Bio metric and Fingerprint Database for Applying Interpolation and Regression Techniques for IoT-based Indoor Localization

Jaya Priyadharisini J
Assistant Professor, Department of Computer Application Christ College of Arts and Science, Kilachery

INTRODUCTION

Internet of Things (IoT) is the networking of physical objects that contain electronics embedded within their architecture in order to communicate and sense interactions amongst each other or with respect to the external environment. In the upcoming years, IoT-based technology will offer advanced levels of services and practically change the way people lead their daily lives. Advancements in medicine, power, gene therapies, agriculture, smart cities, and smart homes are just a few of the categorical examples where IoT is strongly established.
IOT is a system of interrelated things, computing devices, mechanical and digital machines, objects, animals, or people that are provided with unique identifiers. And the ability to transfer the data over a network requiring human-to-human or human-to-computer interaction

Wireless sensor networks

Wireless Sensor Network (WSN) is an infrastructure-less wireless network that is deployed in a large number of wireless sensors in an ad-hoc manner that is used to monitor the system, physical or environmental conditions.
Sensor nodes are used in WSN with the onboard processor that manages and monitors the environment in a particular area. They are connected to the Base Station which acts as a processing unit in the WSN System.

Base Station in a WSN System is connected through the Internet to share data.

ZigBee is an open, global, packet-based protocol designed to provide an easy-to-use architecture for secure, reliable, low power wireless networks. Flow or process control equipment can be place anywhere and still communicate with the rest of the system. It can also be moved, since the network doesn't care about the physical location of a sensor, pump or valve.

The distance-free indoor localization technique, on the other hand, has advantages in reducing the effect of this signal fluctuation caused by the multipath effect in the indoor environment by collecting the spatial information of this signal to record as the database. The common distance-free technique in indoor localization is fingerprinting. The fingerprinting technique requires two phases for the localization, the first phase is called the offline phase, and the second phase is the online phase . In the offline phase, the necessary information related to the spatial information of the area of interest is recorded as the offline fingerprint database. The fingerprint database consists of the signal properties information related to a specific location on the designated grids inside the area of interest: the denser these grids, the more accurate and precise later for the online phase or the localization process. After the data or localization parameters are measured and stored in the database acquisition, the online phase, in which the target or object sends the signal parameter to the system and receive the same parameters as recorded in the database, the pattern matching algorithm will work by comparing the target's signal parameters to those in database. This algorithm then concludes that the target or object belongs to a particular position with similar spatial information recorded in the fingerprint database. Some proposals that applied the interpolation techniques to tackle the database sparsity have been published in The most used interpolation technique for fingerprint-based indoor localization is the Kriging technique. The authors in showed that the Kriging for RSSI, especially in inaccessible areas, can be covered and enhance the overall fingerprint database. Another approach of the database enhancement by path-loss model-based interpolation is available in, the crowdsourcing method, Spatio-temporal similarity and clustering-based and interpolation on. However, most of these approaches have limitations on algorithm complexities. Furthermore, some issues related to the advancement parameter, e.g., Channel State Information (CSI) and its complexity, appear and become the drawbacks of applying this parameter. Thus, as far as our concerns, the simple yet straightforward implementation of interpolation and regression technique is not yet considered—especially when using WSNs-based as the system's core. By utilizing RSSI following the log-loss distance, it views the linear assumption relationships between signal strength and the distance. By considering this, the interpolation and regression can be established as the power-distance relationship in the applied environment. In addition, we consider conducting an actual measurement campaign in our approach. The approach is the algorithm development in simulation or theory and actual implementation both for two- dimensional (2D) and three-dimensional (3D) by using the ZigBee standard as the core of the WSNs system. Our original achievements are enhancing the density database by using a relatively sparse actual measurement database by applying basic interpolation and regression. Our approach is relatively simple compared to previously mentioned publications. Our significant difficulties to overcome is the RSSI fluctuation in some parts of fingerprint position because of the nature of the environment, i.e., near the edge, enormous metal material, and unbalanced obstruction in the environment, making the synthetic database challenging to assure.

The fingerprint technique for both environments is applied, and the offline database fingerprint is obtained by the area of interest 5×5 m^2 for the 2D environment. We utilize the bookshelf as the 3D environment, assuming several floor applications in a multi-story building in the same environment as the 2D settlement. We design the database grid of 1 × 1 m for the 2D and 22×12.5×35 cm for the 3D environment. The cm-scale is to observe how our interpolation and regression algorithm can work well and acceptable in such a short distance. The pattern matching used is the classical minimum Euclidean distance (MED) to know-how is the difference between the predicted and the actual RSSI values. The tiniest error yield from the target-database comparison is assigned as the target's location. From the context we have explained and by proposing the method for actual implementation, we would like to highlight the contribution of our proposed method as follows:

- Implementing the interpolation and regression technique for database enhancement in fingerprint-based indoor localization considering 2D and 3D environments.
- We compare the interpolation and regression techniques to observe the best use of the technique for the case of 2D and 3D environments.

We present the structure of this paper as follows; in the first part, we discuss the introduction and background stated the context and importance of our proposal. In the second part, we present the indoor localization technologies and techniques and the fingerprint technique's comprehensive explanation. For the third part, we detail the material and method, including the ZigBee-based WSNs system, interpolation, regression technique we propose to implement, pattern matching, performance metrics, and the measurement system and setup. Results and discussion will then be presented in the fourth part. Finally, we will conclude our findings and discuss our proposal limitations and plan of our near-future works.

Indoor localization technologies

Indoor positioning via client means that the position is determined directly via the end-user device (“client” – usually a smartphone). This method requires an app. It is most often used for indoor navigation projects and wherever you need to communicate with the user.

Table 1. RF-based indoor localization technology.

Technology	Accuracy (m)	Range (m)	Power (W)
GPS	1 - 20	global	500
RFID	1	1 - 50	0.02 - 0.3
Wi-Fi	1 - 5	<100	0.5 – 1
UWB	<0.3	<300	0.03
BLE	1	<10	0.001
ZigBee	1 - 5	<30	0.02 – 0.04

Indoor localization can act as an "indoor GPS," and some technologies have been introduced to develop indoor localization research. The technologies include radio frequency (RF)-based and other, i.e., mechanical, optical/light wave, acoustic wave, and vision-based, also attract the research further. However, RF-based researches are most common and widely implemented. Table 1 shows the most-used RF-based indoor localization technologies and their features

In this paper, we consider applying the ZigBee technology because of several reasons:

- By comparing the accuracy of other technologies in Table 1, the ZigBee system can achieve an accuracy of 1 – 5 m. UWB technology can achieve better accuracy than ZigBee. However, implementation of UWB need additional hardware, and to achieve <0.3 m accuracy, several advances and complicated algorithm are needed, e.g., Time-of- Arrival (ToA), Ranging Time (RT).

- Observing the range properties of ZigBee technology, the <30 m can cover almost all indoor environments, the scalability issues can be solved.
- Power consumption is also low compared to other technologies; only RFID and BLE give the same or less than ZigBee. However, the range of BLE is relatively short, while the RFID technology also requires the line-of-sight (LoS) communication between tags and the RFID reader.

ZigBee standard is the IEEE 802.15.4 standard working in industrial, scientific, and medical (ISM) bands similar to Wi-Fi in 2.4 GHz. In the measurement system and setup, we will explain how to reduce the effect of the signal interference with Wi-Fi.

2-2- Distance Measurement Technique

Figure 1 shows the several distance measurement or signal properties used in indoor localization.

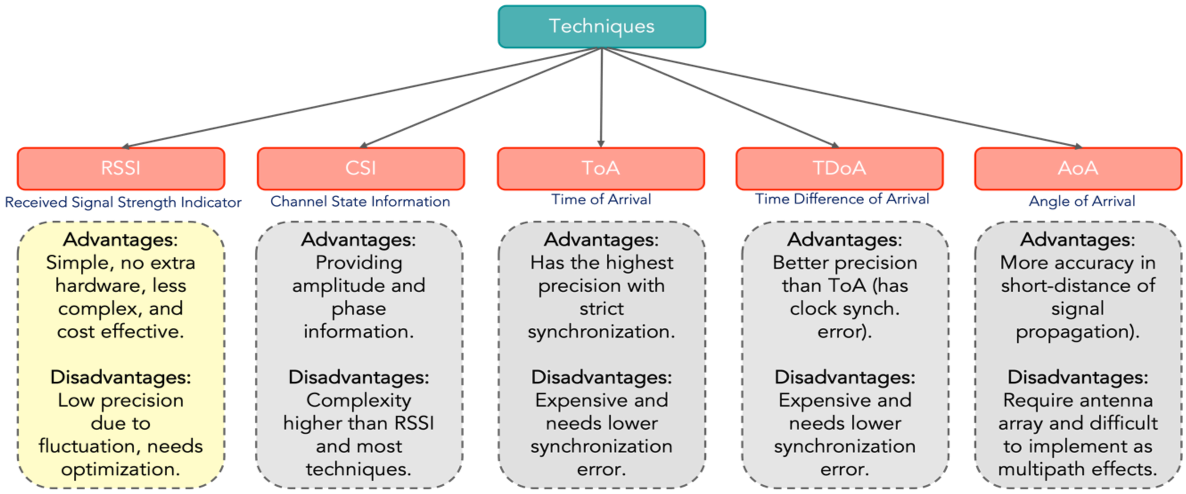


Figure 1. Distance measurement technique.

There are some advantages and disadvantages of each distance measurement technique. However, we can note from the illustration that the RSSI parameter has the most prospective implementation, especially in our approach. First, it has advantages in the straightforward implementation, the second there are disadvantages point that becomes a focal point, especially for the database enhancement in our proposal. RSSI is defined as the received power, and it follows the log-loss distance model [36]:

$$R(dBm) = A - 10 \cdot n \cdot \log_{10} \left[\frac{d_0}{d}\right]. \qquad (1)$$

RSSI in dBm related to the distance, d, is equal to the A, power in dBm of reference distance, d_0 subtracted by the path loss exponent, n, multiply the log distance of d_0 divided by the d; generally d_0 is 1 m. The values of n can be measured by empirical in the particular indoor environment used. In our approach, we collected the RSSI values from the ZigBee standard without considering the RSSI-distance conversion.

2-3- Indoor Localization Methods

Figure 2 shows several basic methods applied in indoor localization systems.

The fingerprint technique gives advantages when there is no need for distance-parameter conversion. Furthermore, the fingerprint technique gives better accuracy than several range-based techniques, i.e., triangulation or trilateration, min-max, and iRingLa compared based on RSSI as localization parameters [37]. The other techniques can also yield high accuracy but need additional support in the algorithm, precise clocking, accurate angle estimation, and the high number of nodes or beacons used for positioning. For previously mentioned advantages of fingerprint technique, however, there is a drawback of fingerprint technique, especially in the offline database construction process, with the burden of the cost and time inefficient, moreover, if there is an application in the large-scale indoor environment, the system will be more complex and needs more human resources. The trade-off of this technique is that if we do not have enough density of fingerprint database, the localization error will be high. On the contrary, when we need to have a very dense fingerprint database, multiple drawbacks will appear related to offline databasing. In this paper, we propose the database enhancement in tackling the drawbacks of the fingerprint technique, especially in the database sparsity, by applying the interpolation and regression technique.

Fingerprint Technique

Biometrics such as fingerprints, voices and ECG signals are unique human characters that cannot be tampered or replicated. This facilitates real-time system implementations. Biometric Attendance systems are commonly used systems to mark the presence in offices and schools as well as in Biometric Security Lock. This project has a wide application in schools, colleges, business organizations, offices where marking of attendance is required accurately with time. Thus, by using the fingerprint sensor, the system will become more secure for the users.

The fingerprint technique is similar to fingerprint pattern recognition in image processing. However, the terms fingerprint here refers to radio fingerprints' spatial information. In the previous section, we discussed the general definition of the fingerprint technique. We have also shown the disadvantages in the fingerprint technique implementation, especially in the database construction process. This section will discuss other related fingerprint techniques, their disadvantages, and solutions. The illustration of the fingerprint technique process can be depicted in Figure 3.

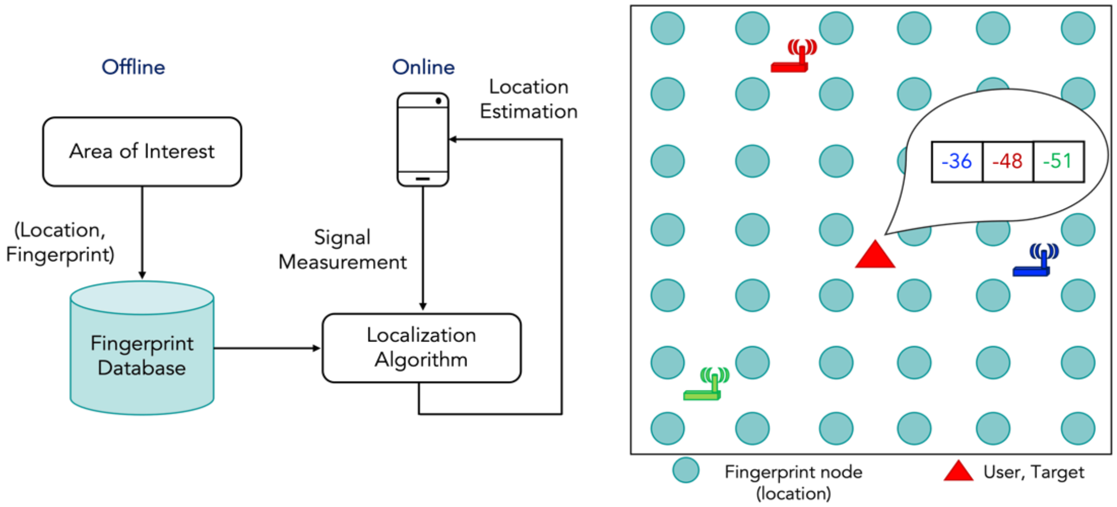


Figure 3. Fingerprint technique.

The fingerprint technique works by constructing the database with spatial information needed to estimate target or object position. This spatial information is the location of the fingerprint point with its corresponding parameter. In this Figure 3, the corresponding parameter is RSSI values from three reference nodes/beacons. The fingerprint technique applies the two phases; the first phase is the offline phase for storing the spatial information as the database, and the second is the signal measurement by an object to be localized by a localization algorithm or pattern matching algorithm. For our proposal, the detail of our fingerprint technique workflow is depicted in Figure 4.

The known fingerprint location depicted in Figure 3 shown in the RSSI-based fingerprint technique in Figure 4 as the $FP\ Location$ 1, $FP\ Location$ 2, … , $FP\ Location\ N$ will collect the RSSI from the reference nodes i to M, with i = 1,2, … , M as the FDi, … . , FDM. After all, the fingerprint points and their corresponding RSSI values are stored in the database. The online phase is when the target calculates the RSSI values from the same i to M nodes and comparing to those in database by applying a pattern matching algorithm. In our proposal, we propose applying the minimum Euclidean distance (MED) algorithm to find the similarity between target parameter, T and the FDi, … . , FDM in the database. Once the i^{th} fingerprint database, which has similar RSSI values, is successfully concluded, the correspondence location of this i^{th} database, the i^{th} FP Location is estimated as the target or object location.

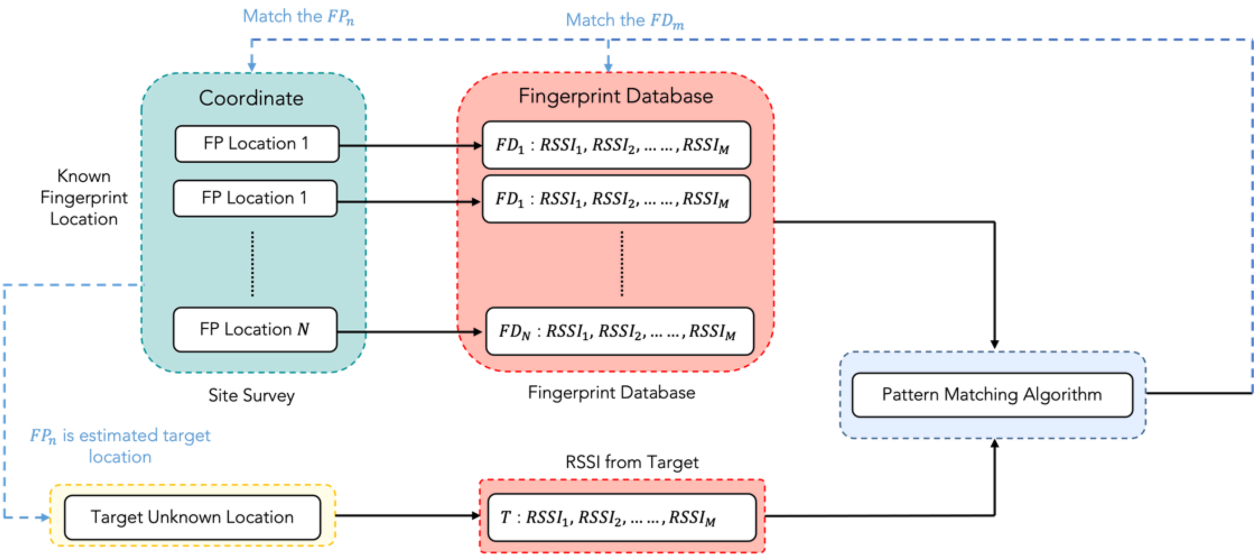


Figure 4. RSSI-based fingerprint technique.

The fingerprint technique performance metrics, i.e., accuracy and precision, are heavily dependent on the quality of the database. One of the factors is the density of the database. However, there will be more burden processes in the offline phase to get a denser database. Some challenges and their prospective solutions can be seen in Figure 5.

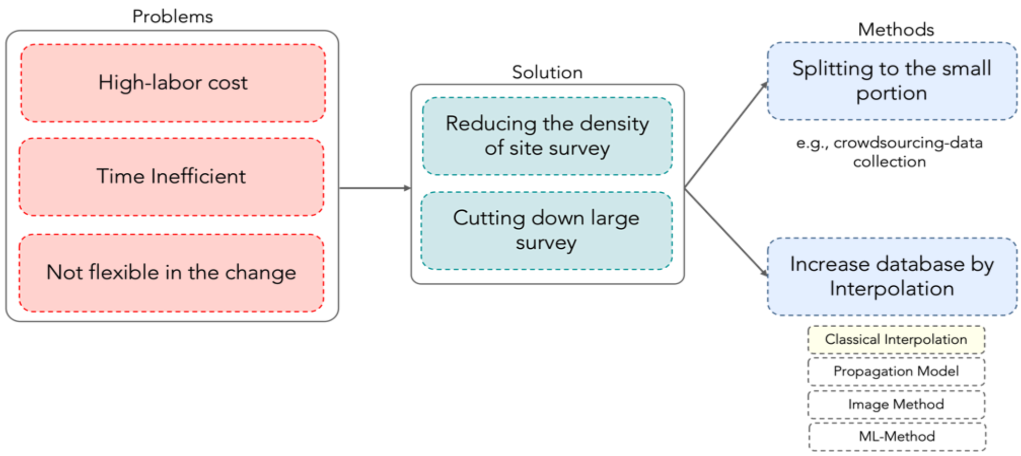


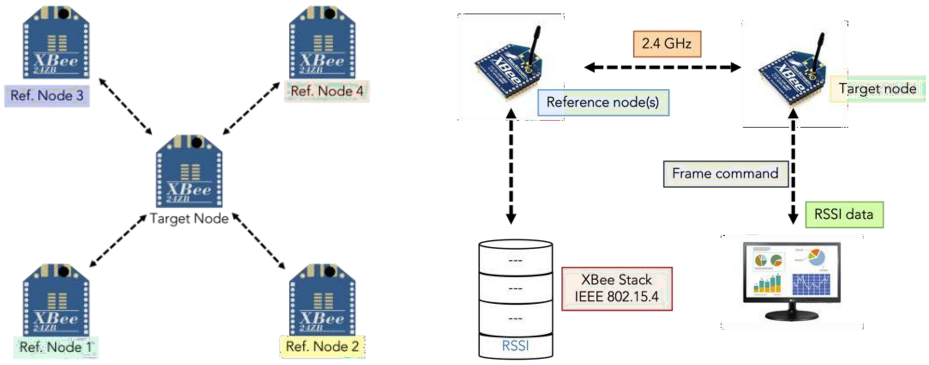
Figure 5. Challenges and prospective solutions of the fingerprint technique.

Cutting down extensive surveys or the database's density improvement can be one of the solutions [39]. Some proposals on the propagation model-based database enhancement have a drawback in the high cost of the channel measurement device and measurement campaigns [40–42]. The other approaches are by using the image method in which the grayscale image of the image processing method is applied to convert the RSSI values in the area of interest [38]. The imaging method is promising but highly complex in the super-resolution image conversion from the RSSI to grayscale image conversion. The ML-based method for database augmentation has abundantly proposed, and because of the threshold in the number of data required in some ML techniques, the complexity is also high [22, 43, 44]. As far as the author is concerned, there are few, or there is no attempt yet in using the classical interpolation technique for this database synthesis. Therefore, this paper proposes utilizing the classical interpolation and regression technique with low complexity and achieving acceptable performance results.

3- MATERIAL AND METHODS

3-1- Wireless Sensor Networks (WSNs) using ZigBee Standard

The ZigBee device, XBee-24ZB, is utilized for both the target and reference nodes in our WSNs setup. We apply the topology star for the localization system. Here, the sink node is the target located inside the area of interest, while the reference nodes are in the corners of the area of interest. Figure 6 shows the illustration of our WSNs based on ZigBee [17, 45].



(a) (b)

Figure 6. a) Topology star illustration, b) The RSSI acquiring process.

The RSSI acquiring process starts with request from the sink node/target to the reference nodes. The reference nodes then get the RSSI packet request by sending back the RSSI from the XBee stack to the sink node. The sink node then translates the packet and store the RSSI values to the storing device/personal computer.

3-2- Interpolation and Regression Technique

The interpolation technique predicts a point or a value between two or more known points/values. The basic interpolation techniques are divided into linear and polynomial interpolation. The linear interpolation technique predicts the values between two points, while interpolation can predict specific points using several known data points.

Bilinear Interpolation Teukolsky et al. [46]

a. Take example of a simple linear interpolation for two data points, y_1 and y_2, as in Figure 7.

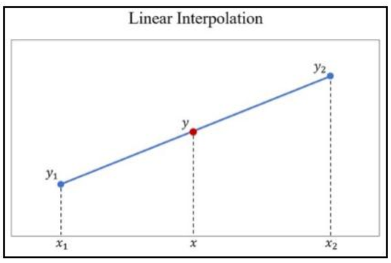


Figure 7. Linear interpolation illustration.

The y between y_1 and y_2 can be predicted as in Equation 2;

$$y = \frac{x - x1}{x2 - x1}(y2 - y1) + y1 \quad (2)$$

b. Bilinear interpolation extends a linear interpolation on two cartesian coordinate axes, the x-axis and the y-axis. Suppose there are 4 points $(x1, y1)$, $(x2, y1)$, $(x2, y2)$, dan $(x1, y2)$, having functions of $(x1, y1)$, $(x2, y1)$, $(x2, y2)$, and $f(x1, y2)$. We can use these values to predict the (x, y) value at (x, y) as illustrated in Figure 8.

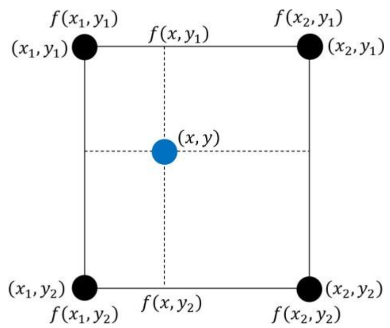


Figure 8. Bilinear interpolation illustration.

The value at (x, y) can be predicted by first, we first interpolate the x-axis as

$$f(x, y_1) = \frac{x - x1}{x2 - x1}(f(x_2, y_1) - f(x_1, y_1)) + f(x_1, y_1). \quad (3)$$

$$f(x, y_2) = \frac{x - x1}{x2 - x1}(f(x_2, y_2) - f(x_1, y_2)) + f(x_1, y_2). \quad (4)$$

Thus, by using Equations 3 and 4, we can the proceed to y-axis interpolation by Equation 5.

$$(x, y) = \frac{y - y1}{y2 - y1}(f(x, y_2) - f(x, y_1)) + f(x, y_1). \quad (5)$$

Polynomial Interpolation Kiusalaas (2013) [47]

c. If we have more than two data points, polynomial interpolation can predict the data points within the data range. Polynomial interpolation can also handle non-linear pattern data. Neville interpolation is one of the polynomial interpolation methods.

d. Predicted value within the data range can be obtained using general equation shown in Equation 6.

$$P_k[x_i, x_{i+1}, \dots, x_{i+k}] = \frac{(x - x_{i+k}) P_{k-1}[x_i, x_{i+1}, \dots, x_{i+k-1}] + (x_i - x) P_{k-1}[x_{i+1}, x_{i+2}, \dots, x_{i+k}]}{x_i - x_{i+k}}. \quad (6)$$

We can solve the Equation 6 by using Neville Method Settlement detailed in Table 2 and k is the degree of the polynomial equation.

Table 2. Neville Method Settlement.

	$k = 0$	$k = 1$	$k = 2$	$k = 3$
$x0$	$P0[x0] = y0$	$P1[x0, x1]$	$P2[x0, x1, x2]$	$P3[x0, x1, x2, x3]$
$x1$	$P0[x1] = y1$	$P1[x1, x2]$	$P2[x1, x2, x3]$	
$x2$	$P0[x2] = y2$	$P1[x2, x3]$		
$x3$	$P0[x3] = y3$			

e. There are three solution steps for solving the Neville interpolation.

1. *First step:* Performing the interpolation of degree 1 by using the Equation 7.

$$P_1[x_0, x_1] = \frac{(x - x1)P0[x0] + (x0 - x)P0[x1]}{x0 - x1}. \quad (7)$$

2. *Second step:* Performing the interpolation of degree 2 by using the Equation 8.

$$P_2[x_0, x_1, x_2] = \frac{(x - x2)P1[x0, x1] + (x0 - x)P1[x1, x2]}{x0 - x2}. \quad (8)$$

3. *Third step:* Performing the interpolation of degree 3 by using the Equation 9.

$$P_3[x_0, x_1, x_2, x_3] = \frac{(x - x3)P2[x0, x1, x2] + (x0 - x)P2[x1, x2, x3]}{x0 - x3}. \quad (9)$$

Polynomial Regression Ostertagová (2012) [48]

f. Polynomial regression is a multiple regression with one independent variable. In one variable polynomial regression equation shown as Equation 10, x is expressed as an independent variable.

$$y = a0 + a1xi + a2xi^2 + \cdots + anxi^k + ei,\ i = 1, 2, \dots, k. \quad (10)$$

3-3- Pattern Matching and Performance Metrics

We applied a simple Minimum Euclidean Distance (MED) to match the RSSI of the target with those in the database. To validate our localization system's accuracy, we used average distance error (ADE) from all errors of target positions in each scenario, both for 2D and 3D environments [5]. Nevertheless, first, we evaluate the RSSI error value between the actual and predicted RSSI form interpolation and regression techniques by Equation 11. This error will be presented in the results and discussion for only 2D, as the RSSI maps from 2D will be easier to interpret. The absolute symbol shows the RSSI discrepancy on the values (because RSSI in dBm is in a negative form). The RSSI error is considered in some points of the 2D fingerprint database to ensure that the proposed method is visible to apply by observing these errors.

$$RSSI\ error\ (dBm) = |RSSI\ value(dBm)actual - RSSI\ value(dBm)predicted|. \quad (11)$$

MED as pattern matching utilizes the Euclidean distance to measure the distance from two points; in our case, the Euclidean distance is the distance of RSSI values in the database and RSSI values of the target. The tiniest error or the minimum error of RSSI by their Euclidean distance for specific fingerprint location is assigned as the target's location.

$$Euclidean\ distance\ (m) = \sqrt{(RSSI_{database} - RSSI_{target})^2}. \quad (12)$$

After we obtain the predicted location based on the similarity of the target and database by the MED, we evaluate the accuracy by the error in the meter of the predicted and the actual target location. Thus, the system's performance metric in our proposal is how the error of the target location prediction, $xpredicted$ and $y_{predicted}$, is compared to the actual target location, $xactual$ and y_{actual}. Suppose there are L number of target locations, we utilized ADE as the performance metric, the mean value of all Euclidean distances for all target locations. If the position of the target node as $i = 1,2, \ldots, L$, we can express the ADE as in Equation 13 [49].

$$ADE = \frac{1}{L}\sum_{i=1}^{L}\sqrt{(xactual,i - xpredicted,i)^2 + (yactual,i - ypredicted,i)^2} \qquad (13)$$

3-4- Methodology Flow Diagram

We design the measurement by applying wireless sensor networks (WSNs)-based data collection. For the flow diagram, we divide into two flows; first flow is an interpolation confirmation step where some of the database points are replaced by the interpolated RSSI values, and the second flow is the interpolation implementation step where the synthetic database from interpolation and regression are combined and evaluated by the localization performance. We start with the confirmation step, where the preparation is to design the measurement for two-dimensional (2D) and three- dimensional (3D) environments, including how we propose the grids of fingerprint and target locations in the area of interest.

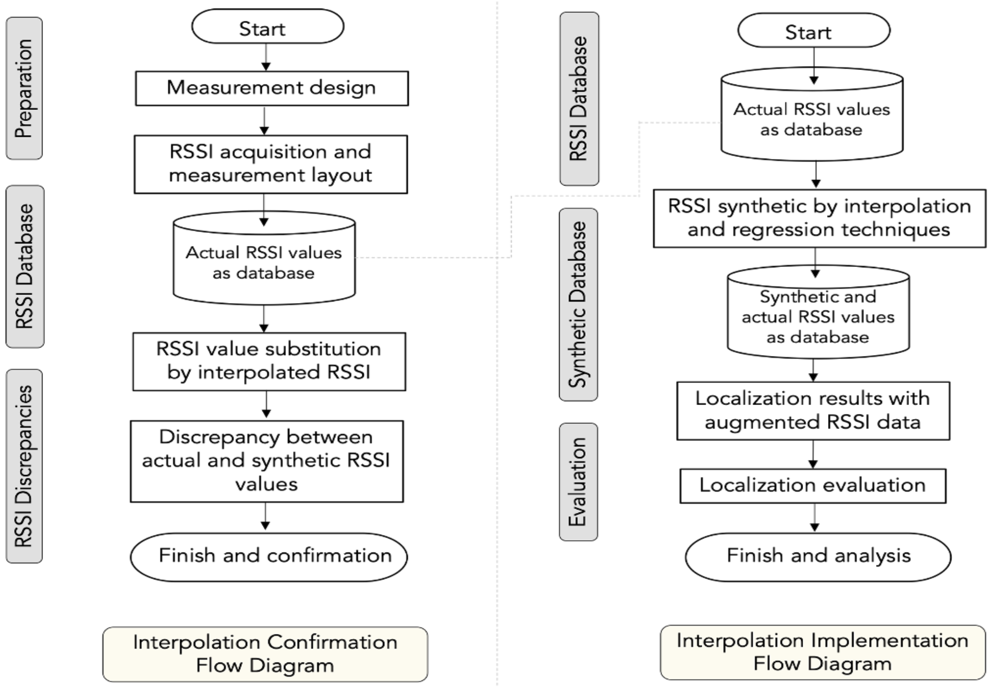


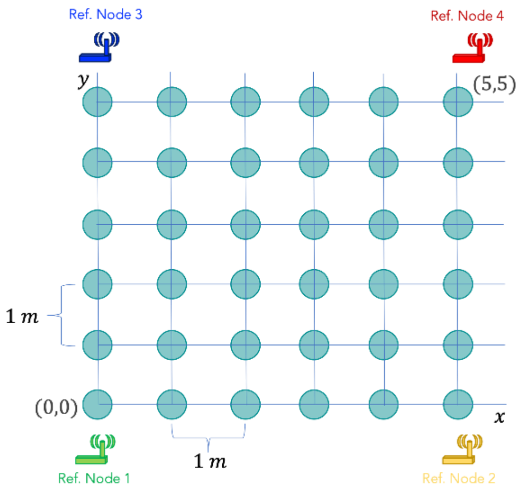
Figure 9. Flowchart of the research Methodology.

3-5- Measurement Campaign

We consider the two 2D and 3D environments for indoor localization system deployment. The first environment is the lobby environment in our department, and we consider only a 2D scenario. For the 3D environment, we simulate the actual multi-story building by using a 4-level bookshelf for elevation parameters. The ZigBee devices are configured as reference nodes and target nodes; both acted as transceivers and were configured as a star topology.

The 2D Measurement Campaign

We conducted a measurement campaign for the 2D environment in the lobby of our building for a 5 × 5 m^2 area of interest. The layout and the actual setup of the measurement campaign for the 2D case are shown in Figure 8. We consider using four reference nodes in a 2D environment; this selection assumes that each reference node will be located in each corner of the rectangular shape measurement area. Thus, the interpolation techniques will likely have a better linear relationship between each reference node and its corresponding distance. We have applied more reference nodes for fingerprint-based in the same measurement area and can be found in [50–52]. Some results suggested that more reference nodes yield better accuracy (scalability). However, our previous publications do not consider interpolation techniques.



The 3D Measurement Campaign

To simulate the multi-story building with different levels/floors, we used a four-level wooden bookshelf. We also considered three scenarios in this 3D environment; clean environment, human body effects, and interference objects of books. The body effect is considered by a standing person in 1m distance from the bookshelf. The bookshelf's dimensions are 92 × 25 × 152 cm^3 and the detailed dimension and illustration are depicted in Figure 10.

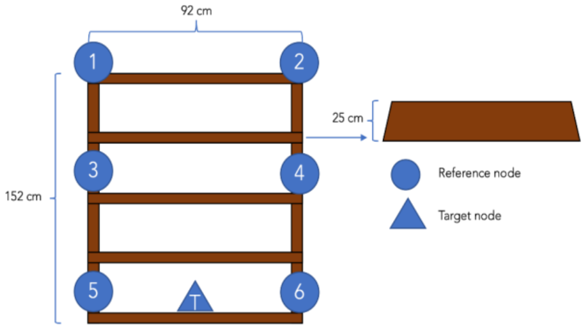
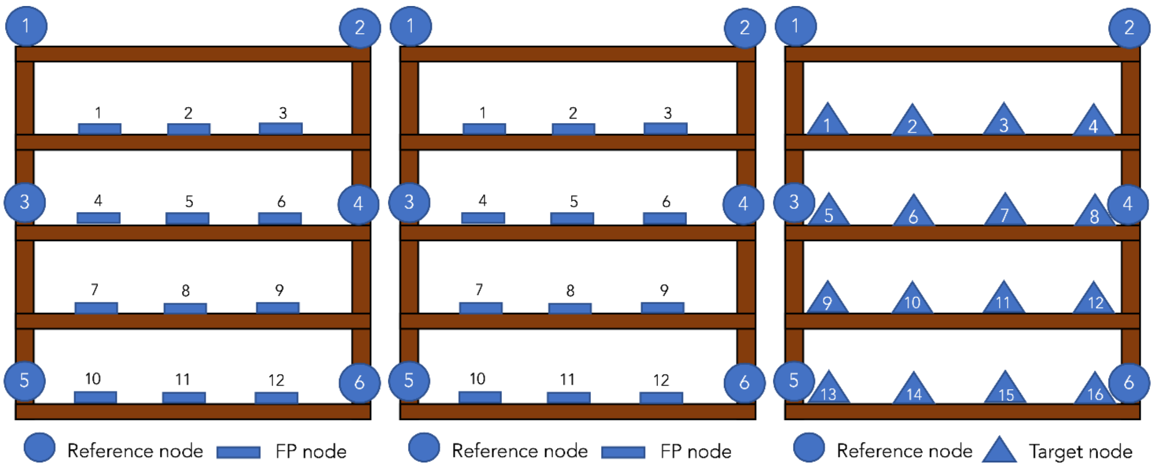
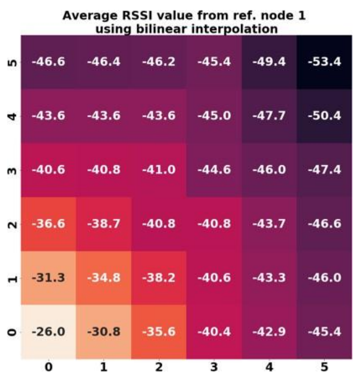


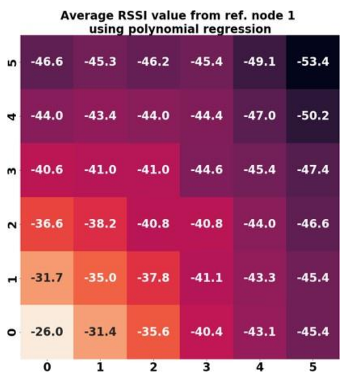
Figure 10. The illustration of a 3D environment in a bookshelf.

The fingerprint node (FP node) and target node can be illustrated in Figure 11. Unlike in the 2D environment, in the 3D environment, we employed six reference nodes. This selection ensures that the elevation properties are well represented as the XBee antenna directional pattern is more horizontally shaped. Moreover, we have only six reference nodes in the dataset's measurement setup and have not yet published the approach. The target, T is placed in the different levels of the bookshelf statically.

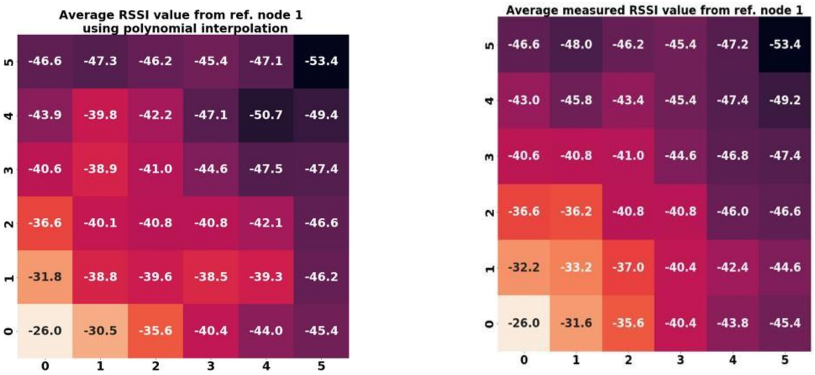




(a)



(b)



(C) (D)

Figure 16. RSSI heatmap reference node 1: a) Bilinear inter., b) Polynomial reg., c) Polynomial inter., d) Measurement.

Figure 17 shows the RSSI values discrepancy for reference node 2. It shows that the majority of the RSSI values prediction using the polynomial interpolation produces a more significant error than polynomial regression and bilinear interpolation. The most significant prediction errors are at orders 11, 14, and 15, with 3.27 dBm, 3.83 dBm, and 4.44 dBm. On the other hand, the polynomial regression method produces the most significant errors at orders 9, 12, and 19, with each point having errors of 2.38 dBm, 2.73 dBm, and 3.14 dBm, respectively. At the same time, the most negligible error results in the prediction of RSSI values using polynomial interpolation, polynomial regression, and bilinear interpolation are 0.09 dBm, 0.02 dBm, and 0.10 dBm. Figure 18 depicts the RSSI heatmap comparison for reference node 2. The brighter colour gradation is origin from the bottom right as the reference node two positions in the measurement.

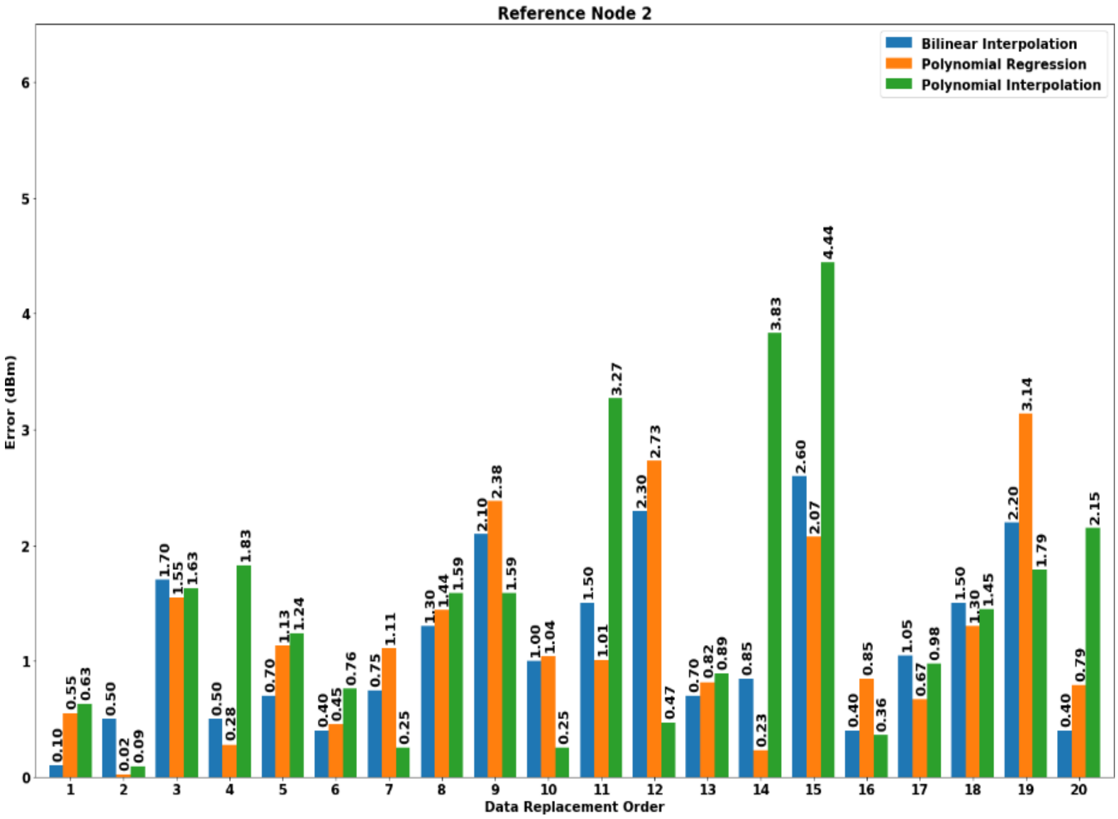


Figure 17. RSSI values error for reference node 2

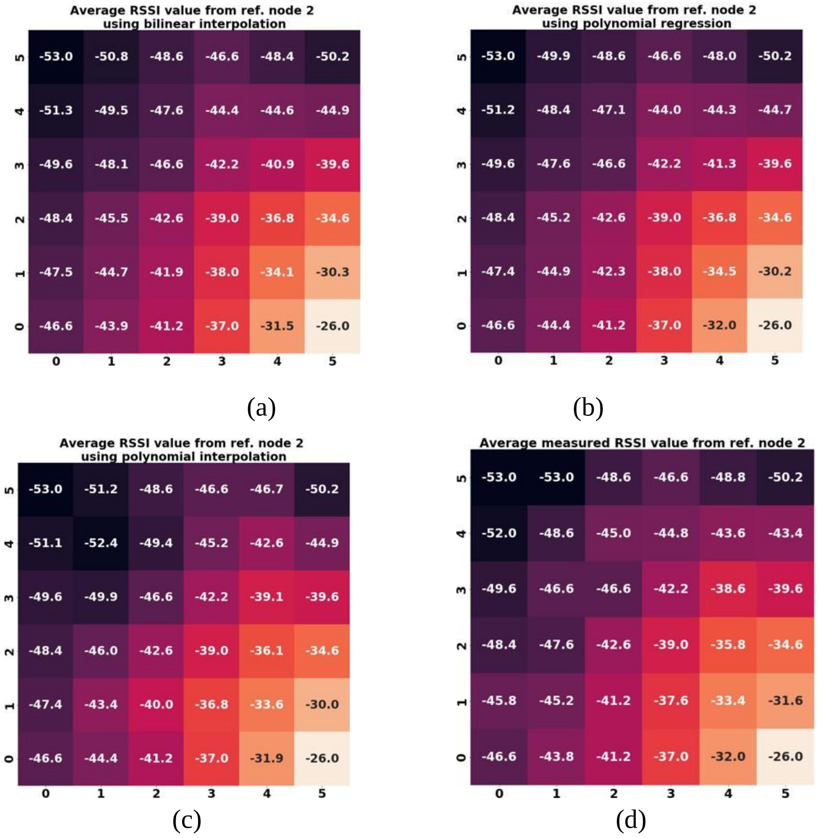
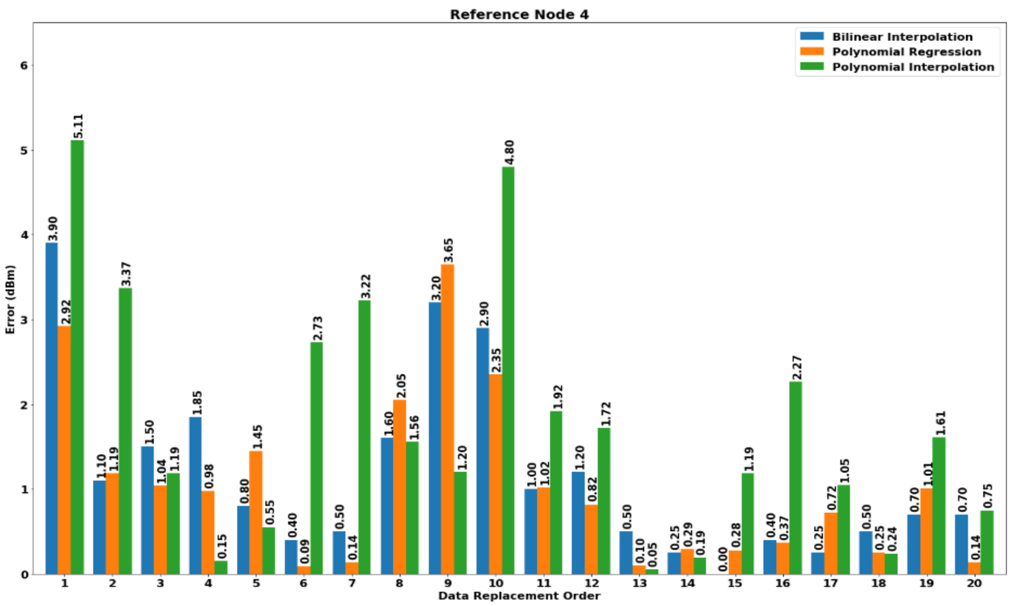


Figure 18. RSSI heatmap reference node 2. a) Bilinear inter., b) Polynomial reg., c) Polynomial inter., d) Measurement

Figure 19 shows the RSSI error values discrepancy comparison between the three proposed methods and actual RSSI values from measurement. We found that the most significant errors are at orders 2, 8, and 18, where the replacement of RSSI values using the polynomial interpolation method produces errors of 3.89 dBm, 4.53 dBm, and 3.47 dBm. In reference node 3, the minor RSSI error from polynomial interpolation, polynomial regression, and bilinear interpolation are 0.06 dBm, 0.08 dBm, and 0.10 dBm. Overall, the mean RSSI prediction error using polynomial interpolation, polynomial regression, and bilinear interpolation methods are 1.81 dBm, 0.95 dBm, and 0.81 dBm, respectively. To observe the precise RSSI values distribution, Figure 20 shows the RSSI heatmap comparison for reference node three, and the brighter to darker color gradation originates from the upper left (the position of reference node 3 in measurement).



The RSSI values discrepancy from three methods compared to the measurement for the reference node four is shown in Figure 21. The results also show that polynomial interpolation produces the most significant error with the largest error at the first order with an error value of 5.11 dBm. In comparison, the polynomial regression method and bilinear interpolation at the same point resulted in an error of 2.92 dBm and 3.90 dBm, respectively. By implementing polynomial interpolation, polynomial regression, and bilinear interpolation methods, the mean RSSI prediction error is 1.74 dBm, 1.16 dBm, and 1.04 dBm, respectively. From these results, we can conclude that the bilinear interpolation performed relatively better than polynomial interpolation and regression. Figure 22 shows the RSSI heatmap comparison for reference node 4.

Localization Results

The synthetic fingerprint databases are labelled as and fp+intBil_db, fp+regPoly_db, fp+intPoly_db, for bilinear interpolation, regression polynomial, and polynomial interpolation, respectively. For 2D environment, we consider the 4 types of target position namely as target type 1–4. We compare the localization results, fp+intBil_db, fp+regPoly_db, fp+intPoly_db databases with the actual measured fingerprint database, fp_meas_db. For the target types 1-4, the comparison of ADE results can be seen in Figure 23. The database used in the positioning results is illustrated in Figure 14. By implementing polynomial interpolation, polynomial regression, and bilinear interpolation methods, new data is added so that a database is formed with the amount of data at points where the RSSI value is not measured. The new database is then labelled as fp+intPoly_db, fp+regPoly_db, and fp+intBil_db. The three databases are then compared with the fp_meas_db database.

Figure 23. Localization results comparison, a) target type 1, b) target type 2, c) target type 3, and d) target type 4.

For target type 1 in Figure 23.a)., the ADE values obtained using fp+intBil_db, fp+regPoly_db, fp+intPoly_db, and fp_meas_db are 0.29 m, 0.34 m, 0.35 m and 0.39, respectively. These results show that applying all the interpolation and regression techniques can reduce the average error of 0.1m in the target type 1 scenario. The database scarcity, or in other words, by reducing the number of database size in the offline database and we compensated with the classical interpolation and regression technique, has been proven to improve the localization performance.

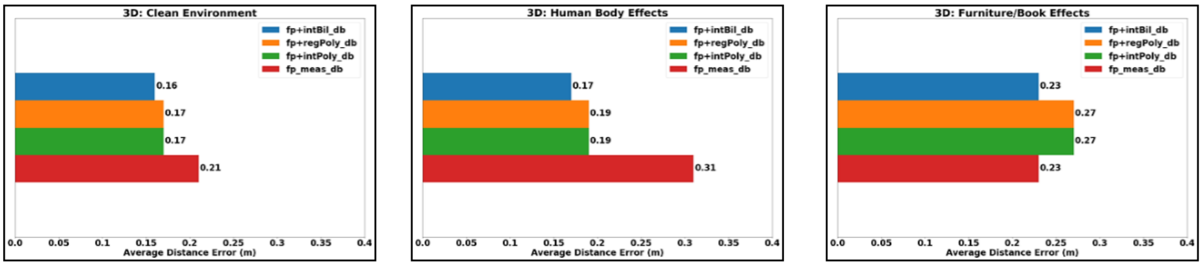
For target type 2, the ADE values obtained using fp+intBil_db, fp+regPoly_db, fp+intPoly_db, and fp_meas_db are 0.42, 0.48, 0.72, and 0.57 m as shown in Figure 23.b). Positioning results with reduced lattice size database using polynomial interpolation method resulted in higher ADE values with an increase of 0.15 m. Meanwhile, polynomial regression and bilinear interpolation reduced ADE by 0.09 and 0.15 m, respectively.

For the horizontal movement of target type 3, the positioning results using fp+intBil_db, fp+regPoly_db, fp+intPoly_db, and fp_meas_db is shown in Figure 23.c). Only fp+intPoly_db obtained the less accurate results compared to fp+regPoly_db, and fp+intBil_db. The 0.2 m error improvement was achieved by the fp+intBil_db, which so far achieved the best results. The last scenario is the target type 4 in the vertical movement. Figure 23.d) shows the results of improvement by all types of the additional database; fp+intBil_db, fp+regPoly_db, and fp+intPoly_db. These results proved that the interpolation and regression for enhancing the database by adding the artificial points between actual measurement points were successfully implemented and validated.

We would like to explore how the RSSI values from 4 reference nodes are placed in each corner of the rectangular shape measurement area by selecting these four target position types. In target types 1 and 2, the diagonal is easier to imagine that the gradation of the RSSI values will follow the diagonal line. However, as the position is somehow in the middle of the area of interest for horizontal and vertical target type positions, it is expected to make the estimation more prone to error, as observed in Figures 23.c and 23.d.

4-2- 3D Environment

Unlike 2D environment, results for the 3D environment were analyzed and validated by considering three types of scenarios in the bookshelf; the first scenario is a clean environment when there is no obstruction objects on the bookshelf, the second scenario, during the measurement campaign, the first author standing close to the bookshelf (0.5 m) facing front. The third scenario is to place books and other stationaries between the reference nodes in every level of the bookshelf having the reference nodes. We expected that by putting a kind of 'noise' in our measurement, we could observe the improvement when applying the interpolation and regression technique. Fingerprint databases generated from the replacement of measurement data are also labeled as fp+intBil_db, fp+regPoly_db, fp+intPoly_db, and fp_meas_db for bilinear interpolation, polynomial regression, polynomial interpolation, and actual fingerprint database, respectively. These techniques not only augment the data inside the data range but are also used to generate data outside the range (i.e., extrapolation). Figure 24 shows the results for three scenarios in the 3D environment.



(a) (b) (c)

Figure 24. Localization results comparison, a) clean environment, b) human body effects, c) furniture/book effects

A clean environment can also be called ideal conditions that we set up without any obstacles in the area of interest. We observe that all enhancement methods can reduce ADE compared to actual fingerprints in this condition. Data quantity enhancement using bilinear interpolation has the most significant ADE reduction. Polynomial interpolation and regression have the same results as shown in Figure 24.a). In the human body effects scenario, we can observe that the localization results were less accurate than the clean environment.

For human body effects, an approach of interpolation applying Kriging for wireless body area network (WBAN) indoor localization has been proposed in [55]. Kriging is a mechanism to predict spatial information for estimating a value at a specific location using actual locations. The paper discussed the possibility of implementing Kriging interpolation to the constructed artificial database and found that by Kriged fingerprint, the accuracy is improved. In this paper, we did not analyze the propagation mechanism. However, we expect that a person standing near the bookshelf will likely affect the localization results. From Figure 24.b), there was a slightly added measured error for 3~8 cm. Polynomial interpolation and regression achieved the same result, but they were less accurate than the actual measurement fingerprint data. Bilinear interpolation has a very tiny accuracy margin better than measurement fingerprint database.For the furniture effects scenario, all enhancement methods can reduce ADE almost twice than using measurement fingerprint data. In other case studies, positioning results using polynomial interpolation and regression gave the same results in furniture obstacles. The bilinear interpolation always gave the minor error position. The object's existence between reference nodes created shadowing effects that even the synthesis database cannot handle. From these results, we can conclude that in the 3D cases, the signal obstruction in the horizontal direction affected the localization results because of the random signal fluctuation. Table 5 compares the results between our proposal and other studies. We emphasized the interpolation technique, the environment, and the performance metric.

Table 5. Performance Comparison with Other studies.

No.	Indoor Localization	Parameter and Technology	Techniques	Environment	Performance Metric
1.	The proposed method	RSSI, ZigBee	Bilinear, polynomial and polynomial regression	Real environment: 2D and 3D	Interpolation reduces estimated error up to 0.2m for 2D and 0.13m for 3D.
2.	Jingxue Bi, et.al. [33]	RSSI, Wi-Fi	Crowdsourcing and interpolation	Real environment: total 3200 m^2	Improving the accuracy results up to 20% manual fingerprint, by using only 25% reference

				area.	points.
3.	Yanwei Li, et.al. [32]	RSSI, Wi-Fi	Classic narrowband path loss model	Dataset from [56]	Refined map with size 0.75×0.75 m^2 (grid) improve accuracy up to 63% from 3.6164 m to 1.3294 m

CONCLUSIONS

The database enhancement by applying interpolation and regression techniques to tackle one of the drawbacks of the fingerprint technique is presented. The actual measurement campaign in 2D and 3D environments was conducted, and RSSI values were used for the database fingerprint for both environments. We used a device for measurement based on IoT technology which cost-effective and straightforward implementation of ZigBee standards. The WSNs-based indoor localization is built and validated by some scenarios; the four target type placements, namely diagonal positions both left and right, the horizontal and vertical position for the 2D environment, while three scenarios tested 3D environment; clean, human body effects, and furniture effects. From all scenarios, almost all results agreed that enhancing the database to expand the database grids by artificial data can reduce localization prediction error. From all scenarios, the simple bilinear interpolation stood out as the best accuracy performance. For the 2D environment, it can reduce to 0.2 m error, and 0.13 m for the 3D environment, which is in this environment the grid size is 0.22 m (22 cm). The performance improvement by replacing and adding the synthesis or artificial database can facilitate actual implementation. Thus, the sparsity issues in the fingerprint database can be tackled, and the offline database construction will be less burden.

Our research interest is in fingerprint database enhancement; for the next step, we consider applying machine or deep learning-based both supervised and unsupervised learning, both offline and online phase to increase the performance metric of indoor localization system and its flexibility and adaptability in the dynamic indoor environment.

REFERENCES

[1] Yassin, Ali, Youssef Nasser, Mariette Awad, Ahmed Al-Dubai, Ran Liu, Chau Yuen, Ronald Raulefs, and Elias Aboutanios. "Recent Advances in Indoor Localization: A Survey on Theoretical Approaches and Applications." IEEE Communications Surveys and Tutorials 19, no. 2 (2017): 1327–46. doi:10.1109/COMST.2016.2632427.

[2] He, Suining, and S. H.Gary Chan. "Wi-Fi Fingerprint-Based Indoor Positioning: Recent Advances and Comparisons." IEEE Communications Surveys and Tutorials 18, no. 1 (2016): 466–90. doi:10.1109/COMST.2015.2464084.

[3] Yang, Chouchang, and Huai Rong Shao. "WiFi-Based Indoor Positioning." IEEE Communications Magazine 53, no. 3 (2015): 150–57. doi:10.1109/MCOM.2015.7060497.

[4] Yan, Dongmei, Bing Kang, Hui Zhong, and Ruolin Wang. "Research on Positioning System Based on Zigbee Communication." In Proceedings of 2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference, IAEAC 2018, 1027–30, 2018. doi:10.1109/IAEAC.2018.8577263.

[5] Jiang, Jehn Ruey, Hanas Subakti, and Hui Sung Liang. "Fingerprint Feature Extraction for Indoor Localization†." Sensors 21, no. 16 (2021). doi:10.3390/s21165434.

[6] Yassin, Ali, Youssef Nasser, Mariette Awad, Ahmed Al-Dubai, Ran Liu, Chau Yuen, Ronald Raulefs, and Elias Aboutanios. "Recent Advances in Indoor Localization: A Survey on Theoretical Approaches and Applications." IEEE Communications Surveys and Tutorials 19, no. 2 (2017): 1327–46. doi:10.1109/COMST.2016.2632427.

[7] Song, Zhenlong, Gangyi Jiang, and Chao Huang. "A Survey on Indoor Positioning Technologies." Communications in Computer and Information Science 164 CCIS (2011): 198–206. doi:10.1007/978-3-642-24999-0_28.

[8] Mehta, Rishika, Jyoti Sahni, and Kavita Khanna. "Internet of Things: Vision, Applications and Challenges." Procedia Computer Science 132 (2018): 1263–69. doi:10.1016/j.procs.2018.05.042.

[9] Carlos-Mancilla, Miriam, Ernesto López-Mellado, and Mario Siller. "Wireless Sensor Networks Formation: Approaches and Techniques." Journal of Sensors (2016).

doi:10.1155/2016/2081902.

[10] Willig, Andreas. “Recent and Emerging Topics in Wireless Industrial Communications: A Selection.” IEEE Transactions on Industrial Informatics 4, no. 2 (May 2008): 102–124. doi:10.1109/tii.2008.923194.

[11] Yiu, Simon, Marzieh Dashti, Holger Claussen, and Fernando Perez-Cruz. “Wireless RSSI Fingerprinting Localization.” Signal Processing 131 (2017): 235–44. doi:10.1016/j.sigpro.2016.07.005.

[12] Mahler, Tom, and Ido Bayda. Product Design and Distance Measurement Using RSSI and LQI. Israel: Tel-Aviv University, 2013.

[13] Parameswaran, Ambili Thottam, Mohammad Iftekhar Husain, and Shambhu Upadhyaya. “Is RSSI a Reliable Parameter in Sensor Localization Algorithms - An Experimental Study.” In IEEE International Symposium on Reliable Distributed Systems, 1–5, 2009.

[14] Sadowski, Sebastian, and Petros Spachos. “RSSI-Based Indoor Localization with the Internet of Things.” IEEE Access 6 (2018): 30149–61. doi:10.1109/ACCESS.2018.2843325.

[15] Li, Guoquan, Enxu Geng, Zhouyang Ye, Yongjun Xu, Jinzhao Lin, and Yu Pang. “Indoor Positioning Algorithm Based on the Improved Rssi Distance Model.” Sensors (Switzerland) 18, no. 9 (2018): 1–15. doi:10.3390/s18092820.

[16] Vo, Quoc Duy, and Pradipta De. “A Survey of Fingerprint-Based Outdoor Localization.” IEEE Communications Surveys and Tutorials 18, no. 1 (2016): 491–506. doi:10.1109/COMST.2015.2448632.

[17] Hatem, Elias, Sergio Fortes, Elizabeth Colin, Sara Abou-Chakra, Jean Marc Laheurte, and Bachar El-Hassan. “Accurate and Low-Complexity Auto-Fingerprinting for Enhanced Reliability of Indoor Localization Systems.” Sensors 21, no. 16 (2021): 5346. doi:10.3390/s21165346.

[18] Ji, Wenqing, Kun Zhao, Zhengqi Zheng, Chao Yu, and Shuai Huang. “Multivariable Fingerprints with Random Forest Variable Selection for Indoor Positioning System.” IEEE Sensors Journal, 2021, 1–1. doi:10.1109/jsen.2021.3103863.

[19] Nessa, Ahasanun, Bhagawat Adhikari, Fatima Hussain, and Xavier N. Fernando. “A Survey of Machine Learning for Indoor Positioning.” IEEE Access 8 (2020): 214945–65. doi:10.1109/ACCESS.2020.3039271.

[20] Wang, Liping, Saideep Tiku, and Sudeep Pasricha. “CHISEL: Compression-Aware High-Accuracy Embedded Indoor Localization with Deep Learning.” IEEE Embedded Systems Letters 0663, no. c (2021): 10–13. doi:10.1109/LES.2021.3094965.

[21] Chen, Guoliang, Xiaolin Meng, Yunjia Wang, Yanzhe Zhang, Peng Tian, and Huachao Yang. “Integrated WiFi/PDR/Smartphone Using an Unscented Kalman Filter Algorithm for 3D Indoor Localization.” Sensors (Switzerland) 15, no. 9 (2015): 24595–614. doi:10.3390/s150924595.

[22] Geok, Tan Kim, Khaing Zar Aung, Moe Sandar Aung, Min Thu Soe, Azlan Abdaziz, Chia Pao Liew, Ferdous Hossain, Chih P. Tso, and Wong Hin Yong. “Review of Indoor Positioning: Radio Wave Technology.” Applied Sciences (Switzerland) 11, no. 1 (2021): 1–44. doi:10.3390/app11010279.

[23] Cheng, Yen Kai, Hsin Jui Chou, and Ronald Y. Chang. “Machine-Learning Indoor Localization with Access Point Selection and Signal Strength Reconstruction.” IEEE Vehicular Technology Conference 2016-July (2016): 2016–. doi:10.1109/VTCSpring.2016.7504333.

[24] Cherntanomwong, Panarat, and Dwi Joko Suroso. “Indoor Localization System Using Wireless Sensor Networks for Stationary and Moving Target.” Conf. Information, Commun. Signal Process 1 (2012): 1–5. doi:10.1109/icics.2011.6173554.

[25] Teukolsky, Saul A., Brian P. Flannery, W. H. Press, and W. T. Vetterling. "Numerical recipes in C." SMR 693, no. 1 (1992): 59-70.

[26] Kiusalaas, Jaan. “Numerical Methods in Engineering with Python 3.” In Numerical Methods in Engineering with Python 3, 3, 2013. doi:10.1017/cbo9781139523899.

[27] Ostertagová, Eva. “Modelling Using Polynomial Regression.” Procedia Engineering 48 (2012): 500–506. doi:10.1016/j.proeng.2012.09.545.

[28] Suroso, Dwi Joko, Panarat Cherntanomwong, Pitikhate Sooraksa, and Jun Ichi Takada. “Location Fingerprint Technique Using Fuzzy C-Means Clustering Algorithm for Indoor Localization.” In IEEE Region 10 Annual International Conference,

Proceedings/TENCON, 88–92, 2011. doi:10.1109/TENCON.2011.6129069.

[29] Suroso, Dwi Joko, Panarat Cherntanomwong, Pitikhate Sooraksa, and Jun Ichi Takada. "Fingerprint-Based Technique for Indoor Localization in Wireless Sensor Networks Using Fuzzy C-Means Clustering Algorithm." In 2011 International Symposium on Intelligent Signal Processing and Communications Systems: "The Decade of Intelligent and Green Signal Processing and Communications", ISPACS 2011, 1–5, 2011. doi:10.1109/ISPACS.2011.6146167.

[30] Cherntanomwong, Panarat, and Dwi Joko Suroso. "Indoor Localization System Using Wireless Sensor Networks for Stationary and Moving Target." Conf. Information, Commun. Signal Process 1 (2012): 1–5. doi:10.1109/icics.2011.6173554.

[31] Suroso, Dwi Joko, Panarat Cherntanomwong, Pitikhate Sooraksa, and Jun Ichi Takada. "Location Fingerprint Technique Using Fuzzy C-Means Clustering Algorithm for Indoor Localization." In IEEE Region 10 Annual International Conference, Proceedings/TENCON, 88–92, 2011. doi:10.1109/TENCON.2011.6129069.

[32] Cherntanomwong, Panarat, Jun Ichi Takada, and Hiroyuki Tsuji. "Signal Subspace Interpolation from Discrete Measurement Samples In Constructing a Database for Location Fingerprint Technique." IEICE Transactions on Communications E92-B, no. 9 (2009): 2922–30. doi:10.1587/transcom.E92.B.2922.