Finger Printing using Applications IoT for Localization Indoor

M. Jothi Lakshmi M.E (CSE
S. Dhanabal
Asp/Head of CSE,
PGP College of Engineering and Technology ,Namakkal.

Abstract-- **Fingerprinting method is one of the preferred method used for indoor localization using Wi-Fi signals because of its low complexity and its cost effectiveness. This paper proposes an indoor localization algorithm using fingerprinting method that is suitable for an indoor IoT application. The proposed algorithm combines the location estimates from two different approaches, deterministic and probabilistic, to estimate the target location. The proposed algorithm was tested for different conditions: stationary and moving IoT targets, line-of- sight and non-line-of-sight indoor environments. The results showed the proposed combined algorithm performed better in terms of localization accuracy, precision and robustness than deterministic and probabilistic methods individually and similar past research.**

Keywords: *IoT, indoor localization, fingerprinting*

I. INTRODUCTION

Indoor localization or indoor positioning is a key enabling technology for IoT applications [1] such as guiding customers or visitors inside a shopping mall or a convention centre, where conventional navigation technologies such as GPS is not available.[2],[3] Even though there are indoor localization solutions that use RFID or BLE beacons with known fixed locations inside a building, it requires additional hardware and installation costs thus making the implementation of these systems costly in terms of time and money. However, using Wi-Fi signals to perform localization makes it a better alternative to the beacon based systems as it does not require the installation of new hardware, thus reducing complexity and cost of the system [4]. In the literature, some research have focused on the study of RF signal propagation in indoor environments while others have developed localization methods that exploit various aspects of RF signal propagation such as propagation time, angle of arrival and received signal strength (RSS) to achieve localization. Methods such as Time of Arrival (TOA) and Time Difference of Arrival (TDOA) use propagation time for localization. These methods require time synchronisation between target device and measuring stations and between measuring stations respectively. Methods that use propagation angle, such as Angle of Arrival (AOA) require measuring stations to have special antenna arrangements in different orientations. In addition to being complex, these systems suffer reduced performance due to propagation time and angle are directly affected by multipath effect existing in indoor environments. Methods using RSS however provide a better alternative. Fingerprinting method uses RSS measurements and is less complex in implementation that it does not require special hardware or the access point locations. It can be implemented in software reducing costs [4]. Performance of localization algorithms are quantified using its accuracy, precision and robustness. Accuracy is a measure of how much the result has deviated from the expected outcome, while precision is a measure of how consistently the result is within a certain value range. Robustness is how well the algorithm perform under poor Radio Frequency (RF) conditions [5]. Past research have resulted in precisions in the range of 90%, but it is also important to know the error value considered when calculating the precision. For example, in [6] the precision values for different algorithms are shown in Table I.

Table I. Precisions of different algorithms in [9]

Method	**Accuracy (m)**	**Precision (< 2m)**	**Precision (< 1m)**
Deterministic	1.6	90%	9%
Probabilistic	1.87	70%	30%
Combined	1.54	65%	30%

As seen above, the precision is 90% when 2m is considered for deterministic (KNN) method while it drops to 9% when
1m is considered. The same applies for Probabilistic and
Combined methods. By comparison, the work in [7] proposed algorithm achieved only 50% precision below 2m error but achieved 30% precision when 1m error was considered. In [8], the accuracy and precision of some commercial indoor localisation solutions are compared. Table II illustrates some solutions using WiFi RSS for localization.

Table II. Commercial products and their performance [8]

Solution	Algorithm	Accuracy	Precision
Microsoft RADAR	KNN	3-5m	50% within 2.5m 90% within 5.9m
Horus	probabilistic	2m	90% within 2.1m
DIT	MLP	3m	90% within 5.12m
MultiLoc	SMP	2.7m	50% within 2.7m

Above table shows that the accuracy and precision of some commercial systems are relatively low. Some solutions, EIRIS and Ubisense, provided accuracy below 1m, but their robustness was lower [8]. After considering different existing solutions and past research the importance of achieving high accuracy, precision along with robustness was identified. For this purpose, this paper uses fingerprinting method in the proposed algorithm.

A. *Fingerprinting method*

Fingerprinting method is one of the most used method for localization because of its above-mentioned benefits. Fingerprinting method involves storing the RF characteristics, known as fingerprints, of locations of the indoor environment in a database and comparing the fingerprint of the unknown target location with the fingerprints in the database to find an approximated location of the target [2]. As such fingerprinting method is composed of two phases:

1) *Offline Phase*

This phase is also called the Data Collection phase during which, the fingerprints of the concerned indoor area are collected and the database is created. The indoor area is divided into an equally spaced grid where the grid points are called reference points (RP), at which the data will be collected. Past studies have showed that multipath effect, reflection, diffraction and scattering cause RSS to randomly vary around a mean value at a location[9]. RSS value is also affected by fading, which consists of two parts, Large-scale fading and small-scale fading. Large-scale fading is caused by attenuation due to signals being absorbed by various materials and objects in the environment. Large-scale fading decides the mean RSS. Small-scale fading is caused by multipath effect. As such, RSS in an indoor environment can be approximated to a Gaussian distribution with a mean and a standard deviation. For a more accurate approximation of mean and standard deviation, large number of samples will need to be collected at each RP. In literature, number of samples collected were as high as 10,000 [7]. After collecting samples the calculated mean and standard deviation will be part of the fingerprint of that RP [9], [10].

2) *Online phase*

In the online phase the algorithm takes a sample fingerprint from the unknown target location and compares it with the fingerprints in the database to classify the RPs that are most likely (or closest) to the target location. There are several known algorithms such as probabilistic, k-Nearest- Neighbour (KNN), neural networks, support vector machine (SVM) etc. This work uses the probabilistic and KNN methods to find two sets of estimated coordinates and finally combine them [8].

In this paper, Section II describes the proposed fingerprinting method in detail. Section III discusses the results and observations of the testing of the algorithm. Finally, Section IV provides the conclusion of this paper.

II. FINGERPRINTING ALGORITHM

This paper implements the fingerprinting algorithm based on past research [6] making modifications with the aim of improving performance in terms of precision, accuracy and robustness. Two algorithms were designed to perform tasks in each phase of the fingerprinting method.

A. *Data collection algorithm*

Figure 1 shows the flowchart of the proposed data collection software used in during the data collection phase. As seen in the flowchart the data collection will be performed for ‘s’ number of times at a particular RP. In this paper 100 is chosen, as the number of samples, due to practical reasons, but larger values will give a better representation of the RF behaviour at the RP. When collecting data at the RP, firstly the Wi-Fi signals will be scanned to obtain the list of available Wi-Fi access points (Cells) and their information such as signal level, signal quality, modulation and MAC address. The second block in the flowchart represent the process of extracting the required information from the list. The list will include Wi-Fi signals from other buildings, but only those from the required building needs to be filtered. There after MAC address and RSS of each cell will be extracted and the total number of times a MAC address (i.e. access point) was received and the total RSS will be saved. When the measurement is done for all ‘s’ number of times, the final fingerprint for the RP will be created by calculating the mean and standard deviation of RSS for each MAC address received and it will them be saved to a log-file.

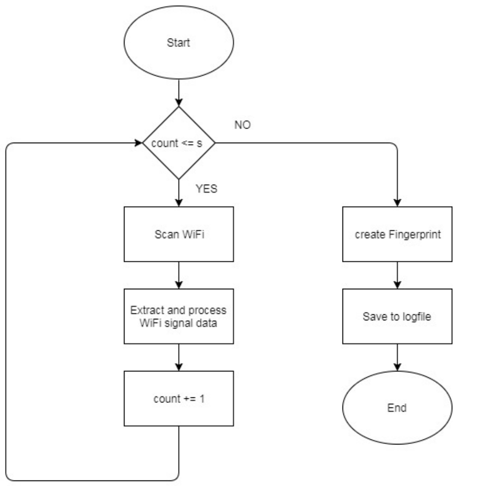

Fig. 1. Data collection software flowchart

After data has been collected for all RP, the fingerprint database (FPDB) can be created using the log-files of each RP. The FPDB consists of three parts named FPDB1, FPDB2 and FPDB3. FPDB1 contains all RP and the list of MAC addresses received at each RP during data collection phase. FPDB2 contains the fingerprints for each RP. Each fingerprint consists of RSS statistics received for all MAC at each RP. The RSS statistics include mean RSS, standard deviation and unique RSS values received during the measurement period and their frequencies. FPDB3 contains coordinates of each RP. The FPDB will be used as an input to the localization algorithm during the online phase, which will be explained next.

B. Localization algorithm

Localization algorithm is executed during the online phase. The localization algorithm proposed in this paper takes five rapid samples at the beginning to create the *'sample'* fingerprint of the unknown target location. This sample fingerprint is said to be of size N, meaning it contains N number of MAC addresses received at the target location and the average RSS of each MAC received during the sampling period. This sampling process allows to get a better representation of RSS, reducing the effect of RSS fluctuation caused by fading. Then the *sample* fingerprint is sent through an above average filter, where MAC addresses whose RSS is higher than the average RSS of the *sample* are selected to create the *'sample_n'* fingerprint of size *n* where n<N. The resulting *sample_n* MAC address list is then matched against the FPDB1 in the pre-match phase. In the pre-match phase RP that contain all the MAC in the *sample_n* are selected from FPDB1 to create the *prematch* set of RP. This reduces the number of RP to m<M where M is the number of RP in the test area. The *prematch* set of RP is then used to calculate the target location using two different methods, deterministic method and probabilistic method. The difference between the two methods is that in deterministic method RPs that are closest to the target location are found while in probabilistic method RPs that are most probable to be the target location are found.

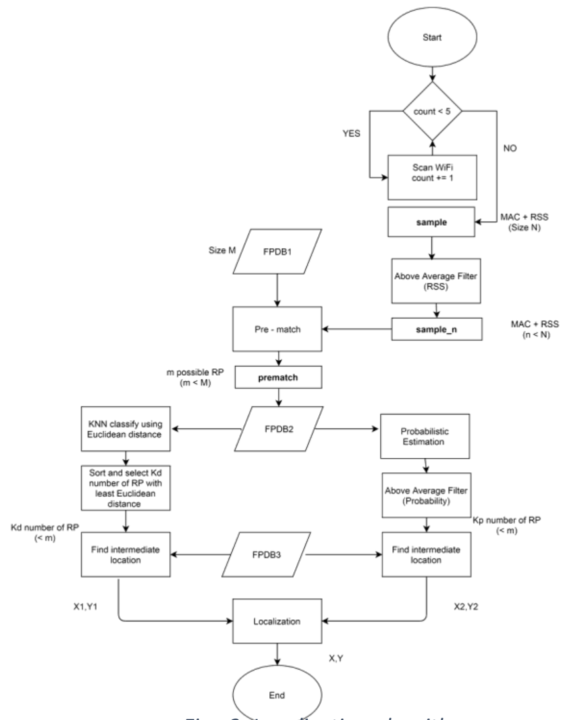

Fig. 2. Localization algorithm

i. Deterministic method

This method calculates the Euclidian distance between the *sample* and each RP in the *prematch*.

[illegible] (1)

In Eq. 1, [illegible] and [illegible] are RSS values of ith MAC in *sample_n* and corresponding fingerprint in FPDB2 where i=1,2,3…n. The Euclidian distance [illegible] is found for all RP in *prematch* where j=1,2,3….m. Then from the prematch, K_d number of RPs that have the lowest Euclidean distance are selected. The value of K_d that gives the best performance must be experimentally found beforehand. [6]

[illegible]

of being the target location. The conditional probability of the ith RP is found using Bayes' rule [11] as shown in Eq. 3.

$$P(RP_i|S) = \frac{P(S|RP_i) * P(RP_i)}{\sum_{j=1}^{m} P(S|RP_j) * P(RP_j)} \quad (3)$$

Where P(RP) is the prior probability of the target being at a given RP. This value depends on various factors such as user speed, user movement patterns but here it is assumed that each RP in the *prematch* is equally probable making P(RP) = 1/m. $P(S|RP_i)$ is the likelihood of the *sample_n* (i.e. S) occurring at the i^{th} RP. Value of $P(S|RP_i)$ is given by Eq. 4.

$$P(S|RP_i) = P(RSS_1|RP_i) P(RSS_2|RP_i) \ldots P(RSS_n|RP_i) \quad (4)$$

Where i=1,2,…m and $P(RSS_n|RP_i)$ are Gaussian probabilities of RSS of i^{th} MAC address, $P(RSS_n|RP_i)$, in the *sample_n* modelled by Eq. 5.

$$P(RSS_i|RP_j) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (5)$$

Where x, μ and σ are the RSS of the i^{th} MAC in *sample_n*, the mean RSS of the MAC in RP and the standard deviation of RSS of the i^{th} MAC. After calculating $P(RP_i|S)$ for each RP in *prematch*, the K_p number of RPs with the highest probabilities will be selected. Using this, the intermediate coordinates of target, (x_p, y_p) will be found using the following equation: [6]

$$(x_p, y_p) = \frac{\sum_{j=1}^{K_p} w_j * (x_j, y_j)}{\sum_{j=1}^{K_p} w_j} \quad (6)$$

Where the weight w=p and (x_j, y_j), the coordinates of each j^{th} RP in the *prematch* set, will be retrieved from FPDB3. After (x_d, y_d) and (x_p, y_p) are found the two results will be combined as shown in Eq. 7. to get the final estimated coordinate (X,Y) of the target location.

$$(X, Y) = \frac{\frac{1}{E1}}{\frac{1}{E1}+\frac{1}{E2}}(x_d, y_d) + \frac{\frac{1}{E2}}{\frac{1}{E1}+\frac{1}{E2}}(x_p, y_p) \quad (7)$$

Where E1, E2 are errors of (x_d, y_d) and (x_p, y_p) with respect to the test location (x_t, y_t) respectively. E1 and E2 are found as in Eq. 8 and Eq. 9 respectively.

$$E1 = \sqrt{(x_t - x_d)^2 + (y_t - y_d)^2} \quad (8)$$

$$E2 = \sqrt{(x_t - x_p)^2 + (y_t - y_p)^2} \quad (9)$$

III. TESTS, RESULTS AND OBSERVATIONS

This section discusses tests performed to measure the

Thereafter Weighted K-Nearest Neighbour (WKNN) algorithm is used to calculate the intermediate coordinates (x_d, y_d) of the target as in Eq. 2 where $w_i = 1/d_i$ for the i^{th} RP with the lowest distance. The (x_i, y_i) values are retrieved from FPDB3.

ii. *Probabilistic method*

The main idea behind the probabilistic method is to find the RPs in the *prematch* set, which have the highest probability

A. *Line of Sight (LOS) scenario*

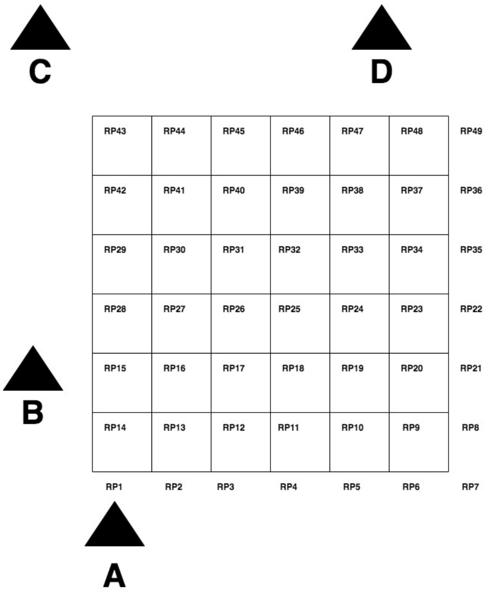

Fig. 3. Test area layout for LOS scenario

The test area is located in an open area with five access points having clear LOS with the entire test area. Figure 3 shows the test area with the approximate location of access points. (Note that the exact locations of access points are irrelevant when using fingerprinting method)Test parameters are shown in Table below.

Table III. Stationary test parameters for LOS scenario

Parameter	Value
RF propagation	LOS
No of RP	49
Area size	6m x 6m
Origin	RP1
X-axis direction	RP1→RP7
Y-axis direction	RP1→RP43
Test location (xt,yt)	(3.5, 3.5)
User speed	N/A
Readings	100
K_d and K_p	K

Access points A and D have clear LOS with entire test area while B and C have obstructions to parts of the area. Test results are shown in Table IV.

Table IV. Results for different K in LOS scenario

K	Deterministic		Probabilistic		Combined	
	Error (m)	Pre (%)	Error (m)	Pre (%)	Error (m)	Pre (%)
3	1.5546	9	2.1247	6	0.4757	88
4	1.5613	1	1.4414	24	0.5383	84
5	1.5784	16	1.5851	29	1.1923	46
6	1.3382	19	1.3812	21	0.9635	49
7	1.7811	5	1.5638	34	1.4347	31

The combined method has an improved the accuracy and precision when compared to the individual methods. The maximum precision of 88% for error below 0.9m and lowest error of 0.4757m were observed by the combined method for K=3. The results of the proposed method for K=3 is shown in Fig. 6.

As seen in Fig. 4 the results are mostly clustered within 1m radius of the test location (3.5, 3.5). A moving test was performed to track the moving target with the following test parameters:

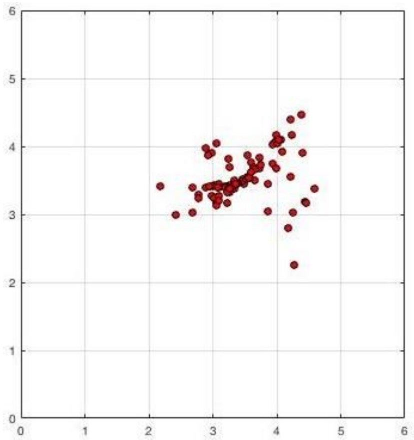

Fig. 4. Stationary test K=3 using combined method for LOS scenario

Table V. Moving test parameters for LOS scenario

Parameter	Value
RF propagation	LOS
No of RP	49
Area size	6m x 6m
Origin	RP1
X-axis direction	RP1→RP7
Y-axis direction	RP1→RP43
Between points	(3.5,0) ←→ (3.5,5)
User speed	0.31 m/s
Readings	100
K_d and K_p	3

The apparatus was moved in a straight line, back and forth, slowly at a speed of 0.31 m/s for the 100 readings. A low speed was selected to simulate IoT application where a customer walks in a shopping mall. As seen in Fig. 5, the resulting points from the proposed combined method are mostly above 0.5 m from the actual path of the target, but comparatively, the combined method has more points that are closer to the actual path than the other two method.

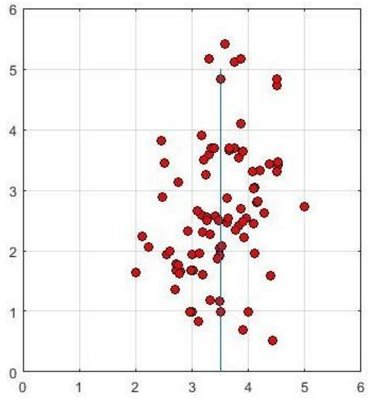

Fig. 5. Moving test with K=3 using combined method for LOS scenario

B. Non Line of sight scenario

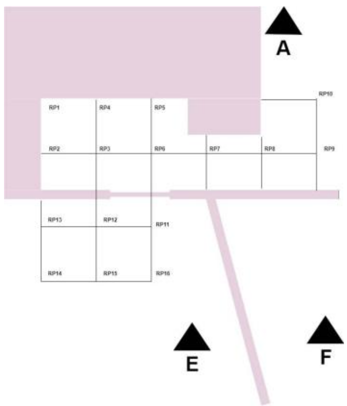

Fig. 6. Test area layout for non-LOS scenario

-

To test the robustness of the algorithm, it was tested under non-LOS conditions. For this, a room, which is located at the end of a narrow corridor, was selected as the test area. Fig. 6 shows this area with the locations of the nearby access points. The test location was chosen as (2.6, 2) such that it does not have LOS from any of the access points.

Table VI. Stationary test parameters for non-LOS scenario

Parameter	Value
RF propagation	Non-LOS
No of RP	16
Area size	5m x 3m
Origin	RP14
X-axis direction	RP14→RP16
Y-axis direction	RP14→RP1
Test location (xt,yt)	(2.6, 2)
User speed	N/A
Readings	100
K_d and K_p	K

The test results in Table VII shows that the proposed algorithm performs better in terms of both accuracy and precision over the other two individual methods.

Table VII. Results for different K in non-LOS scenario

K	Distance		Probabilistic		Combined	
	Error (m)	Pre (%)	Error (m)	Pre (%)	Error (m)	Pre (%)
3	0.6745	73	0.7814	73	0.5362	91
4	0.6916	68	0.9521	29	0.6019	87
5	0.703	91	0.7676	77	0.5678	95
6	0.6282	82	0.6845	99	0.4968	99
7	0.5939	87	0.471	98	0.4025	99

The lowest average error of 0.4025m and highest precision of 99% for error below 0.9m were obtained by the combined method when K=7, and the results are illustrated in Fig. 7.
In Fig. 7, it can be seen how the results of combined method are clustered closer together. This implies that the precision of the combined method is higher as seen in Table VII. A moving test was performed with the test parameters shown in the following Table VIII.

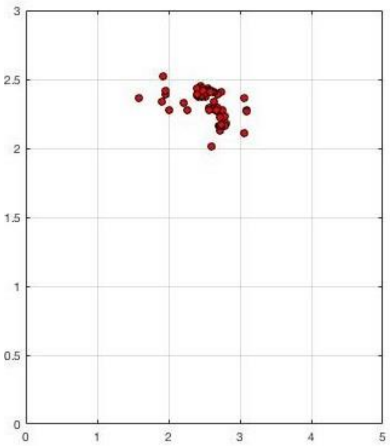

Fig. 7. Stationary test with K=7 using combined method for non-LOS scenario

Table VIII. Moving test parameters for non-LOS conditions

Parameter	Value
RF propagation	Non-LOS
No of RP	16
Area size	5m x 3m
Origin	RP14
X-axis direction	RP14→RP16
Y-axis direction	RP14→RP1
Between points	(0,2.2) ←→ (4.5,2.2)
User speed	0.29 m/s
Readings	100
K_d and K_p	7

As in previous section, the apparatus was moved in a straight line back and forth for the 100 readings. The results are shown in Fig 8.

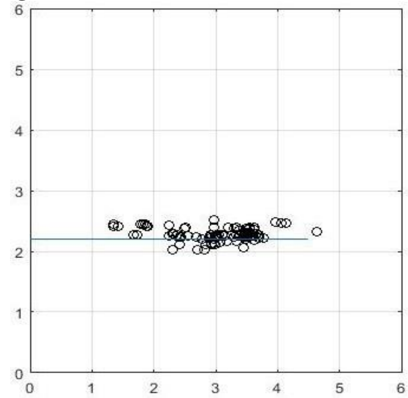

Fig. 8. Moving test (K=7) using deterministic method for non-LOS

Unlike in the LOS scenario, in the non-LOS case the algorithm tracks the target more precisely along the actual path. All three methods provide target location within 0.5m of the actual path. However, only the deterministic method tracked the target along the path for more distance than the other two methods whose results were concentrated. Further coordinates were not received when the target was near the wall at location (0, 2.2).

C. Observations

After the tests, the accuracy and error results can be used to compare the combined method's performance in terms of accuracy and average error with those of deterministic and abilistic methods.

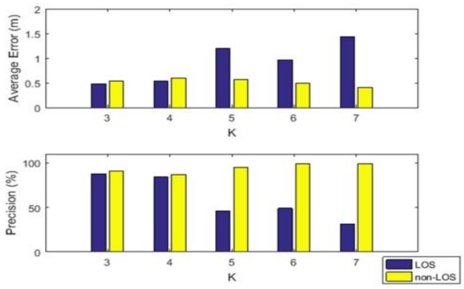

Fig. 9. Accuracy and Precision comparison of the combined method for LOS and non-LOS scenarios for combined method

As seen in Fig. 9, when K<5, the combined method performs with an average error <0.7m and precision >84% for both LOS and non-LOS conditions. For the same conditions, the deterministic and probabilistic methods have a higher accuracy and lower precision than the combined method as seen in Fig. 10 and Fig. 11 respectively.

For all K under LOS conditions, the performance of the combined method is better than the deterministic and probabilistic methods. However K<5 values provide the best performance for all situations.

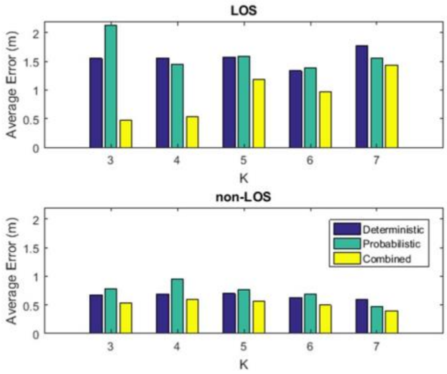

Fig. 10. Accuracy comparison of deterministic, probabilistic and combined methods for LOS and non-LOS scenario

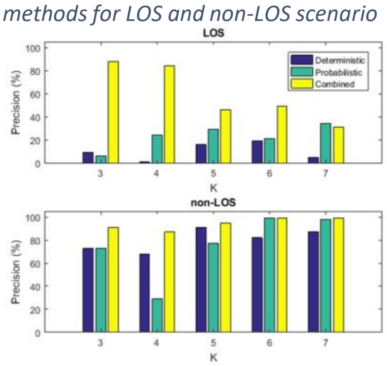

Fig. 11. Precision comparison of deterministic, probabilistic and combined methods for LOS and non-LOS scenarios

IV. CONCLUSIONS

In conclusion, for a stationary target device, the proposed combined algorithm achieved a maximum precision of 88% under LOS conditions with K=3 and a maximum precision of 99% was achieved under non-LOS conditions was with K=6 and K=7. Accuracy of the proposed algorithm remained stable around 0.5m for all K under non-LOS conditions, while it degraded with increasing K for LOS conditions. In addition, the proposed fingerprinting algorithm with combined method achieved a 91% precision and accuracy of less than 1m when K=3 and K=4 for both LOS and non-LOS conditions. Therefore, it was concluded that the proposed algorithm can be used for localization under any RF conditions using K<=4 with satisfactory overall performance with high accuracy, precision and robustness. Overall, the combined method performed better than both deterministic and probabilistic methods for all situations. For a moving target, the algorithm, using deterministic method, performed better under non-LOS conditions by tracking it with less deviation from the actual path. However, further research needed to be done to track a mobile target precisely along its path in an indoor environment

REFERENCES

[1] D. Macagnano, G. Destino, and G. Abreu, "Indoor positioning: A key enabling technology for IoT applications," in *Internet of Things (WF-IoT), 2014 IEEE World Forum on*, 2014, pp. 117-118: IEEE.

[2] P. Prasithsangaree, P. Krishnamurthy, and P. Chrysanthis, "On indoor position location with wireless LANs," in *Personal, Indoor and Mobile Radio Communications, 2002. The 13th IEEE International Symposium on*, 2002, vol. 2, pp. 720-724: IEEE.

[3] K. Kaemarungsi and P. Krishnamurthy, "Modeling of indoor positioning systems based on location fingerprinting," in *INFOCOM 2004. Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies*, 2004, vol. 2, pp. 1012-1022: IEEE.

[4] Z. Farid, R. Nordin, and M. Ismail, "Recent advances in wireless indoor localization techniques and system," *Journal of Computer Networks and Communications*, vol. 2013, 2013.

[5] C. Rizos, A. G. Dempster, B. Li, and J. Salter, "Indoor positioning techniques based on wireless LAN," 2007.

[6] R. Ma, Q. Guo, C. Hu, and J. Xue, "An improved WiFi indoor positioning algorithm by weighted fusion," *Sensors*, vol. 15, no. 9, pp. 21824-21843, 2015.

[7] X. Song, F. Yang, L. Ding, and L. Qian, "Weight adjust algorithm in indoor fingerprint localization," in *Signal Processing and Communication Systems (ICSPCS), 2012 6th International Conference on*, 2012, pp. 1-5: IEEE.

[8] H. Liu, H. Darabi, P. Banerjee, and J. 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, pp. 1067-1080, 2007.

[9] K. Kaemarungsi and P. Krishnamurthy, "Properties of indoor received signal strength for WLAN location fingerprinting," in *Mobile and Ubiquitous Systems: Networking and Services, 2004. MOBIQUITOUS 2004. The First Annual International Conference on*, 2004, pp. 14-23: IEEE.

[10] A. Bose and C. H. Foh, "A practical path loss model for indoor WiFi positioning enhancement," in *Information, Communications & Signal Processing, 2007 6th International Conference on*, 2007, pp. 1-5: IEEE.

[11] V. Patmanathan, "Area localization using wlan," *KTH Electrical Engineering*, 2006.