วารสารวิจัย UTK ราชมงคลกรุงเทพ

Analysis of indoor Wi-Fi localization using gaussian process regression and K-nearest neighbor algorithms

Myo Myint Maw[1*], Hnin Mya Nandar Myo Tint[1], Sarun Duangsuwan[2]

[1] Department of Computer Engineering and Information Technology, Mandalay Technological University, Mandalay, Myanmar
[2]Department of Information Engineering, Faculty of Engineering, KMITL, Prince of Chumphon Campus, Thailand
*Corresponding author. Email: myomyintmawphdit5@gmail.com, myomyintmaw@mtu.edu.mm

ABSTRACT

Global positioning system (GPS) cannot well localize in indoor environment. Nowadays, indoor Wi-Fi localization is very attractive in positioning and localization research areas because of existing Wi-Fi infrastructure can be used to conduct, so there is no need extra hardware requirements and cost is not expensive. In indoor Wi-Fi localization, the received signal strength indicator (RSSI) fingerprinting played a key role in the access point performance services. This paper proposed the accuracy analysis of indoor Wi-Fi signal based on localization using machine learning approach. Four accept points (APs) were used to measure RSSI and the measured RSSI data were configured as RSSI fingerprint database which was composed of RSSI data of each AP and the position of receiver point. Four corridors which has 1m x 1m has were researched for indoor Wi-Fi localization. 3000 training RSSI and 234 testing RSSI data points were applied in indoor localization. The machine learning algorithms: Gaussian process regression (GPR) and K-nearest neighbor (KNN) approaches were proposed to analyze the accuracy of indoor localization. The results of Wi-Fi localization accuracy were shown using GPR and KNN in indoor environment. The accuracy and mean square error were discussed for indoor Wi-Fi based on GPR and KNN. In this proposed Wi-Fi indoor localization, the accuracy using GPR performed 75%. Moreover, the accuracy using KNN was 83% when K neighbor value was five. When K neighbor value was 10, the accuracy also outperformed 83%. Finally, according to the results, the analysis of corridors using KNN could provide more accuracy than GPR for testing environment in this study.

Keywords: Fingerprint, Wi-Fi, Gaussian process regression, K-nearest neighbor, Received signal strength indicator, Indoor localization

1. Introduction

Global positioning system (GPS) is widely used in outdoor localization but indoor localization does not commonly used GPS signal because of this signal cannot well reach to the destination. Blocking signal is also occurred in this system. However, the accuracy, precision and the position factors are also important in indoor localization. Therefore, most of indoor localizations are employed Wi-Fi signal because these are already installed in existing environments. The cost in indoor localization is effective using existing infrastructure; e.g. access points (APs). There are many indoor Wi-Fi applications such as shopping malls, airports, university campus, offices, hospitals and factories, [1]. In these localization techniques such as angle of arrival (AOA), time of arrival (TOA), time of flight (TOF), time

Received 17-03-2020
Revised 19-05-2020
Accepted 22-05-2020

difference of arrival (TDOA), triangulation, received signal strength indicator (RSSI) [2]. Among these techniques, RSSI based fingerprint indoor localization has been found in many research papers. The RSSI based fingerprint approach could not find more accurate solution [3]. Therefore, this paper has been studied the RSSI fingerprint using machine learning approach: supervised learning such as K-nearest neighbor (KNN) and gaussian process regression (GPR).

Due to the accuracy of Wi-Fi localization system, most of machine learning based algorithms are applied in most of researches. Gaussian process regression has been developed using received signal strength based on Wi-Fi localization system. The estimation of localization error has been proposed in industrial environments using gaussian process [4]. This research has been examined a large number of machine learning algorithms for indoor localization based on the sensors available in smartphones [5]. The authors proposed hybrid approach which has been achieved accuracy by the best offline instance-based methods and the speed of non-instance-based methods. In these approaches, there are two main parts: passive and active localization [6]. This paper has been studied passive localization based on machine learning approach, Naïve Bayes classification. For active technique, the example of smart phones is used in real environment. So RSSI data and channel state information (CSI) data are applied in this research. The purpose of this research is to find out how the precision of indoor localization using RSSI fingerprinting is affected by different levels of number of connected Wi-Fi devices to APs [7]. Experiments have been carried out at two different locations, which give a chance to find similarities in results between real world environments [7]. In this research, indoor positioning using Wi-Fi access point is investigated as the main usage of location based service applications. The models are designed using KNN algorithm for indoor positioning. This paper compares the performance of the data of four access points and six access point data [8].

The aim of this paper is to develop Wi-Fi based indoor positioning approach using for GPR and KNN which has acceptable accuracy. This paper investigates the prospect of RSSI with fingerprint database indoor localization. Besides, the investigation and analysis among RSSI fingerprinting is also studied in real environment inside Mandalay Technological University (MTU) using existing AP's data. RSSI data is measured by using laptop, in a selected place in main building. Additionally, the comprehensive details of indoor localization using GPR and KNN algorithms are presented in this paper.

2. Methodology

For localization in indoor environment based on fingerprint using machine learning approach, two phases were required: data collection and analysis. The APs were required as the transmitter and laptop as the receiver data before data collection in the proposed model. In this proposed indoor localization model, firstly, the collection of RSSI data were stored as training data in fingerprint database. For data collection, the DELL laptop was used with Windows 7. The laptop model code was INSPIRON 1464. The model of the Wi-Fi adapter or network card was DELL wireless 1397 WLAN Mini card. To collect the training data, the net-spot software was used with DELL laptop which was core i3 and 4-GB RAM. The averaged RSSI for each reference point from each access point (AP) is shown in this software. The type of access point was Unifi which was UAP-LR, and operating frequency was at 2.4 GHz, data speed was 300 Mbps. In this paper, the same type of Wi-Fi access point was used. The measurement model used RSSI values from four Wi-Fi APs operated at frequency of 2.4-GHz. The transmission power of each Wi-Fi access point was 24 volt. The measurement area had four access point has already been installed at main building, MTU. Secondly, the collections of RSSI fingerprint data were analyzed using GPR and KNN. And then, the comparison results were described for proposed indoor localization system.

The data collection phase consisted of taking from one point to next point was apart 1m to receive Wi-Fi signal strength and storing the data to the fingerprint database. In indoor proposed localization, the corridors regions were measured. We collected the RSSI about 10 times for each point and got then the averaged RSSI data along the four parts of corridors and stored the RSSI to the database. In experimental regions, the corridor of indoor area was approximately 3462 m^2. Each of the test area, corridor, at second floor, MTU was divided by blocks. Each block was approximately 1m long and 1m wide.

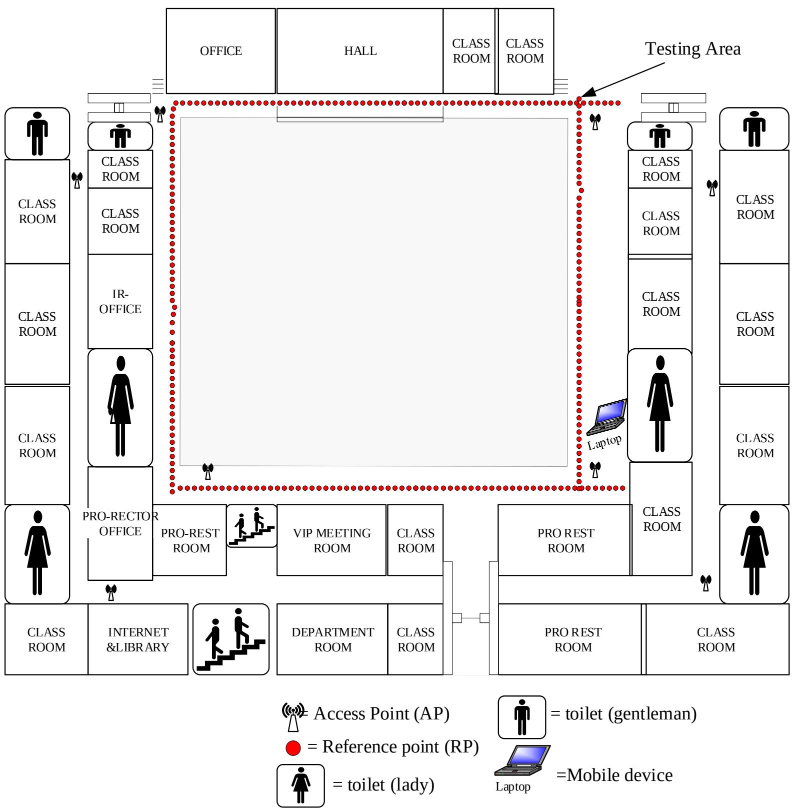

Figure 1. Measurement plan for corridors at second floor, main building, MTU.

2.1 RSSI data collection

The collected the total data point for each was 234 points. For each point, the reference distance was separated from 1m apart. These RSSI data of 234 points were used as testing data for simulation. The training fingerprint database composed of each column had RSSI of four APs and point location as train.csv file. In testing fingerprint database, each column had RSSI of four APs and point location as test.csv file. Location file as location.csv included X coordinate as first column, Y coordinate as second column and reference point number. The collected measurement area for corridors are shown in Figure 1.

2.2 Indoor localization fingerprinting

In fingerprinting database, there are primarily two phases is shown in Figure 2: calibration phase is also called offline phase or training phase and positioning phase is also called online phase or testing phase. In the calibration phase, Laptop device was used as a receiver to measure RSSI values in dBm from several access points at the chosen calibration point in the area of interest. Then, the fingerprint database was also called radio map includes $(RSSI)_n$ values of all $(APs)_m$. n is the number of calibration points, m is the number of APs. Usually, the average samples were recorded for each location. Online phase determined the location of a user by comparing the measured RSSI values with the training database and applied a location estimation algorithm to estimate its current location. In this phase, we developed fingerprinting methods using GPR and KNN algorithms. Then, the accuracy of these methods was analyzed and the comparison results were discussed for proposed indoor localization. The schematic overview of the system design model is illustrated in Figure 2. This fingerprint database comprised of RSSI of each AP and location.

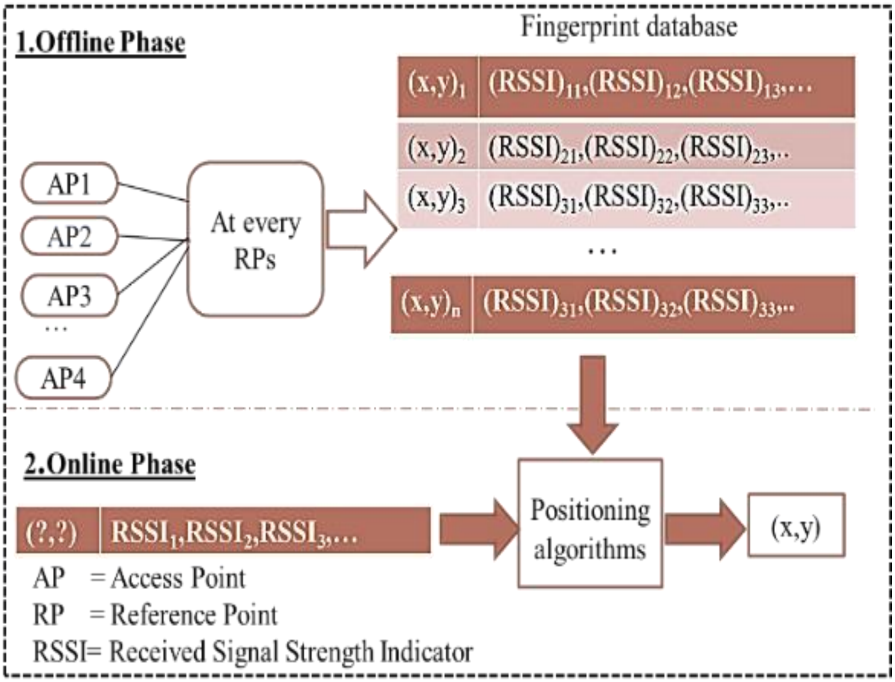

Figure 2. Proposed model for indoor localization.

2.3 Gaussian process regression

In this section, the indoor proposed localization models is implemented by GPR. In indoor localization based on fingerprint, the GPR is used to construct the model using RSSI. In this paper, let D = {(x_1, y_1), (x_2, y_2), . . . , (x_n, y_n)} be a set of training samples with RSSI for indoor corridors, locations which drawn from a noisy process

$$y_i = f(x_i) + \varepsilon \quad (1)$$

where each x_i is an input sample and each y_i is a target value, or observation, in R. ε is zero mean, additive Gaussian noise with known variance σ_n^2. GPR estimates Gaussian distributions over function based on training RSSI data and each point. The GPR is fully specified by a mean function and a covariance function such as [9]:

$$m(x) = E[f(x)] \quad (2)$$

$$k(x_i, x_j) = E[(f(x_i) - m(x_i))(f(x_j) - m(x_j)] \quad (3)$$

where x_i and $x_j \in$ R are random variables. Without loss of generality, mean can be considered as zero. The kernel function in Equation 4 considered in this work is expressed as

$$k(x_i, x_j)\sigma^2 \exp\left(-\frac{\|x_i - x_j\|}{2l^2}\right) \quad (4)$$

For GPR, we only put attention on a finite subset of function values $f = (f(x_1), f(x_2), ..., f(x_n))$ which follows a regular gaussian process (GP) distribution such that

$$f \approx GP(m(x), k(x_i, x_j)) \quad (5)$$

where μ = 0; $k = k(x_i; x_j)$. Let x^* be a matrix with on each row a new input point X_i^*, i = 1…n. To sample a function, compute firstly the covariance between all inputs in X^* and collect these in $n \times n$ matrix:

$$K(x^*, x^*) = \begin{bmatrix} k(x_1^*, x_1^*) & k(x_1^*, x_2^*) & \cdots & k(x_1^*, x_n^*) \\ \vdots & \vdots & \ddots & \vdots \\ k(x_n^*, x_1^*) & k(x_n^*, x_2^*) & \cdots & k(x_n^*, x_n^*) \end{bmatrix} \quad (6)$$

where choosing the prior mean function m(x) = 0 to simplify the matrix algebra shown in Equation 6, then sample values of f at inputs X^* from the GP by sampling from a multivariate normal distribution.

$$f \approx N(0, k(x, x^*)) \quad (7)$$

where the notation $f_* = \left[f(x_1^*), ..., f(x_n^*)\right]^T$. f_* is a sample of the function values. To sample observations y_*, we would have to add an additional and independent sample of the noise term ε.

For posterior prediction form a GP, the collected observations $D_i = (x_i, y_i)$ and make predictions for new inputs X_* by drawing f_* from the posterior distribution $P(f, D_i)$. By definition, previous observations y_t and function values. f_* follows a joint (multivariate) normal distribution. This distribution can be written as

$$\begin{bmatrix} y_t \\ f_* \end{bmatrix} = GP\left(0, \begin{bmatrix} k(x_t, x_t) + \sigma^2 I & K(x_t, x_*) \\ K(x_*, x_t) & K(x_*, x_*) \end{bmatrix}\right) \quad (8)$$

where $K(X_i, X_i)$ is the covariance matrix between all observed points so far, $K(X^*, X^*)$ is the covariance matrix between the newly introduced points as described earlier, $K(X^*, X_i)$ is the covariance matrix between the new input points and the already observed points and $K(X_i, X^*)$ is the covariance matrix between the observed points and the new input points. Moreover, I is an identity matrix and σ^2 is the assumed noise level of observations. This posterior is also a GP with mean function

$$m_t(x) = K(x, X_t).\left[K(X_t, X_t) + \sigma^2 I\right]^{-1} y_t \quad (9)$$

and the covariance function

$$k_t(x, x) = k(x, x^*) - K(x, X_t) \left[K(x_t, x_t) + \sigma^2 I\right]^{-1} K(x_t, x^*) \quad (10)$$

This means that calculating the posterior mean and covariance of a GPR involves first calculating the four different covariance matrices above and then combining them.

2.4 K-nearest method

The received signal strength of Test Points (TPs) was compared with the data ofFingerprint database in terms of corresponding KNN in online stage. To build GPR, this research used the MATLB 2019b fitcknn module.

The estimated position was calculated for proposed system. We used KNN algorithm based on Euclidean Distance to find the nearest point. Consider RSSI fingerprinting data base: {$RSSI_i$, RP_{xi}}. The RP_{xi} was shown the location of i^{th} RP, and $RSSI_i$ of i^{th} RP was a vector of RSSs which was: {$RSSI_1$, $RSSI_2$,…,$RSSI_n$}. This algorithm simply chose the fingerprint that had the minimum distance to the current measurement in the signal space. In fact the K-nearest neighbors distance and the signal strength vector of the mobile user could be {$RSSI_1$, $RSSI_2$…$RSSI_j$,…,$RSSI_n$}, where $RSSI_j$ refers to the RSSI of AP j^{th}. So, the Euclidian distance of RSSIs of i^{th} RP and user's is calculated by [10]:

$$D_i = \sqrt{\sum_{j=1}^{n} (RSS_{ij} - RSS_j)^2} \quad (9)$$

3. Results and discussion

In this experiment, this paper analyzed the accuracy of Wi-Fi localization based on fingerprinting with GPR and KNN methods for indoor environments. In this proposed model, we simulated the process using MATLAB programming.

There are a total of 234 RSSI data for each AP, 234 locations for X coordinate and Y coordinate. Therefore, the total RSSI of four APs, each RSSI was three times, 3000 RSSI for all APs. The RSSI data was randomly separated with a ratio of 90:10. Total of 3000 data used to train the machine were used for testing the optimum machine learning regression model. Ten-fold cross-validation was used to train the machine learning regression model, therefore 3000 data were divided into ten folds, with a total of 300 data for each fold. Before the process begins, the 3000 data randomly separated with a total of 2700 data for the training set and a total of 300 data for the validation set. Based on the ten folds means there were nine folds containing 2700 training set data and 300 fold containing 2400 validation set data. Then, the process using data that divided by folds type was repeated ten times (K = 10). The position of the validation set fold was different in each iteration. Figure 3 is the process of ten-fold cross-validation. were found by calculating the distance of each RP of database vector from the tag based on their RSSs considering Euclidian

The training data of RSSI values is 3000 and the testing data is 234 in fingerprint database using GPR and KNN.

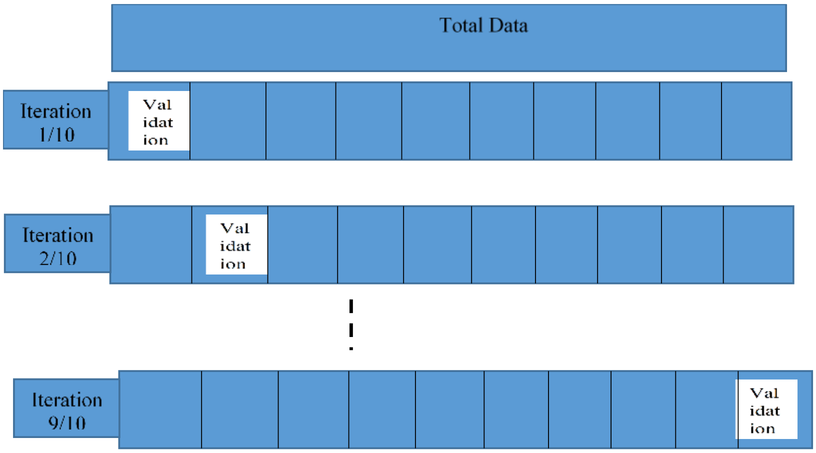

Figure 3. Process of ten-fold cross validation in GPR and KNN.

The following Figure 4 is shown for training and testing RSSI values for four corridors of four APs. The points of X, Y locations are shown in Figure 4. The X point is maximum is 65 m and Y point is maximum 55 m. The red line color is training signal and the bule line color is for testing signal in the following Figure 4. The location point in X coordinate and Y coordinate is shown in Figure 5. The blue circle point is training points and the red plus sign is testing pints for location to collect the RSSI of four APs.

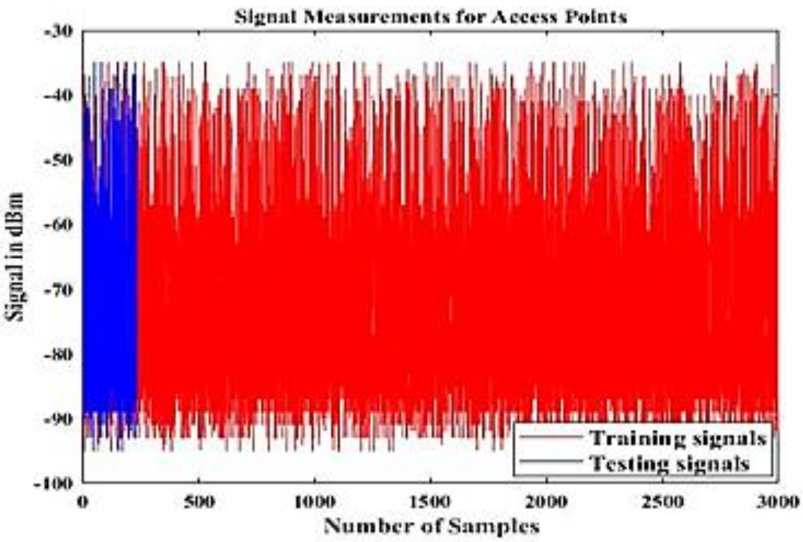

Figure 4. RSSI Signal for training and testing

In Figure 6, the simulation result of performance accuracy for KNN was described. The green circle point is correct point, the red cross point was fail point and the black dot was training point in simulation as shown in Figure 6. Using KNN, the 194 correct points were occurred in testing environment. For 4 APs model, the most precise result occurred when K=5, cross validation is 5 fold, and K=10, cross validation is 10 fold with 82% and 83% accuracy, respectively.

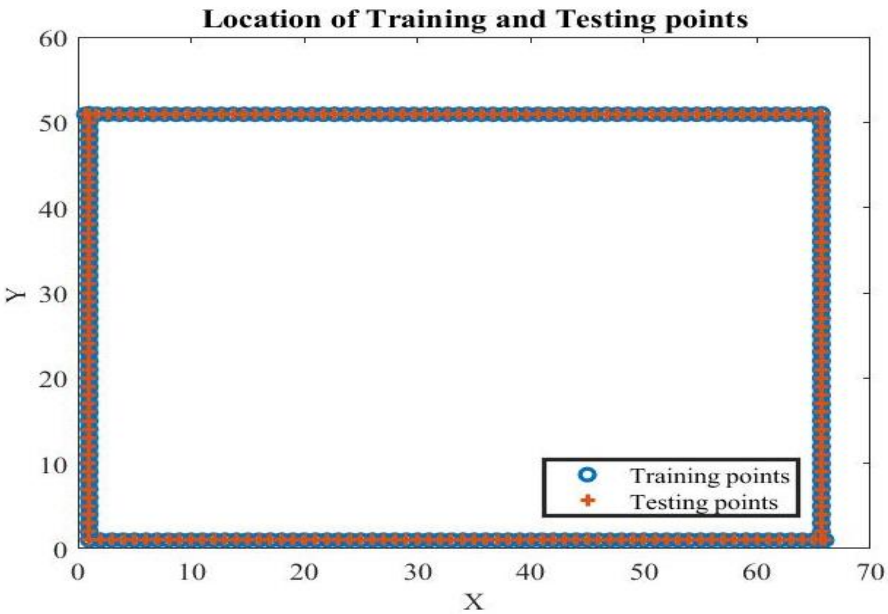

Figure 5. Location of X,Y points for training and testing

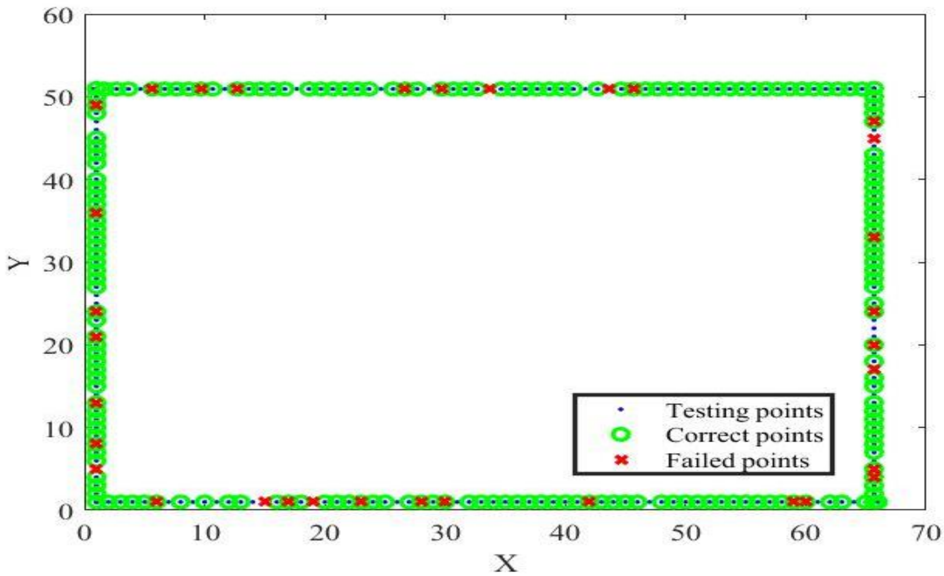

Figure 6. The estimation result of data position getting from using KNN

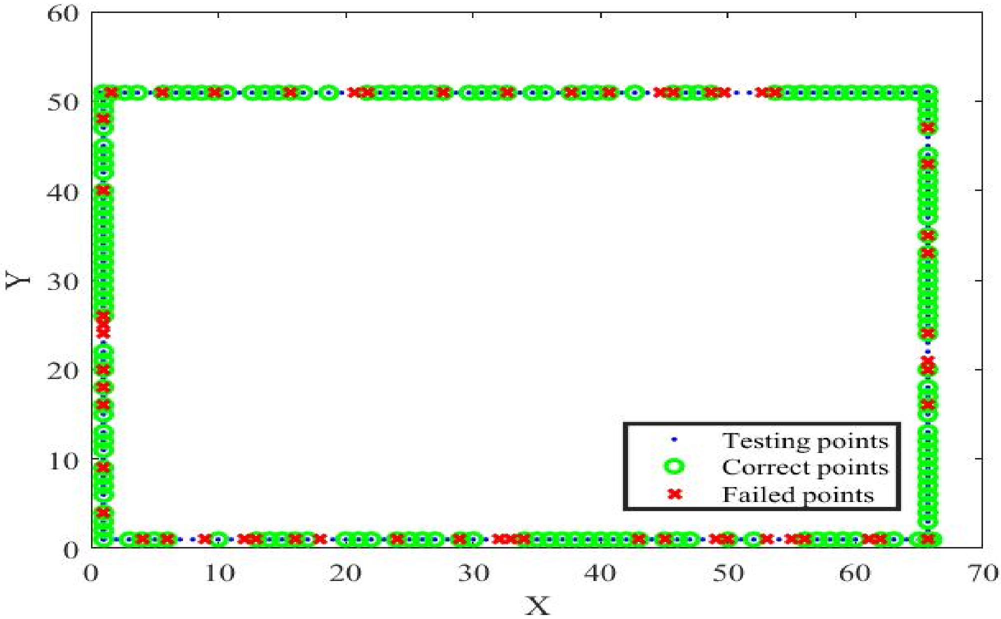

Figure 7. The estimation result of data position getting from using GPR

In Figure 7, the estimation result of localization points based on GPR method was displayed. GPR model using the squared exponential kernel function with default kernel parameters. Training dataset was used to build the GPR model and validation dataset was used to evaluate the prediction of the GPR model compared to the measured data are required. To implement KNN, this research used the MATLB 2019b fitrgp module. The 176 correct points among testing points of 234 points were resulted using GPR. The positioning accuracy acquired just about 75.23% in simulation result for localization for four APs.

4. Conclusions

In this paper, Wi-Fi signal strength in indoor localization was implemented using KNN and GPR approach for proposed system. The simulation results showed that the KNN algorithm can bring enhancement compared with the GPR for the fingerprint matching approach of localization accuracy for testing indoor environment. According to the accuracy results, the KNN approach was better performance in indoor proposed environment. In the future, we will incorporate the control input information, such as the removing noise using the Bayes filter.

5. Acknowledgements

We thank our colleagues from Mandalay Technological University who provided insight and expertise that greatly assisted the research. We thank reviewers for comments that greatly improved the manuscript.

6. References

[1] Wang C, Kao T, Fang S, et al. Robust Wi-Fi Location Fingerprinting Against Device Diversity Based on Spatial Mean Normalization. Proceedings of Asia-Pacific Signal and Information Processing Association Annual Sumit and Conference (APSIPA ASC) [Internet]. Kaohsiung, Taiwan; 2013. Available from: http://hb.diva-

portal.org/smash/get/diva2:1095554/FULLTEXT01.pdf
[2] Zafari F, Gkelias A, Leung King K, A Survey of Indoor Localization Systems and Technologies, IEEE Communications Surveys and Tutorials, 2019. p 2568 - 99.
[3] Akram BA, Akbar A, Wi-Fi Fingerprinting Based Room Level Indoor Localization Framework Using Ensemble Classifiers. [Internet]. Mehran,2019. DOI: 10.22581/muet1982.1901.13.
[4] Duvallet F, Tews A D, WiFi Position Estimation in Industrial Environments Using Gaussian Processes. [Internet]. IEEE; 2008: DOI: 10.1109/IROS.2008.4650910.
[5] Mascharka D, Manley E, Machine Learning for Indoor Localization Using Mobile Phone-Based Sensors[Internet]. 2015. Available from: https://arxiv.org/pdf/1505.06125.pdf
[6] Wu Z, Xu Q, Li J, et al. Passive Indoor Localization Based on CSI and Naive Bayes Classification. [Internet]. IEEE; 2018. DOI: 10.1109/TSMC.2017.2679725.
[7] Christoffer O, Tobias O, The precision of RSSI-fingerprinting based on connected Wi-Fi devices [dissertation]. Högskolan i Borås; 2017.
[8] Amirsoori S, Daud SM, Ahmad N A, et al. WI-FI Based Indoor Positioning Using Fingerprinting Methods (KNN Algorithm) in Real Environment. International[Internet]. Journal of Future Generation Communication and Networking. 2017; 10(9): 23-36 Avialable from: http://dx.doi.org/10.14257/ijfgcn.2017.10.9.03.
[9] Rasmussen CE, Gaussian processes in machine learning, in Advanced Lectures on Machine Learning[Internet]. https://www.cs.ubc.ca/~hutter/EARG.shtml/earg/papers05/rasmussen_gps_in_ml.pdf
[10] Ding X, Gao L, Wang Z, Modified Fingerprinting Algorithm for Indoor Location[Internet], Journal of Communications. 2017; 12(3): 145-51 Available from: http://www.jocm.us/uploadfile/2017/0331/20170331050729488.pdf.