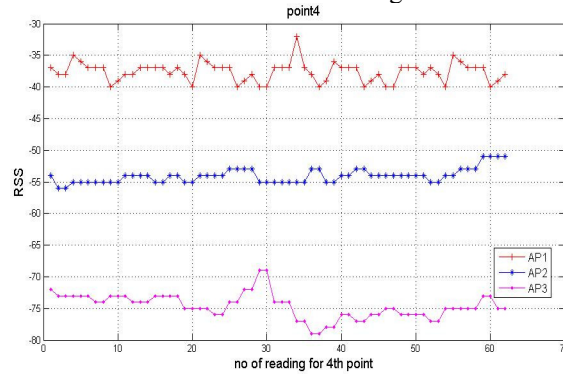

Fig. 5 variation of RSS samples at the the same point from 3 Access point.

4.1. Location using Trilateration Method:

The numbers of APs and the locations depend on the HDOP, The smallest DOP value means the best APs geometry for calculating user position. To minimize the HDOP (2 Dimension x and y), the volume must be maximized that contained by numbers APs.

So there are many trails to achieve the best geometry of the APs that give best HDOP as shown in table (2.2), the 1st trial, three APs were located as a triangle shape to achieve the best geometry, the HDOP distribution shown in Fig. 6.
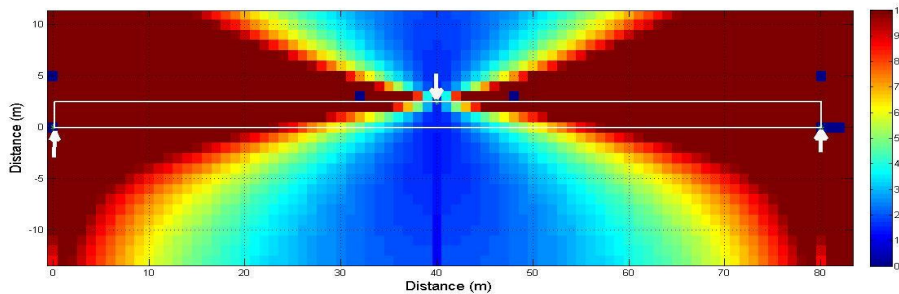
Fig. 6 HDOP distribution over the corridor using 3 Aps and
the distribution of the APs inside the corridor

So using three AP is not prefer due to the bad HDOP , the 2nd trial using 4 APs each AP at each corner as shown in Fig. 7
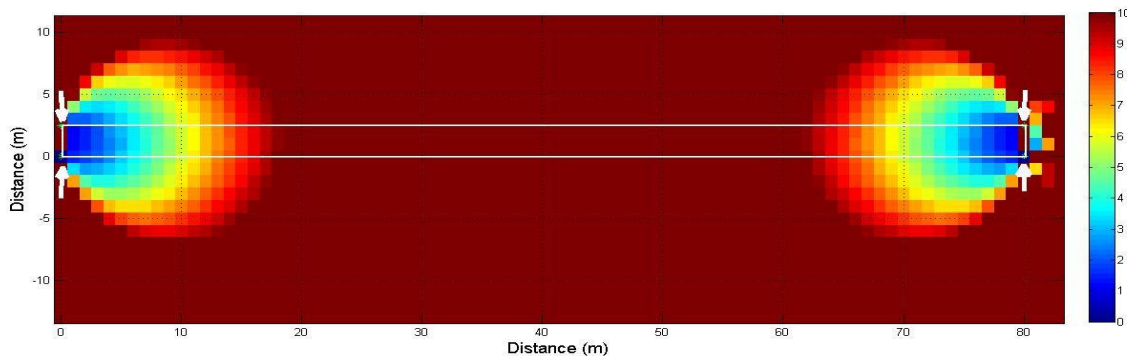

Fig. 7 HDOP distribution over the corridor using 4 Aps

So using 4 AP is not prefer due to the bad HDOP due to the small width , the 3rd , 4th and the 5th trial using 5, 6, 8 APs consequently , the distribution of Aps shown in the Fig. 8 and Fig. 9.
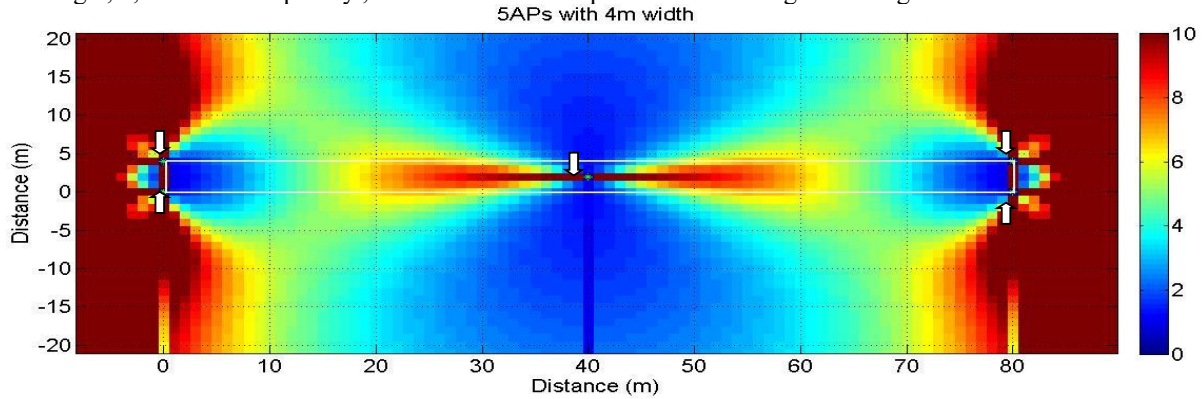

Fig. 8: HDOP distribution over the corridor using 5 Aps and
the distribution of the APs inside the corridor

Using 5 APs as shown in Fig. (8) not prefer due to the bad HDOP, due to the small width but it is better than 3 and 4 APs.
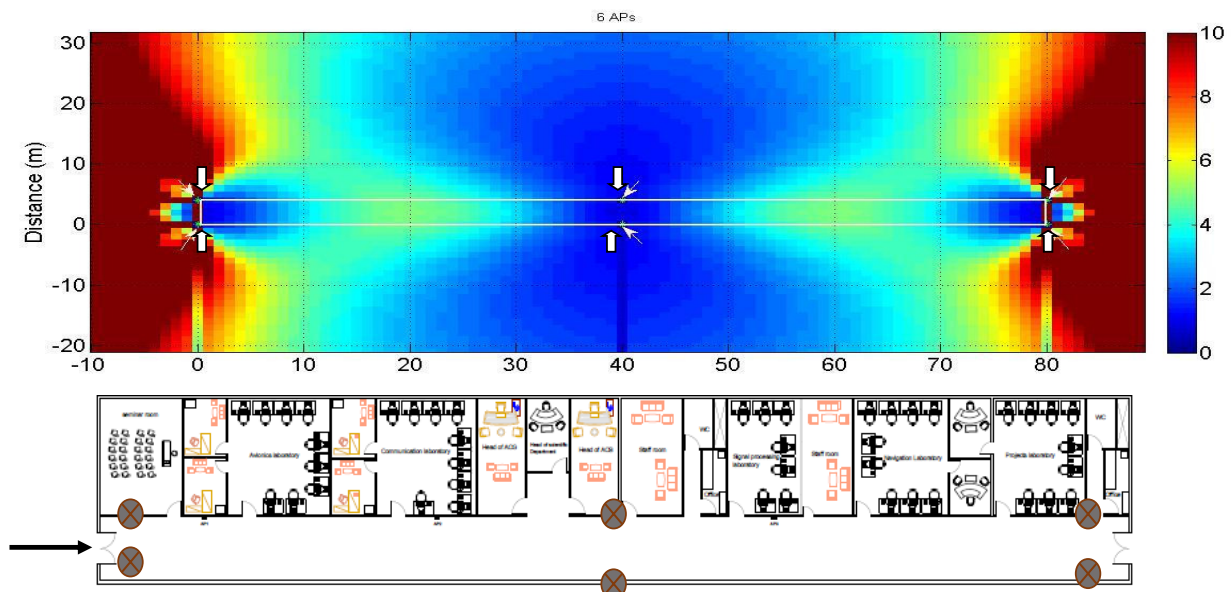

Fig. 9: HDOP distribution over the corridor using 6 Aps and the
distribution of the APs inside the corridor

Using 6 APs as shown in Fig. (4.9) is better than 3, 4 and 5 Aps due to the best HDOP.
AP1= [0 0]; AP2= [4 0]; AP3= [0 80];
AP4= [4 80]; AP5= [4 40]; AP6= [0 40].

Estimated a value of n that will minimize this MSE (n) value (MMSE) minimum mean square error. So the value of the **path loss exponent (n=2.489)** the value of the standard deviation that represent the **shadowing parameter is σ= 4.097**

The distance between the transmitter (Access point) and the receiver can calculated from Equation (8)

$$d = d_0 \times \exp\left(\frac{P_r(d) - P_r(d_0) - \chi_\sigma}{10n}\right)$$

Using different parameters for nearest neighbor K (2,3,4) and parameter p leads to different norm values to obtain the most suitable parameters which give more suitable position Using RMSE estimator .

$$RMSE= \sqrt{(\frac{1}{n}\ \sum_{j=1}^{n}(p_j - \widehat{p_j})^2)}$$

There are three chosen stationary test cases, test points 1, 2 and 3 to test which   parameters are suitable for the experimental as depicted in Fig. 10, The results of the different No of Access point shown in Fig. 11.
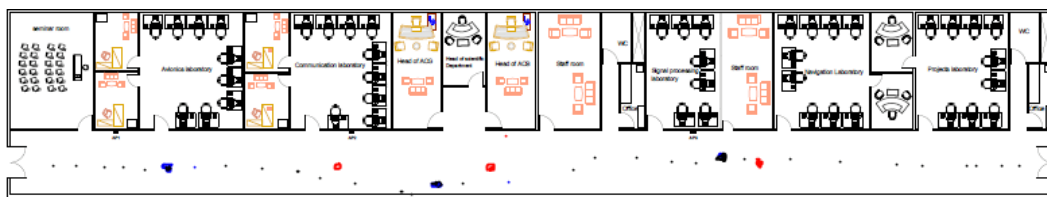

Fig. 10 shows the error of the user location inside the corridor

where ● is the actual position and   ● the estimated position

Table 1. The average error due to NO of Access point

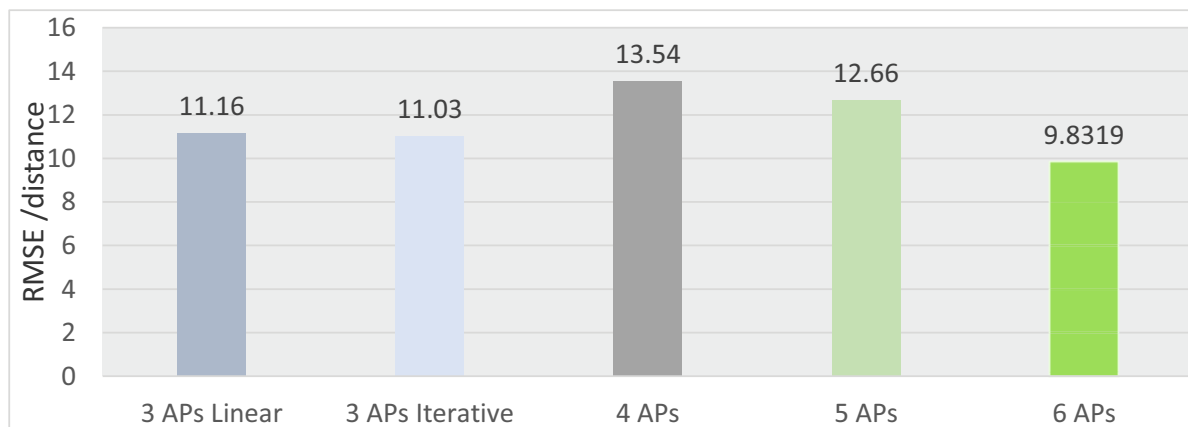| No of Aps	Avg error
3APs linear	10 m
3APs iterative	8 m
4Aps	7.5 m
5Aps	6.4 m
6 Aps	5.5m

Fig. 11 RMSE for Alternative No of Aps

4.2. Location using Fingerprinting method

4.2.1. Radio Map: The radio map was collected in a department corridor with a dimension 2.5×85 m^2 which divided into 20 grid cells have the same area and CPs in the center of the cells as shown in Fig. 12. Due to the width of the corridor is small so the localization in one Dimension (1 D), RSS values from one AP at the CPs varied through the grids due to obstacles such as human, doors…etc., all the grids has the same area.

To achieve a suitable Radio Map inside the corridor that give best performance according the experimental, there were three Radio maps constructed before any processing to reduce the size and the time of the operation. The 1st radio map has 60 RSS samples with 60sec measurement period, the 2nd has 100 RSS samples with 100 sec measurement period and the 3rd has 300 RSS samples with 300 sec measurement period using personal laptop device. Vistumbler software was used such that the basic information of each APs (SSID, RSS) and RSS values are retrieved in dBm scale. There are 3 APs were placed inside the corridor so that all APs are heard at each CPs

Fig. 12: The distribution of APs and CPs inside the corridor

 Is the AP

Note that the samples of default reading more than -90dBm and lower than -29dBm are excluded and the samples that exceed the median value of the RSS samples by 10dBm UP DOWN to avoid the unexpected values due to any reason and guarantee a good performance.

Effect of parameters in algorithms

Deterministic approach: There are several parameters in algorithms such as {1-No of APs 2- No of neighbors K parameter K=2, 3, and 4. 3- P parameter at measuring P-norm distance between the RSS at the two phases. 4- No of RSS samples that constructed the Radio Map 60,100 and 300 samples 5- Radio map RSS samples represented by scalar values such as (Mean, Median and Mode)} SO there were a lot of trials as shown in Fig. 13, to find the best parameters that work together to achieve the best performance inside the department's corridor
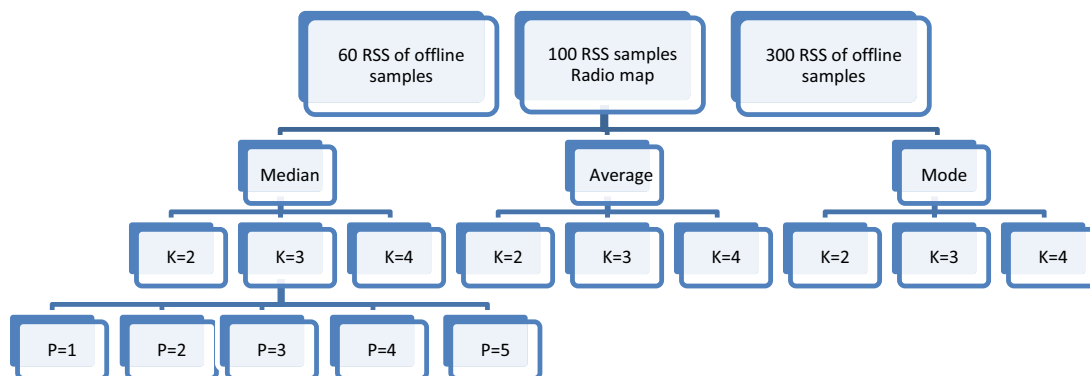

Fig. 13. Effects of the parameters on performance

The K-NN method and WKNN were tested. The Results after all the experiments that reach to 135 trials with 3 APs by varying all the parameters P, K, raw data and raw map samples obtained the best estimated position for KNN algorithm at raw map 100 samples with mode raw data at P=2 Euclidian distance and K= 2 that gives the smallest value of RMSE is 1.2 as shown in Fig. 14

The best estimated position for WKNN algorithm at raw map 100 samples with mode raw data at P=5 and K= 2 that gives the smallest value of MSE is 1.08 as shown in Fig. 15
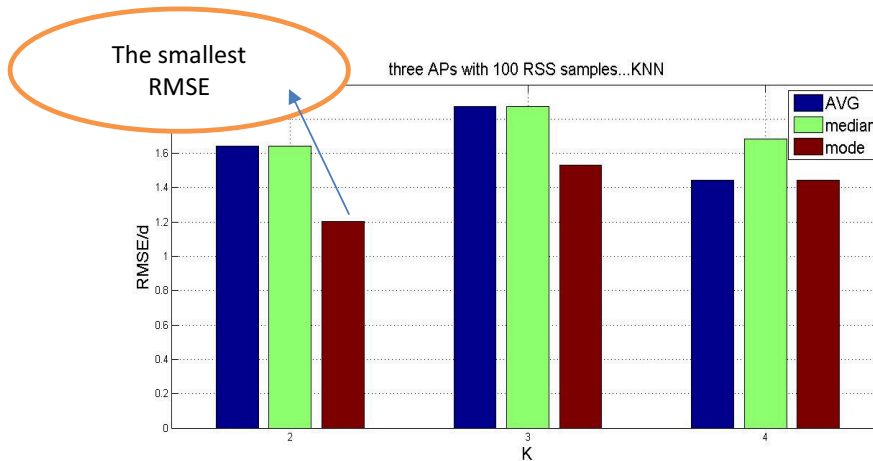

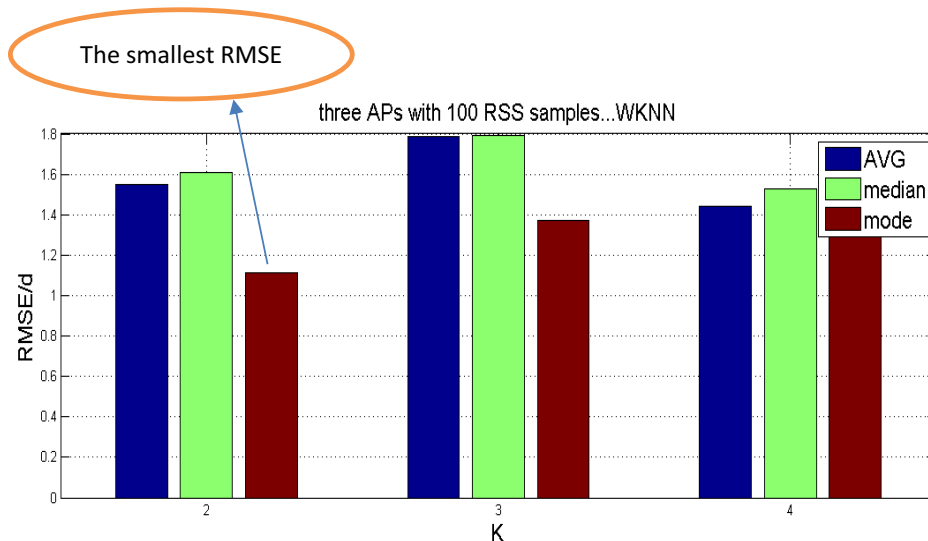Fig. 14 The best value of RMSE at KNN algorithm

Fig. 15. The best value of RMSE at WKNN algorithm

Using 3 Aps is more suitable than one or two Aps and this decision was based on the least error at the same parameters, for KNN algorithm at raw map 100 samples P=2 Euclidian distance which gave the best results at KKN and more experimental as shown in Fig. 16.

Probabilistic Approach: The measurement likelihood was computed using different parametric approximations of the normalized histograms. In the tests, the exponential and Gaussian approximation produced the smallest RMSE as listed in Table 2
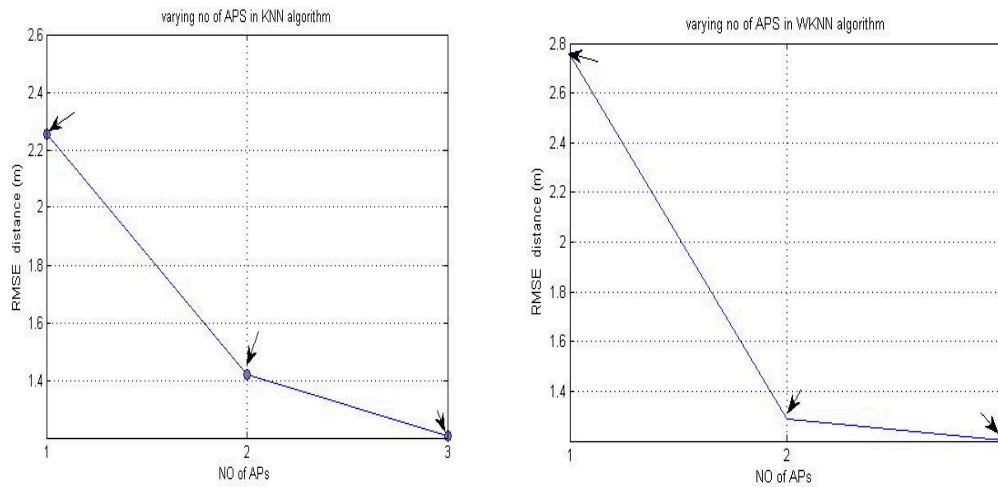
Fig. 16. Comparison between using one, two or three APs in KNN and WKNN

Table 2. The exponential and Gaussian approximation produced the smallest RMSE

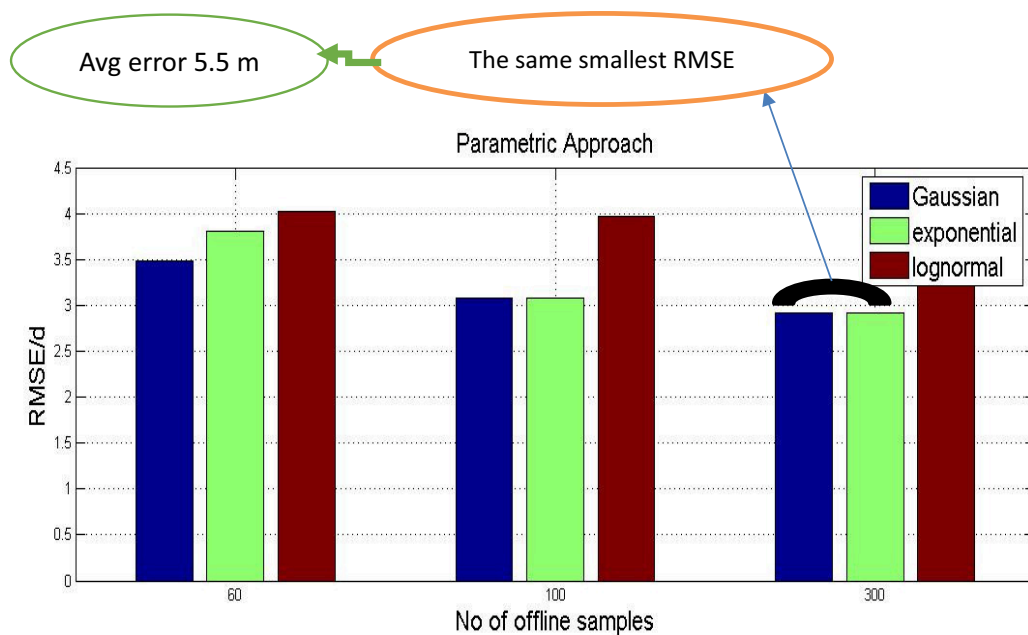| Function	At raw map 60 samples RMSE (m)	At raw map 100 samples RMSE (m)	At raw map 300 samples RMSE (m)
Gaussian Method	3.48	3.0822	2.9155
Exponential Method	3.8079	3.08	2.9155
Log normal Method	4.023	3.969	3.632

Fig. 17: The exponential kernel function is claimed to provide good results
at raw map 100 samples

The exponential kernel function is claimed to provide good results at raw map 100 samples and in this work, it also gave better performance compared to the another kernel function as shown in Table 3, and Fig. 18.

Table 3. The exponential kernel function is claimed to provide good results at raw map 100 samples

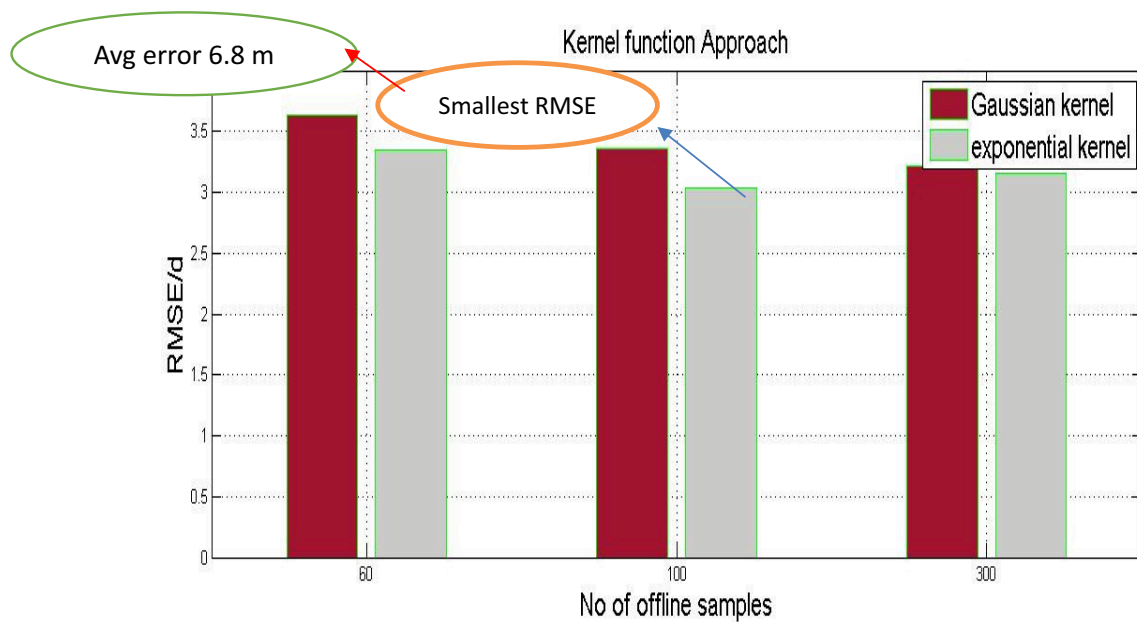| Function	At raw map 60 samples RMSE (m)	At raw map 100 samples RMSE (m)	At raw map 300 samples RMSE (m)
Gaussian kernel	3.632	3.356	3.21
Exponential kernel	3.35	3.032	3.156

Fig. 18. shows the best performance at Kernel exponential

After comparing between all techniques of the fingerprinting method, deterministic WKNN is the best from the point of average error in meters as listed in Table 4.

Table 4: comparison between all fingerprinting methods

Methods	Avg error
Deterministic KNN	3.75m
Deterministic WKNN	2.8m
Parametric Approach (Exponential-Guassian)	5.5m
Non parametric Approach (Exponential Kernel)	6.8m

After comparing between all techniques of the indoor navigation method, fingerprinting method is the best from the point of average error, number of access points and the cost as listed in Table 5.

Table 5. Comparison between trilateration and fingerprinting methods

Methods	Avg error	No of APS	Cost
Trilateration Methods Using 6APs	5.5m	6	High
Fingerprinting Methods Deterministic WKNN	2.8m	3	Low

5. Conclusions

In this paper, the indoor localization was introduced as a very important technology last decades and also the different techniques to achieve the experimental work using Trilateration technique using RSS Ranging and Fingerprinting technique.

Trilateration technique used the RSS Ranging to measure the distance between the AP and the target after computing the path loss exponent and the shadowing parameter , there were a lot of APs (WI-FI) in different position inside the corridor to choose the best positions of the APs and the no of the APs according to the value of the HDOP which is the error in x-axis and y-axis , the best performance was obtained with 6 APs distributed as shown in Fig.(4-24) but the no of APs that used was large to do an indoor localization inside this corridor ,so using an alternative method to avoid the large no of APs called Fingerprinting method.

Various position fingerprinting algorithms were considered by presented the mathematical basis of the algorithms and testing with WLAN Received-Signal-Strength measurement data. The mathematical formulation was executed from several points of view, the parameters were varied in the experimental to find the best performance. The environment variables, such as the number of Access-Points and the RM density, were also changed and the algorithms were compared also in these varying circumstances. The basic goal was to model the location as a random variable which led to partition the area of interest into rectangular cells instead of considering only the individual CPs. Deterministic methods, such as the (KNN) or (WKNN) method, were formulated with the discrete location variable, as they are presented in the literature .probabilistic methods were formulated by estimation of likelihood functions with several approaches (non-parametric and parametric) are also explained. After comparing each other the best technique is Deterministic WKNN that achieve the lowest RMSE.

6. References

[1]   M. Vossiek, L. Wiebking, P. Gulden, J. Wieghart, C. Hoffman, and P.   Heide, Wireless local positioning. IEEE Microwave magazine, 2003, pp. 77–86.2003

[2]   H. Liu, H. Darabi, P. Banerjee, J. Liu, "Survey of wireless indoor positioning techniques and systems", IEEE Trans. Syst. Man Cybern. C: Applications and Reviews, vol. 37, no. 6, pp. 1067-1080, Nov. 2007.

[3]   A. Bensky, Wireless Positioning Technologies and Applications,   Norwood, MA, USA:Artech House, 2008

[4]   V. Honkavirta, T. Perala, S. Ali-Loytty, and R. Piche, "A comparative survey of WLAN location fingerprinting methods," in Positioning, Navigation and Communication, 2009. WPNC 2009. 6th Workshop on, March 2009, pp. 243 –251.2009

[5]   Theodore Rappaport, Wireless Communications: Principles and Practice, Prentice Hall PTR, Upper Saddle River, NJ, 2001

[6]   Pu, C.-C. (2009). Development of a New Collaborative Ranging Algorithm for RSSI Indoor Location 2009

[7]   J. Vander Stoep, "Design and implementation of reliable localization algorithms using received signal strength" in , Diss. University of Washington, 2009

[8]   Tracking in WSN, PhD Thesis, Dongseo University, South Korea.Fundamentals of Global Positioning System Receivers, A Software Approach Kay, S. M., *Fundamentals of Statistical Signal Processing* Estimation Theory, Chapter 8, Prentice Hall, Englewood Cliffs, NJ 1993.

[9]   Spilker, J. J. Jr., "Satellite constellation and geometric dilution of precision," Chapter 5, and Axelrad, P., Brown, R. G., "GPS navigation algorithms," Chapter 9 in Parkinson, B

[10]   J. Safar, et al., "Analysis, Modeling, and Mitigation of Cross-Rate Interference in eLoran," Navigation, vol. 63, pp. 295-319, 2016

[11]   Jiang Landu, "A WLAN fingerprinting based indoor localization   technique", *Diss. University of Nebraska*, 2012

[12]   V. Honkavirta, "Location fingerprinting methods in wireless local area networks," M.Sc. thesis,   Tampere   University   of   Technology,   2008.   [Online]. Available:http://math.tut.fi/posgroup/2008

[13]   ParamvirBahl and Venkata N. Padmanabhan. Radar: An in-building rf-based user location and tracking system. INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE, 2(10):775‑784, March

[14]   TeemuRoos, Petri Myllym¨aki, Henry Tirri, Pauli Misikangas, and JuhaSiev¨anen. A probabilistic approach to wlan user location estimation. International Journal of Wireless Information Networks, 9(3):155‑163, July 2002.

[15]   BAHL, P., AND PADMANABHAN, V. N. RADAR: An In-Building RF-based User Location and Tracking System. In IEEE Infocom 2000 (March 2000), vol. 2, pp. 775–784.

[16]   CASTRO, P., CHIU, P., KREMENEK, T., AND MUNTZ, R. A Probabilistic Location Service for Wireless Network Environments. Ubiquitous Computing 2001 (September 2001).

[17]   CASTRO, P., AND MUNTZ, R. Managing Context for Smart Spaces. IEEE Personal Communications (OCTOBER 2000).

[18]   HAEBERLEN, A., FLANNERY, E., LADD, A., RUDYS, A., WALLACH, D., AND KAVRAKI, L. Practical Robust Localizationover Large-Scale 802.11 Wireless Networks. In 10th ACM MOBICOM (Philadelphia, PA, September 2004).

[19]   Binghao Li, James Salter, Adrew G. Dempster, and Chris Rizos. Indoor positioning techniques based on wireless lan. Technical report, School of Surveying and Spatial Information Systems, UNSW, Sydney, Australia, 2006.

[20]   Rehim, A. A. Y. A. and M. Amin (2004). HORUS: A WLAN-based indoor location determination system.2004

[21]   K. Kaemarungsi, "Distribution of WLAN received signal strength indication for indoor location   determination," National Electronics and Computer Technology Center,Thailand, Tech. Rep., 2006.

[22]   B. Li, J. Salter, A. G. Dempster, and C. Rizos, "Indoor positioning techniques based on wireless LAN," School of Surveying and Spatial Information Systems, UNSW,Sydney, ustralia, Tech. Rep., 2006..

[23]   R. O. Duda, P. E. Hart, and D. G. Stork, Pattern Classi-fication. John Wiley, Sons Inc., 2001.