Machine Learning-based Advanced Localization Method in Wireless Communication

San Hlaing Myint
Graduate School of Fundamental Science and Engineering
Waseda University
Tokyo, Japan
shm2007@gmail.com

Takuro Sato
Graduate School of Fundamental Science and Engineering
Waseda University
Tokyo, Japan
t-sato@waseda.jp

***Abstract*—In existing localization research area, most of the localization methods suffer from propagation loss because of multi-path effects and sensitivity of the components of wireless technology, e.g. Received Signal Strength Indicator (RSSI). It is leading to miss estimation of localization methods and degradation of estimation accuracy. In this paper, a new advance localization method is proposed in different point of view. K-nearest neighbor (KNN) algorithm is adopted to get the best estimation accuracy and RSSI is used as main feature for estimating unknown position of mobile user. We extend to estimate a circle region instead of one coordinate position with the aid of circle theory as a new idea. The objective of this research is to improve the accuracy of localization methods and to mitigate the sensitivity of the existing methods. In this study, an advance localization method is proposed by considering existing finger printing method from a different point of view. In addition, Radius value optimization problem is investigated to be able to analyze the effects of different radius values on localization accuracy. On the other hand, the sensitivity of estimation accuracy of proposed method is analyzed in terms of different norm functions.**

Keywords—Fingerprint localization; machine learning; wireless communication

I. Introduction

Recent advance on location-aware wireless communication systems is leading to integration of heterogeneous technology by improving the advantages and mitigating the drawback of existing technologies to ensure high accuracy and efficiency in all possible scenarios, such as, Line-of-Sight (LOS) or Non-Line-of-Sight (NLOS) of multipath propagation in indoor and outdoor environments [2]. Generally, localization methods can be divided into two main groups, geographic information based localization method and fingerprinting localization method which is based on machine learning classifiers.

The geometric localization methods are developed for improving localization accuracy in outdoor or suburban area by computing the desired location based on geometric information and classical geometric principles. Triangulation, trilateration and multilateration are significant in both GPS (Global Positioning System) and LPS (Local Positioning System) [11], [12]. However, there are some drawbacks of this technology. It needs to use several transmitters to enable a receiver to calculate its geographical position of Mobile User (MU). Another important difficulty is the correctness in measuring the distance between MU and Access Point (AP). It might be correct in Line of Sight (Direct Path) situation and might not be correct in complicated indoor signal propagation (NLOS) situation. Because the distance is computed from RSSI in Trilateration approach, it has been suggested to add a learning procedure that means collecting some RSSI data at some points with the known coordinates to solve the variation of RSS depending on environmental changes [5].

On the other hand, with the development of next generation wireless communications technology, machine learning and artificial intelligence become promising research trends for smart and intelligent localization system. The estimation of the accuracy of fingerprinting localization methods relies on a pre-constructed database (e.g., a radio map) [1]. These databases include collected values of Localization Measurement Unit (LMU), e.g. RSSI, Angle of Arrival (AOA), Time of Arrival (TOA) and so on [13]. The robustness of these recorded values is controlled by the capturing time, sensitivity of environmental situation and sensitive properties of wireless components. Thus, some additional analysis on their representation by using data standardization methods such as standard deviation, mean and variance are necessary to find data granularity of variables.

The existing localization technologies emphasize on the estimation of a certain position of unknown MU by applying different methods. We would like to highlight two problems. The first one is the difficultly of estimation in crowded obstruction area because there are so many obstructive objects and multipath interferences. In this situation, it is very hard to define the exact position of unknown MU. The next problem is that it is very hard to estimate the exact position of moving objects, e.g. driving cars, because of their speed.

In order to mediate the above problems, we propose a new advance localization method by considering different points of view. K-nearest neighbor (KNN) algorithm is adopted to provide the best estimation accuracy and RSSI is used as a main feature for estimating unknown position of mobile user. As a new idea, we approximate the position of unknown MU by examining a circle region applying circle theory in which the unknown MU might be situated rather than a coordinate point.

The steps of our proposed advance localization method are summarized as follows:

- Fingerprinting localization using K-Nearest Neighbor (KNN) classifier.
- Drawing the desired circle region by using parametric equations.
- Investigation of radius value optimization problem because radius is the key parameter in estimating the circle area.
- Controlling the locality of the circle region by varying the radius R value. The larger R value, the more localization accuracy.
- Calculating a simple radius of curvature by observing different degrees of curvature to find the optimized value of R for estimating circle region. The observation starts with degree of curvature, D= 45^o and gradually increments it by 45^o up to 180^o.

In the following sections, background theory of our study is introduced in Section II. The preprocessing stage for Fingerprinting localization is presented in Section III. Our proposed advance localization method follows in Section IV. Radius values optimization problems are explained in Section V. And, our simulation results and discussion and conclusions of our paper are described in Section VI and Section VII, respectively.

II. Background Theory

A. RSSI-beased Propagation Model

In wireless communication, there are the typical localization measurement units, e.g. RSSI, TOA or TDOA (Time Difference of Arrival), AOA and so on [13]. Among them, receive signal strength (RSS) plays as an important indicator to measure quality of wireless signals and to describe the relationship between power of transmitter and power of receiver. Therefore, the higher the RSSI values indicate the stronger the received signal. RSSI provides channel information and position information for localization and tracking target receiver because RSSI is a function of LMU, RSSI is widely used use because of cost effective and lower complexity [6]. The power of received signal (P_R) of free space propagation model can be express by the following formula:

$$P_R = \frac{P_T G_T G_R \lambda^2}{(4\pi)^2 d^2 L} \qquad (1)$$

Where P_R is received power of signal, P_T is transmitted power; L is the loss factor relates to P_R power to P_T. Antenna gains are expressed by G_T and G_R, and λ is carrier wavelength. In multipath propagation, the power of received signal decreases and some part of received signal is lost along a long signal path distance because of multipath interference and shadowing effect which occur by some obstacles along the path of signal (such as trees and buildings). Log-normal shadow model is an obvious generic model in wireless propagation environment. It extends to Friis Free space model because it can describe the propagation loss for a wide range of environments. The net path loss in dB when moving from distance d_0 to d for an arbitrary distance d > d_0 can be expressed as follows:

$$PL(d)(dB) = PL(d_0) + 10_\alpha \, log\left[\frac{d}{d_0}\right] + \chi_\sigma \qquad (2)$$

Here d_0 is near-earth distance. χ is a normally (Gaussian) distributed random variable (in dB) with standard deviation σ and α is path loss index, variation of α depends on the propagation environment and its values becomes larger when there are more obstacles between the transmitter and receiver.

B. RSSI-based Fingerprinting Localization

Generally, fingerprinting localization methods include two main phases, offline training phase and online localization phase. The main important part of the first offline phase is creating fingerprint database (power map). The construction and maintenance of a sufficient fingerprint database could be laborious and problematic. Since numerous studies have been done in this filed, some of research created their own database by using special calibrated modem with extra software [6] and by installing Android application on their devices and by taking records in their university campus [7], [8] and some of them used the legally published database as a benchmark [10]. In this work, UJIIndoorLoc dataset is used as a benchmark database for our experiment.

In the database, there are two main parts, training dataset and testing dataset. In the training set, RSSI values are recorded at pre-determined reference point (RP), specific coordinate (x, y), from all possible wireless access point (WAP). It contains the set of reference point RP, RPi = {RP_1,RP_2,……,RP_M}, M is the number of reference point in the data set. The training data set is expressed as follows:

RP_i = { (x_i, y_i) , $RSSI_{i1}$, $RSSI_{i2,......,}RSSI_{iN}$} as a form of RSSI Vector,

where (x_i, y_i) is the ith specific coordinate of ith reference point, RPi, here, longitude and latitude, and N is the number of RSSI value between the ith reference point, RPi and the N number of wireless access point, WAPi = {WAP_1, WAP_2 ,….., WAP_N}.

In the testing data set, we have the set of unknown point (UP), coordinates of mobile unites (MU), UPi = {UP_1, UP_2,……, UP_N} and their RSSI value. Therefore, we can express the testing data set as follows:

UP_i = { (x_i, y_i) , $RSSI_{j1}$, $RSSI_{j2,......,}$ $RSSI_{jN}$}

where (x_i, y_i) is the ith specific coordinate of ith reference point, MU_i, here, longitude and latitude, and N is the number of RSSI value between the unknown mobile user, MUi and the N number of wireless access point, WAPi = { WAP_1, WAP_2 ,….., WAP_N }.

In online localization phase, the unknown position of the mobile user can be predicated by calculating the Euclidean distance between the RSSI values of the UPi and RPi, in (3). For each unknown point, the best match searching is performed by calculating Euclidean distance with all reference points from training data set and ranging the distance values by descending order.

$$D_i = \sum_{j=1}^{n} \sqrt{(RSSI_i - RSSI_j)^2} \quad ; D_i=\{D_1, D_2,\ldots, D_M,\} \quad (3)$$

In KNN algorithm, the estimation accuracy depends on K value and the estimation process of unknown position is calculating by KNN positions where appropriate K values are selected by doing some experiment. Coordinate of estimated position is obtained as follows:

$$EP = \frac{\sum_{i=1}^{K} \frac{1}{D_i} RP_i}{\sum_{i=1}^{K} \frac{1}{D_i}}; D_i= [D_1, D_2,\ldots, D_K] \quad (4)$$

Where EP is the coordinate of estimated position, D_i is the distance for all selected reference points and RP_i is the position of selected reference points.

C. Parametric Equation of a Circle

Circle geometry theorem is very simple and familiar with us. Parametric equations are commonly used to build a geometric object such as a curve or a surface. Parametric equations are a set of equations that express a set of quantities as explicit functions of a number of independent variables, known as "parameters". A circle is the set of all points in a plane that are equidistant from a fixed point of the plane called the center of the circle [3]. Therefore, a circle which centered at the origin can be defined as the coordinated of all points that satisfy parametric equation.

$$x = r\cos(t) \quad (5)$$

$$y = r\sin(t) \quad (6)$$

In the above equations, x and y are the coordinate of any point on the circle, r is the radius of the circle and t is the angle of the circle parameter (Fig. 1).

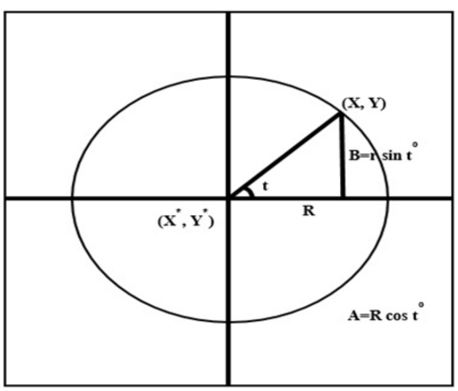

Fig. 1. Drawing a circle by using parametric equations.

III. PREPROCESSING FOR FINGERPRINTING LOCALIZATION

A. Dataset Preparation and Analysis

This database was created by Jaume I University, Spain. The dataset was built with recorded 520 RSSI fingerprints by more than 20 users with 26 different devices in Jaume I University campus. The database covers for 108703m2 including 3 buildings with 4 or 5 floors depending on the building and contains 19937 sampled points have been captured form 520 wireless access points (WAPS), 19938 records for training and 1111 records for testing. There are 259 attributes in this database, 1 to 520 for RSSI level, 521-522 for real-world position, longitude and latitude, and the others for additional information for indoor localization [7], [9].

In this experiment, the additional attributes like building ID or space ID are removed because we want to highlight how the relativity of mobile user position and RSSI Level effects on the estimation accuracy of localization process. Therefore, totally 522 RSSI Level attributes are used for our experiment and additional field is added in the database as a class label for real world position. To investigate the effect of the granularity of the RSSI values, the statistical analysis is performed for determining the appropriateness for use of certain scale of variables, such as mean, standard deviation and variance, on the baseline KNN algorithm. According to the result of analysis, we get the best estimation accuracy on the original RSSI values of dataset as shown in Fig. 2. Therefore, we decided to use the original RSSI values for our experiment.

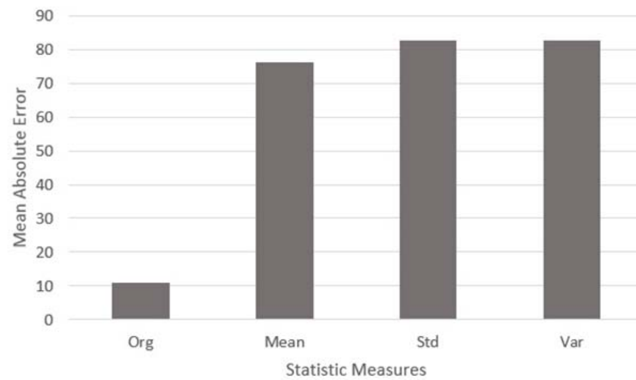

Fig. 2. Estimation accuracy on different data representations.

B. Model Selection

In this research, weka tool is used for model selection process and possible machine learning method are implement for testing with the base line data set. We chose five machine learning algorithms with the lowest estimation error, including K-Nearest Neighbor (kNN), random forest, Random tree, REP tree and bagging, by evaluating with mean absolute error. Experimentation is performed with 10fold cross validation and K factor is 5.

According to the experimentation result, we can see clearly in Fig. 3 that KNN is obviously the best one among others even though the mean absolute error of random forest is as low as KNN. In the existing research [9], researchers performed the model selection with five machine leaning models and showed that KNN is the best candidate for RSSI based localization.

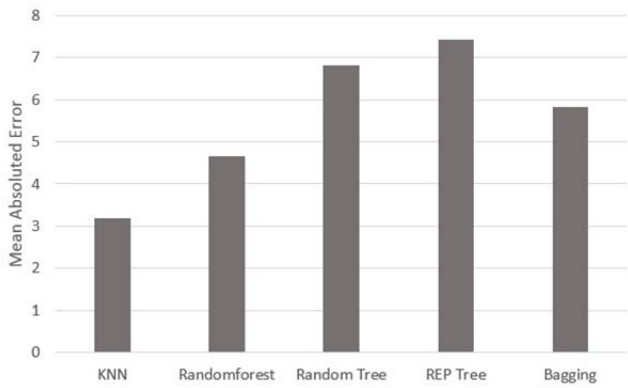

Fig. 3. Machine learning model selection.

IV. PROPOSED ADVANCED FINGERPRINTING LOCALIZATION METHOD

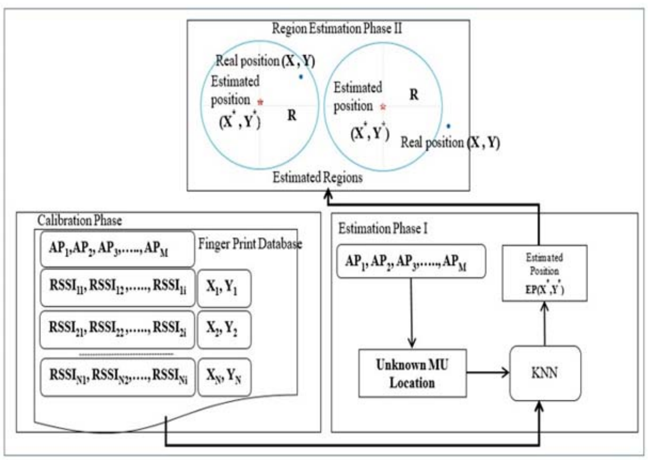

Fig. 4. Proposed two-phases region estimation method.

In this work, we propose a two-phase region estimation method to resolve the highlighted problems. Firstly, fingerprinting-based KNN algorithm is implemented to estimate the unknown user position (X*,Y*). Secondly, we used this estimated position to estimate the desired region by using parametric calculation methods (Fig. 4).

During the first phase, we used the KNN algorithm, which has been used as the baseline classifier in many existing localization methods with a fixed number of nearest neighbors (K). The algorithm calculated to estimate the unknown position is based on the fixed K values. The estimation accuracy increases when K values become larger. KNN algorithm is based on the ranking of the error measures which is the distance between the training data points of RSSI values and the testing data points of RSSI values. According to the KNN algorithm, the position of the unknown MU can be estimated as follows:

$$(X^*, Y^*) = \frac{\sum_{i=1}^{K} \frac{1}{D_i} RP_i}{\sum_{i=1}^{K} \frac{1}{D_i}} \tag{7}$$

In the second phase, in order to estimate the desired region, the parametric calculation is performed by using parametric equations of the circle. Since the estimated position of the MU is defined as the center of the circle, the center of circle is not at the origin. By using the following equations, the center of circle can move from the origin to its proper location.

$$x = X^* + r\cos(t) \tag{8}$$

$$y = Y^* + r\sin(t) \tag{9}$$

where X*and Y* are the coordinates of the proper location, estimated location of the MU.

V. RADIUS VALUES OPTIMIZATION PROBLEM

In this section, we explore the radius values optimization problem which is that the greater value of the radius, the higher the accuracy of estimation. In the proposed methods, the locality of the circle region is controlled by varying the value of radius, R. Therefore, the radius plays a role as a key parameter in the estimation of the desired region.

A. Theory of Curvature

The Radius can easily determine a circle and also a circular curve or arc. The radius of curvature, R, is the reciprocal of the curvature. For a curve, it equals the radius of the circular arc which best approximates the curve at that point. For surfaces, the radius of curvature is the radius of a circle that best fits a normal section [4].

On the other hand, the degree of curvature is defined as the central angle to the ends of an arc of a certain arc length which can normally be defined in different values in different scenarios. The values of angle increasingly change in forward direction to the arc as the value of the arc length is constant. Generally, the degree of curvature can be defined by 360^o as following:

$$\frac{D}{360^o} = \frac{L}{C} \tag{10}$$

Here, D is the degree of curvature, L is the length of arc and C is the circumference. The arc and the circumference have the same length of radius. The circumference of a circle can be expressed as follows:

$$C = 2\pi R \tag{11}$$

By substituting (11) in (10),

$$\frac{D}{360^o} = \frac{L}{2\pi R} \tag{12}$$

Finally, the relationship of the radius of curvature and the degree of curvature can be defined as follows:

$$D = \frac{180^o L}{2\pi R} \tag{13}$$

$$R = \frac{180^o L}{2\pi D} \tag{14}$$

B. Simulation Test-Bed for Radius Value Optimization

In order to find the optimized value of R for estimated circle region; a simple radius of curvature calculation is performed by using curvature theory. According to the relationship between degree and radius of curvature, a simple radius of curvature calculation is performed by observing different degrees of curvature. The observation starts with degree of curvature, D= 45^o and gradually increments it by 45^o up to 360^o.

In American railway standard measurement, the 100 feet (or meter) is commonly used as a standard length of a curve. In this work, 100 millimeters are adopted as a standard length of a curve. And then the observed degree values are sliced by 45^o between the range of minimum angle, we assume that 32^o is minimum, and 180^o. Finally, different observed radius values are calculated over the slices of the degree of curvatures as shown in Fig. 5.

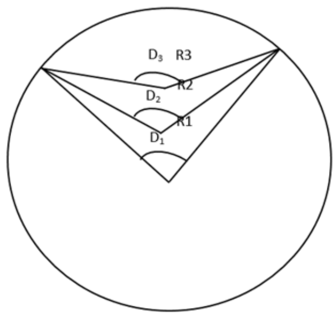

Fig. 5. Relationship between degree curvature and radius of curvature.

C. Evaluation Measurement

In order to measure the estimation accuracy, firstly we need to define the evaluation methods and measure units. Since the proposed localization method estimates the circle region in which the unknown MU can exist, we defined a hit count when the unknown MU is interior of the estimated circle region and a miss count when the unknown MU is exterior of the estimated circle region. In this evaluation, the total hit count and total miss count for all testing data sets are expressed in percentage (hit% and miss%). A circle separates a plane into two sets of points, here, origin of a circle, O (X*, Y*) and the point on the boundary of a circle, A (X, Y). Therefore, the length between O (X*, Y*) and A (X, Y) can be defined as the radius R.

In this scenario, we add third point, the real position of MU, MU (X, Y). The distance between the real point (X, Y) and the origin point, estimated point, O (X*, Y*) can be calculated as follows:

$$d = \sqrt{(X - X^{*})^{2} + (Y - Y^{*})^{2}} \quad (15)$$

For a given circle with radius R, we can set two rules as follows:

1) The interior of a circle is the set of all points whose distance from the center of the circle, d, is less than the length of the radius of the circle, R; (D < R).

2) The exterior of a circle is the set of all points whose distance from the center of the circle, d, is greater than the length of the radius of the circle, R; (D > R).

VI. Simulation Results and Discussion

The simulation of the proposed region estimation was carried out using MATLAB. In Fig. 6, a sample simulation result is demonstrated in a scenario with 2norms (Euclidean) distance function, radius of curvature R=64 and degree of curvature D=90^o by testing over 19937 training data records and 1111 testing data records.

According to simulation results, we show the estimation accuracy in Fig. 7, hit rate of proposed region estimation method over four slices of R values, R= [32,42,64,127] by observing of D values, D= [180^o,135^o,90^o,45^o]. According to the estimation results, estimation accuracy is constantly remained over 90% and the best estimation accuracy is significantly over 98% when the radius value R is 127 and degree value D is 45^o. Finally, we can approve that the proposed method can reduce the mean absolute error than existing KNN based fingerprint localization method as shown in Fig. 8.

On the other hand, we need to investigate the impact of various distance function (different norms) on the estimation accuracy of KNN based proposed method because the algorithm is built on measuring distance between vectors and it decides final output by ranging distance vectors. The relationship of proposed method and different norms (different functions) will be discussed in next section.

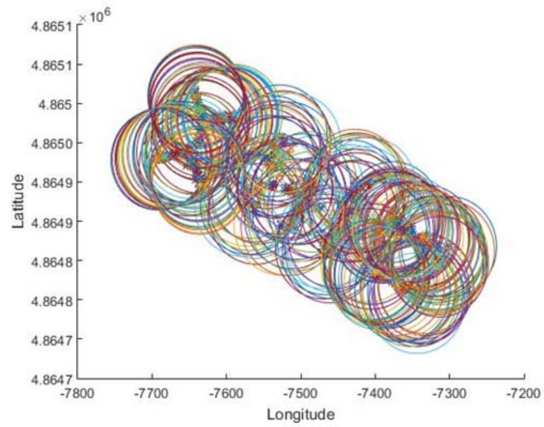

Fig. 6. Sample simulation results for estimating regions.

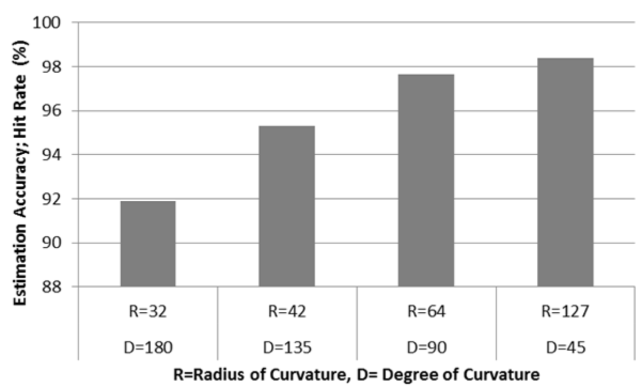

Fig. 7. Experimental result on various degree of curvature and radius of curvature with 2norms different functions.

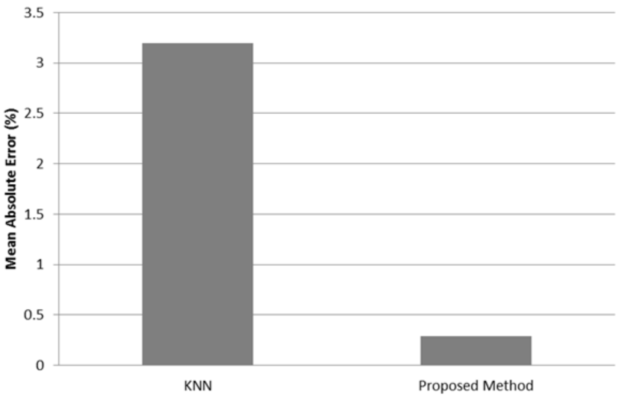

Fig. 8. Experimental result on based line KNN vs Proposed method.

A. *Estimation Accuracy Analysis with Different Norms*

Since the essence of KNN algorithm is established on calculating the distance between the test data and each of the training data to make a decision of the final estimation, the estimation accuracy may vary with different distance measures, in terms of norms, Lp-norms with p ∈ [1, ∞]. ṽ is defined as vector space for all vectors x ∈ ṽ. The size of the error vector (ṽ) is called norm of vector, x. Therefore, the measurements of different norms affect the estimation accuracy of algorithm. There are a few different kinds of norm for measuring the size of error vector, such as L1-norm, L-2norm (Euclidean-norm), Lp-norm and infinity-norm.

L1-norm is defined as the sum of the absolute values of its data points as follows:

$$\|\tilde{v}\|_1 = \sum_{i=1}^{n} |v_i| \quad (16)$$

L2-norm is defined as the square root of the sum of square of absolute values of its data points as follows:

$$\|\tilde{v}\|_2 = \sqrt{\sum_{i=1}^{n} |v_i|^2} \quad (17)$$

∞-norm is the limit of Lp-norm for p→∞ is defined as follows:

$$\|\tilde{v}\|_p = \sqrt[p]{\sum_{i=1}^{n} |v_i|^p} \qquad , p \in [1, \infty] \quad (18)$$

In this study, the experiment is based on ten different norm functions including L1-norm, Euclidean (L2-norms) and Lp-norm with p € [3, 10]. Lp-norms is defined as follows:

$$\|\tilde{v}\|_p = \sqrt[p]{\sum_{i=1}^{n} |v_i|^p}\ ,\ p \in [3, 10] \quad (19)$$

B. *Evaluation Results with Different Norms*

How the different norms in distance function have direct impact on the estimation accuracy of KNN algorithm is very interesting point in location estimation process. Otherwise, how the influence of the different norms in different radius of curvature values and degree of curvature values impacts on the estimation accuracy of proposed method is very interesting problem in this experiment.

We conducted our experiment with L1-norm, L2-norm and Lp-norm, here, Lp-norm is set up with p € [3, 10]. The experimental result shows sensitivity of estimation accuracy of each radius R values and degree D values on 10 different norms. The experimental results are shown in Fig. 9.

We can see that the estimation accuracies of even norms, such as 2norms, 4norms, 6norms, 8norms and 10norms, are increasingly remained over 80% and the estimation accuracy decreases when the value of norm is greater. The estimation accuracy dramatically decreases in odd norms, such as 3norms, 5norms, 7norms and 9norms, except 1norm which accuracy is over 90% as well as other even norms.

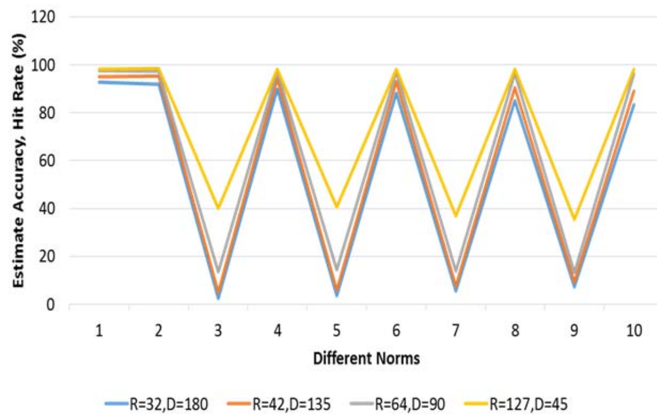

Fig. 9. Experimental result on different norms.

VII. CONCLUTIONS

Although the existing localization methods achieved the satisfying estimation accuracy in literature, they still have critical issues in sensitivity and robustness of localization estimation. In this study, an advanced localization method is proposed concerning RSSI values. The proposed idea is based on circle theory and intend to improve estimation accuracy analyzing and optimizing radius values. According to the simulation results, the proposed method gives a better accuracy in location estimation than KNN based localization method. In addition, experimentation is performed in terms of different norms functions, such as Lp-norm for examining the sensitivity of the proposed method. The experimental results show that the best estimation accuracy is found on L1-norm and L2-norm with radius value = 127 and degree = 180^o. To improve the robustness of proposed localization method, this analysis encourages to penalize a constraint on the curvature of radius values because of the relationship between the robustness of the classifiers and the curvature of the radius values. In near future, we will focus on building a propagation model with our own scenario to create a fingerprinting database by expecting to gain better simulation results and mathematical proof.

REFERENCES

[1] Chao Liu, Aroland Kiring, Naveed Salman, Lyudmila Mihaylova, Inaki Esnaola, “A Kriging Algorithm for Location Fingerprintingbased on Received Signal Strength,” Sensor Data Fusion: Trends, Solutions, Applications (SDF), 2015, 15666434, 6-8 Oct. 2015.

[2] Peter Brida, Juraj Machaj, Jozef Benikovsky, “WLAN Based Indoor Localization System,” Vol. 2, NO. 4, December 2011.

[3] Weisstein, Eric W. “Parametric Equations”. MathWorld.

[4] Alfred Gray, “Modern Differential Geometry of Curves and Surfaces with Mathematica,” 2nd ed. Boca Raton, FL: CRC Press, 1997.

[5] Binghao Li , James Salter , Andrew G. Dempster , Chris Rizos, “Indoor Positioning Techniques Based on Wireless LAN,” First IEEE International Conference on Wireless Broadband and Ultra Wideband Communications, Sydney, Australia, March 2006.

[6] Mussa Bshara, Leo Van Biesen, “Fingerprinting-based Localization in WiMAX networks depending on SCORE measurements,” 2009 fifth International Conference on Telecommunication, 12 June 2009.

[7] UJIIndoorLoc: A New Multi-building and Multi-floor Database for WLAN Fingerprint-based Indoor Localization Problems, Joaquin Torres-Sospedra, 2014 International Conference on Indoor Positioning and Indoor Navigation, 27th-30th October 2014.

[8] Qiang Yang, Sinno Jialin Pan, and Vincent Wenchen Zheng, “Estimating Location Using Wi-Fi,” IEEE 2007 ICDM Contest, Vol.23, No.1, January/February 2008.

[9] Dan Li, Le Wang, Shiqi Wu, “Indoor Positioning System Using Wifi Fingerprint,” Stanford University.

[10] Gints Jekabsons, Vadim Zuravlyov, “Refining WI-FI Based Indoor Positing,” Institute of Applied Computer Systems, Riga Technical University.

[11] Labyad Asmaa; Kharraz Aroussi Hatim; Mouloudi Abdelaaziz, “Localization algorithms research in wireless sensor network based on Multilateration and Trilateration techniques,” 2014 Third IEEE International Colloquium in Information Science and Technology (CIST), 20-22 Oct. 2014.

[12] Oguejiofor O.S, Aniedu A.N, Ejiofor H.C, Okolibe A.U, “Trilateration Based localization Algorithm for Wireless Sensor Network,” International Journal of Science and Modern Engineering (IJISME), Volume-1,Issue-10, September 2013, pp-21-27.

[13] Shahrzad Khodayari, Mina Maleki, Elham Hamedi, “A RSS-based Fingerprinting Method for Positioning based on Historical Data,” 2010 International Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS), 11-14 July 2010.