An Algorithm Based on Bayes Inference And K-nearest Neighbor For 3D WLAN Indoor Positioning

Hitalo J.B. Nascimento, Emanuel B. Rodrigues, Francisco R. P. Cavalcanti and Antonio Regilane L. Paiva.

Abstract—**This paper proposes a hybrid algorithm based on Bayesian inference and K-Nearest Neighbor to estimate the three-dimensional indoor positioning implemented from a fingerprint technique. Additionally, a comparison was made between the main algorithms discussed in literature. The experiments were conducted in a typical building with two floors with $180\mathrm{m}^2$ and four access points.**

The proposed solution showed a precision in the location of the rooms of 97% and 90% the estimates were at maximum three meters away from the actual location, furthermore, such method has lower variability than other algorithms, with deviation in relation to the mean reaches of 37.62%.

***Keywords*— 3D Indoor positioning, Fingerprint, Bayes inference, K-Nearest Neighbor**

I. INTRODUCTION

With the technological development it became possible the location people and objects in outdoor environments, using tools such as Global Positioning System (GPS). However, in dealing with indoor locations, this tool is not efficient, because of the high variability in this type of environment, compared to outdoor environments, because besides the power and gain antennas, the variability is dependent on the type of construction and internal structures, such as walls, floors and partitions walls. In this sense and considering that indoor location has various applications, such Emergency System, Paths at malls, Schools, University Campus, Airports, Hospitals, Mobile robot localization in indoor environment, just to name few examples, several works have addressed to indoor location. A survey of the main techniques classified three typical location estimation schemes (triangulation, scene analysis and proximity), which are discussed in [1] [2]; whereas [3], [4], [5] and [6] propose the application of artificial neural network in the problem. K-Nearest Neighbor (KNN) and Bayes methods are addressed in [7], [8], [9], [10], [11] and [12].

In our work, we propose a solution to the problem through a hybrid algorithm based on Bayesian inference and K-Nearest Neighbor, using the Received Signal Strength (RSS) as input. Furthermore we compare this method with Artificial Neural Network (Multilayer Perceptron - MLP and Radial Basis Function - RBF), K-Nearest Neighbor. Finally, these algorithms, implemented from a fingerprint technique, are compared to the method based in geometric properties triangulation.

Hitalo J.B. Nascimento, Emanuel B. Rodrigues, Francisco R. P. Cavalcanti and Antonio Regilane L. Paiva are membres of GTEL - Wireless Telecom Research Group, Department of Teleinformatics Engineering, Universidade Federal do Ceará, Fortaleza, Brazil. E-mails: hitalo@gtel.ufc.br, emanuel@gtel.ufc.br, rodrigo@gtel.ufc.br, lane.p.l@hotmail.com.

This article is organized as follows: section II gives an overview of the algorithms used in this work. The test results are presented in Section III. Finally, Section IV summarizes the results and suggests guidelines for future work and references.

II. POSITIONING ALGORITHMS

In literature, Wi-Fi indoor positioning can be classified into two main categories: triangulation (method based on geometric properties) and fingerprint (signal data collection or pattern recognition technique). This classification is shown in Figure 1

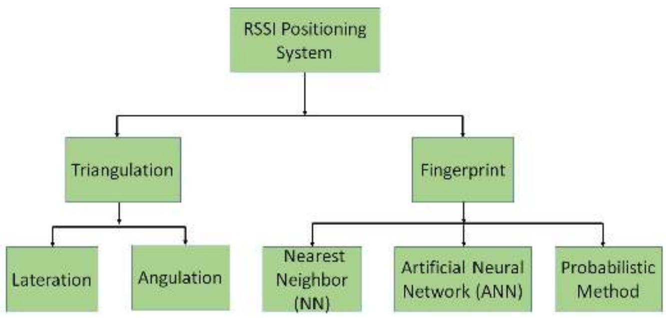

Fig. 1: *Taxonomy of RSS positioning techniques in Wi-Fi.*

A. TRIANGULATION

Triangulation uses geometric properties of triangles to estimate the target location [1]. This method is classified into two categories, lateration and angulation. Lateration positioning method measures the distance between the mobile terminal and a set of at least three reference points (RP), as shown in Figure 2 [13] [14]. This technique extends to the three-dimensional case by simply adding a fourth reference point. The conceptualization would then start with a sphere associated with the first reference point. This sphere would then be reduced to a circle after the introduction of the second reference point, reducing the situation to the two-dimensional case. This method is similar to the technique used by Global Positioning System (GPS) [6] [7] [8]. The most common methods used for estimating of distance are: Time of Arrival (TOA), Angle of Arrival (AOA), and Received Signal Strength.

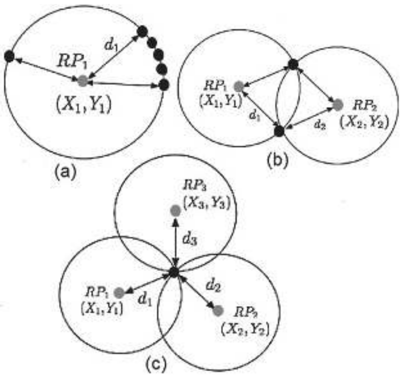

Fig. 2: *Lateration positioning method. The grey circles represent the reference points and the black circles symbolize the mobile terminal in (a) single reference point (b) two reference points and (c) three reference points [13] [14].*

The angulation to determine the position of an object from multiple intersections resulting angles [1] . The figure 3 shows this procedure.

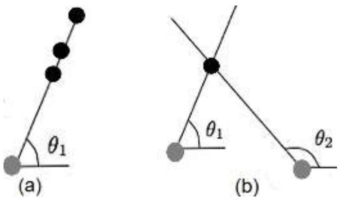

Fig. 3: *Angulation technique. The grey circles represent the reference points and the black circles symbolize the mobile terminal in (a) one reference point and (b) two reference points [13][14].*

B. FINGERPRINT

Fingerprint is a technique based on pattern recognition, dividing the location system into two phases, off-line and online. In the offline phase, the RSS measures database is developed. The online phase consists of reading a value RSS and through a classification algorithm to compare this RSS with the values stored in the database during the off-line phase, thus obtaining location. Figure 4 summarizes this procedure, whereas in subsection to 1, 2 and 3 are discuss the main algorithms based in this technique, including the method proposed in this work.

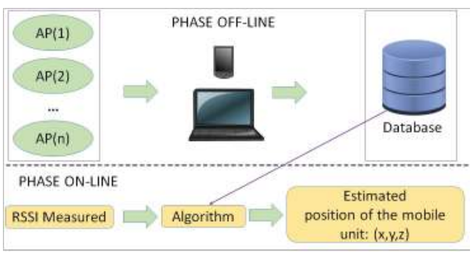

Fig. 4: *Indoor Positioning based on Fingerprint*

1) K- NEAREST NEIGHBOR (KNN): KNN is a non parametric supervised learning algorithm where new objects are classified based on a similarity measure. The main idea of the KNN technique is based on the calculation of the distance between the measured RSS value Obtained in the online phase and the location information of the RSS database – Fingerprinting. The most common metric for similarity are [15]:

.

- Euclidean Distance:

$$Dist(r,s) = \sqrt{(x_r - x_s)(x_r - x_s)'} \tag{1}$$

- Standardized Euclidean Distance:

$$Dist(r,s) = \sqrt{(x_r - x_s)D^{-1}(x_r - x_s)'} \tag{2}$$

- Mahalanobis Distance:

$$Dist(r,s) = \sqrt{(x_r - x_s)V^{-1}(x_r - x_s)'} \tag{3}$$

- Manhattan Distance:

$$Dist(r,s) = \sum_{j=1}^{n} |x_{rj} - x_{sj}| \tag{4}$$

- Minkowski Distance:

$$Dist(r,s) = \sqrt[p]{\sum_{j=1}^{n} |x_{rj} - x_{sj}|^p} \tag{5}$$

- Cosine Distance:

$$Dist(r,s) = \left(1 - \frac{x_r x_s'}{\sqrt{x_r' x_r}\sqrt{x_s' x_s}}\right) \tag{6}$$

- Correlation Distance:

$$Dist(r,s) = 1 - \frac{(x_r - \overline{x}_r)(x_s - \overline{x}_s)'}{\sqrt{(x_r - \overline{x}_r)(x_r - \overline{x}_r)'}\sqrt{(x_s - \overline{x}_s)(x_s - \overline{x}_s)'}} \tag{7}$$

where x and x' denote a vector column and its transpose, respectively. x_r and x_s, are the r^{th} and s^{th}, respectively. x_{rj} indicate j^{th}, features of the r^{th} sample in the data set. $\overline{x}_r$ indicates the mean of all features of the r^{th}. D is diagonal matrix with diagonal elements given by v_i^2. V is a sample of covariance matrix.

2) ARTIFICIAL NEURAL NETWORK (ANN): An artificial neural network can be defined as a mathematical model that simulates the behavior of the brain. The basic structure of an ANN is composed by a set of artificial neurons that receive as input signals $(x_1, x_2, \cdots x_n)$. Each input is assigned a weight and then is obtained the weighted inputs sum. In the next step is applied an activation function and a threshold value that will generate the output result. There are several activation functions. Some are shown in figure 5 [16]. In a general way, ANN can be classified into two categories: feed-forward, in which the graph is acyclic and recurrent networks, opposed to feed-forward, cycles and loops occur for feedback connections.

Typical examples of recurrent ANN's are: Kohen Son and Hopfield network.

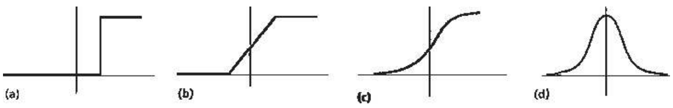

Fig. 5: *Different types of activation functions: (a) threshold, (b) piecewise linear, (c) sigmoid, and (d) Gaussian.*

In indoor positioning, ANN receives as input one RSS vector and coordinates related to the same. The output is a vector two elements or three elements to 2D space and 3D, respectively.

3) BAYES INFERENCE: The probabilistic method is based on the bayes theorem given by:

$$\mathbb{P}(\theta|x) = \frac{\mathbb{P}(x|\theta)\mathbb{P}(\theta)}{\int \mathbb{P}(x|\theta)\mathbb{P}(\theta)d\theta} \tag{8}$$

Here, $\mathbb{P}(\theta)$ is called prior probability, it is the probability of θ before X is observed. $\mathbb{P}(\theta|x)$ it is the posterior probability. It represents the probability of θ given X and $\mathbb{P}(X|\theta)$ is the likelihood function.

The Bayesian model applied to the problem of indoor location has the following rule:

Let $L_1, L_2, \cdots, L_n$., a set of possible locations, Select (L_i), if $P(L_i|RSS) > P(L_j|RSS)$, for $i, j = 1, 2, \cdots, n$ and $i \neq j$. $\mathbb{P}(L_i|RSS)$, The likelihood of a set is at a given location in an observed RSS. The probability in question, can be orbited as following:

$$\mathbb{P}(L_i|RSS) = \frac{\mathbb{P}(RSS|L_i)\mathbb{P}(L_i)}{\sum_{k=1}^{n} \mathbb{P}(RSS|L_k)\mathbb{P}(L_k)} \tag{9}$$

There are several methods for estimating of likelihood function $\mathbb{P}(RSS|L_i)$. Some of these methods are shown below:

.

- Histogram approach: The histogram approach is probably the oldest method for estimating density functions. This approach subdivides the domain into bins and counts the number of samples n_b which fall into each bin. To estimate $\mathbb{P}(RSS|L_i)$ obtains a frequency distribution of the signal in the environment.The aim is to verify the percentage of each range RSS with respect to all locations of the indoor environment.

- Kernel Method: In probability and statistics, the estimation kernel is a non-parametric method for a random sample of distribution with unknown density given by:

$$f_X(x) = \frac{1}{nh}\sum_{i=1}^{N} K\left(\frac{x - X_i}{h}\right) \tag{10}$$

Where K (.) is kernel function: $\int_{\mathbb{R}} K_X(x)dx = 1$, h is a smoothing parameter non-negative called the bandwidth or window width and $X_i, i = 1, 2, ...n$ are random variables independent and identically distributed.

The density estimate for an observation RSS given L_i modeled by normal distribution is given by:

$$f_x(x) = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{h\sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{RSS - RSS_i}{h}\right)^2\right] \tag{11}$$

4) PROPOSED METHOD: The proposed method consist in estimate the posterior probability using the K-Nearest Neighbor. The basic idea of such an estimate is consider a hypersphere with volume V that surrounds the K neighbors of the point RSS, given by:

$$\mathbb{P}(RSS|L) = \frac{k_i}{N_i V} \tag{12}$$

The priors probabilities are approximated by:

$$\mathbb{P}(L) = \frac{N_i}{N} \tag{13}$$

$$\mathbb{P}(RSS) = \frac{k}{NV} \tag{14}$$

$$\mathbb{P}(L|RSS) = \frac{\mathbb{P}(RSS|L)\mathbb{P}(L)}{\mathbb{P}(RSS)} = \frac{\frac{k_i}{N_i V}\frac{N_i}{N}}{\frac{k}{NV}} = \frac{k_i}{k} \tag{15}$$

III. EXPERIMENTAL RESULTS

Our experiments were conducted in a typical building with two floors ($180m^2$) and four access points from different vendors (Figure 6). The RSS measurements were performed in 90 collection points distributed on two floors with a notebook Pentium Dual-Core 2.3 GHZ, memory 2GB, Realtek adapter rt1819se wireless lan, 802.11b/g/n. All techniques were implemented in R language, using the following hardware: AMD FX - 8120 - Desktop. Memory: 4GB, DDR 3. For each collection point were obtained 100 measurements with the Vistumbler software [17], generating the RSS map shown in figures 7.

In figure 8 shown the variations of signal strength values from four access points in a fixed point in a period of 300 seconds. Note that even in a fixed position, the fluctuations of the signal level reaches 10dBm, expected variability for indoors environments.

The estimated location was obtained in two different ways:

.

- Estimated per room: In this case, the algorithms return to the room based on the measured RSS. The table 1 show the results for such an estimate.
- Estimate for coordinated: In this case, the algorithm returns to three-dimensional position of the user. results are summarized in table II and in the the Figures 9 and 10.

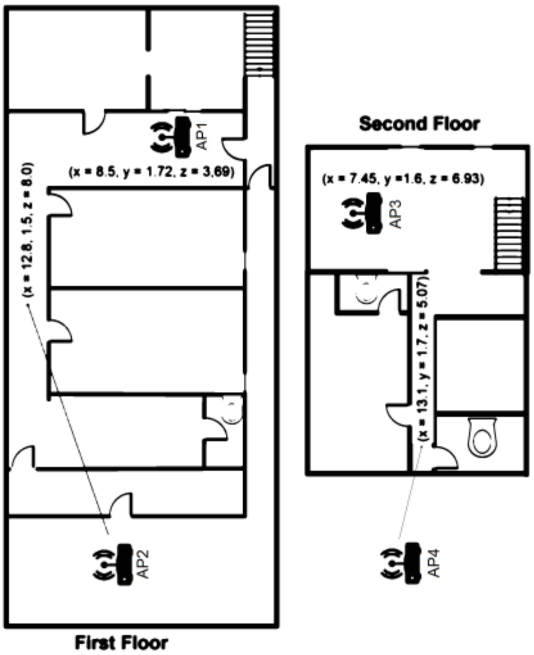

Fig. 6: Floor Plan - The test area used in the experiments

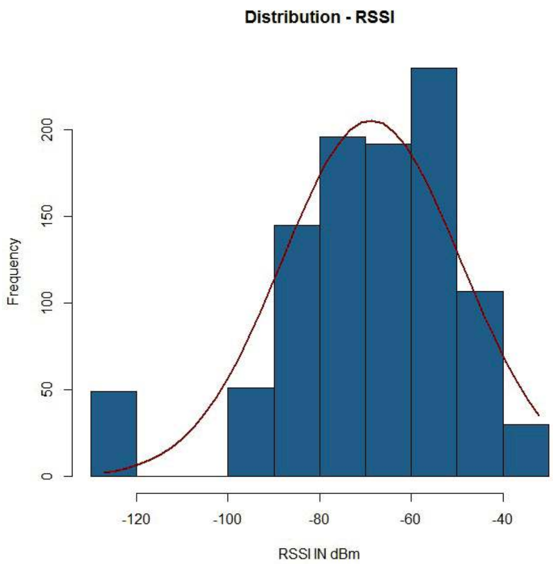

Fig. 7: RSS Histogram

For triangulation technique, we estimated the distance to the target using the RSS. Such technique was not efficient when compared to others methods. Only 13.5% of the estimates showed lower or equal errors to one meter and 65.5% lower or equal to three meters. Similarly, for the estimate for rooms, the triangulation has not obtained a good performance, in this case, the correct identification room occurred in only 73 %.

The K-Nearest Neighbor algorithm was implemented with euclidean distance with K = 2 and 3. The method reached 82% and 83% accuracy in 3D Positioning and precision in the location of the rooms 85% and 89% for k=2 and 3, respectively

The ANN methods was were trained with back-propagation algorithm with 50,100,150,200,250 and 300 neurons in hidden layer. method was slightly better than the KNN method, with 84% and 86 % for MLP and RBF model, respectively (estimated per room) and 69% and 70% for estimated for 3D coordinated, however, with a lower performance than the methods based on Bayesian inference, which were implemented using histogram and kernel methods, with results of 87% and 89% for the estimates with lower or equal errors to three meters and 94% and 95% of precision in the location of the rooms

The proposed algorithm presented better performance than others methods with mean error of 1.86m, precision in the location of the rooms of 97% and 90% the estimates were at maximum three meters away from the actual location.

The above results are summarized in table II and in the the Figures 9 and 10.

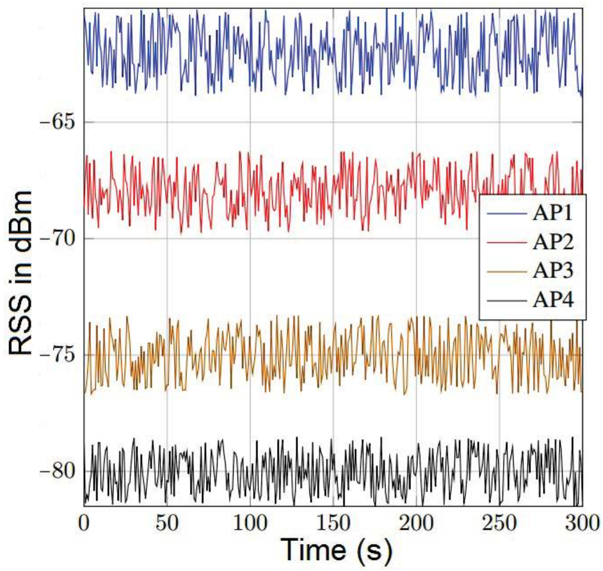

Fig. 8: Variations of RSS values from four AP in a fixed point.

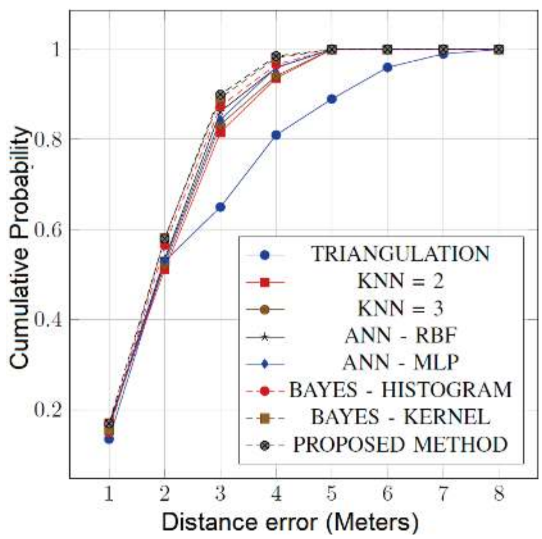

Fig. 9: *Error in Meters - Cumulative*

TABLE I
PRECISION RATE IN LOCATION OF THE ROOMS (%)

AGORITHM	%	AGORITHM	%
Triangulation	73	RNA(RBF)	91
KNN = 2	85	Bayes - Histogram	94
KNN = 3	89	Bayes - Kernel with Gauss	95
RNA(MLP)	90	Proposed Method	97

TABLE II
LOCATION ESTIMATION ERRORS (AVERAGE, COEFFICIENT OF VARIATION, Q1, Q2 AND Q3 QUARTILES IN METERS)

AGORITHM	ME	C_v	Q1	Q2	Q3
Triangulation	2.67	57.33%	1.47	1.94	3.83
KNN = 2	2.07	45.29%	1.37	1.73	2.51
KNN = 3	2.04	44.41%	1.36	1.72	2.43
RNA(MLP)	2.00	42.40%	1.36	1.72	2.38
RNA(RBF)	1.98	42%	1.35	1.71	2.3
Bayes - Histogram	1.94	40.99%	1.35	1.7	2.2
Bayes - Kernel with Gauss	1.89	38.52%	1.34	1.68	2.1
Propose Method	1.86	37.62%	1.33	1.66	2.00

The proposed method has a lower variability than other algorithms presented in this paper. The deviation in relation to the mean reaches 37.62% against 38.52%, 40.99%, 42%, 42.40%, 44.41%, 45.29% and 57.33% for the methods Bayes (Kernel with Gauss), (Bayes - Histogram), RNA (RBF), RNA(MLP), KNN = 3, KNN = 2 and triangulation, respectively, which shows greater accuracy in estimating the location.

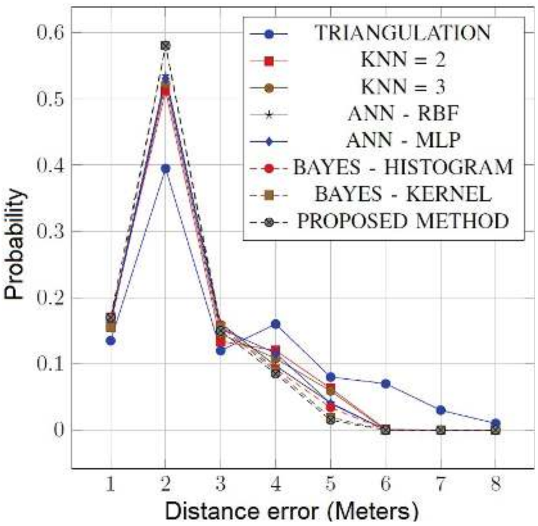

Fig. 10: *Error in Meters - Probability distribution*

IV. CONCLUSIONS

In this paper, we have proposed a hybrid algorithm based on Bayes inference to locate people and objects in 3D indoor environments. with exception the triangulation method, all techniques have made use of a structure based on pattern recognition, called fingerprinting.

The proposed solution in this work presented better performance than others algorithms. The proposed algorithm presented better performance than others methodswith mean error of 1.86m, precision in the location of the rooms of 97% and 90% the estimates were at maximum three meters away from the actual location.

As future work, we intend to implement to a method based on angulation and other hybrid algorithms, in order to maximize the accuracy of the estimated location, and finally comparing the building algorithms in other models.

REFERENCES

[1] Hui, L. and Houshang, D. and Banerjee, P. and Jing. L. *Survey of Wireless Indoor Positioning Techniques and Systems* IEEE Transactions on System, Man, And Cybernetics-Part C: Applications and Reviews, Vol. 37, NO. 6, 45-61, November 2007.
[2] Sindhuja, S. and Ilakiya, P. *Fundamentals of Data Mining in Genomics and Proteomics* Journal of Emerging Technologies and Innovative Research (JETIR), Vol. 2, Issue: 2, February 2015.
[3] Chih-Yung Tsai, Shuo-Yan Chou , Shih-Wei Lin, Wei-Hao Wang *Location determination of mobile devices for an indoor WLAN application using a neural network* Knowledge and Information Systems, Vol. 20, Issue: 1, pages 81-93, Jully 2009.
[4] Vilovic, I. and Burum, N. *A Comparison of MLP and RBF Neural Network Architectures for Location Determination in Indoor Environments* 2013 7th European Conference on Antennas and Propagation (EuCAP), Publisher: IEEE, pages 3496-3499, April 2013.
[5] Nerguizian, C. and Despins, C. and Affes S. *Geolocation in mines with an impulse response fingerprinting technique and neural networks* IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS,Vol. 5, NO. 3, pages 603-611, March 2006.
[6] Hsu-Yang Kung and Sumalee Chaisit and Nguyen Thi Mai Phuong. *Optimization of an RFID location identification scheme based on the neural network* International Journal of Communication Systems, Vol. 28, Issue: 4, pages 625-644, March 2015.
[7] Feng Yu and Minghua Jiang and Jing Liang and Xiao Qin and Ming Hu and Tao Peng and Xinrong Hu *5G WiFi Signal-Based Indoor Localization System Using Cluster K-Nearest Neighbor Algorithm* International Journal of Distributed Sensor Networks, Volume 2014 (2014), Article ID 247525, 12 pages.
[8] Feng Yu; Minghua Jiang; Jing Liang; Xiao Qin; Ming Hu; Tao Peng; Xinrong Hu "An indoor localization of WiFi Based on branch-bound algorithm", Information Science, Electronics and Electrical Engineering (ISEEE), 2014 International Conference on, On page(s): 1306 - 1308 Volume: 2, 26-28 April 2014
[9] Roos, T. and Myllymäki, P. and Tirri, H. and Misikangas. P. and Sievänen J. *A Probabilistic Approach to WLAN User Location Estimation* International Journal of Wireless Information Networks, pages 45-61, Vol. 09, 2002.
[10] Honkavirta, V. and Perälä, T. and Ali-Löytty, S. and Piché. R. *A Comparative Survey of WLAN Location Fingerprinting Methods* Proc. of the 6th Workshop on Positioning, Navigation and Communication, pages 243-251, 2009.
[11] Maja STELLA, Mladen RUSSO, Dinko BEGUŠIĆ. *RF Localization in Indoor Environment* RADIO ENGINEERING, VOL. 21, NO. 2, JUNE 2012.
[12] Dong, B. and Xing, J. and Wu, F. and Zhou. Y. *Probabilistic Algorithm based on Fuzzy Clustering for Indoor Location in Fingerprinting Positioning Method* International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 6, No. 8, 2015.
[13] Abdullah Al-Ahmadi and Tharek Abd. Rahman (2012). One Stage Indoor Location Determination Systems, New Approach of Indoor and Outdoor Localization Systems, Dr. Fouzia Elbahhar (Ed.), ISBN: 978-953-51-0775-0, InTech, DOI: 10.5772/50522. Available from: http://www.intechopen.com/books/new-approach-of-indoor-and-outdoor-localization-systems/one-stage-indoor-location-determination-systems
[14] Axel, K. [2005]. Location-Based Services: Fundamentals and Operation, John Wiley & Sons, Hoboken, NJ, USA.
[15] M. Hauskrecht, R. Pelikan, M. Valko, J. Lyons-Weiler. Feature selection and dimensionality reduction in genomics and proteomics. In Berrar, Dubitzky, Granzow, eds. Fundamentals of Data Mining in Genomics and Proteomics , Springer, pages 149–172, Fall 2006.
[16] Anil, K.J. and Jianchang, M. and Mohiuddin, K.M. *Artificial neural networks: a tutorial* IEEE Computer Society, Vol. 29, Issue: 3, pages 31-44, March 1996.
[17] Vistumbler, http://www.vistumbler.net/. Accessed 16 Jan 2016.