International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 10, No.1 (May-2021)

http://dx.doi.org/10.12785/ijcds/100166

# Random Forest and Interpolation Techniques for Fingerprint-based Indoor Positioning System in Un-ideal Environment

**Dwi Joko Suroso[1], Alvin S. H. Rudianto[1], Muhammad Arifin[2] and Singgih Hawibowo[1]**

*[1] Dept. of Nuclear Engineering and Engineering Physics, Universitas Gadjah Mada, Yogyakarta, Indonesia*
*[2]Radio Engineering,PT. Delameta Bilano, East Jakarta, Indonesia*

*Received 27 Dec. 2020, Revised 27 Mar. 2021, Accepted 03 Apr. 2021, Published 02 May. 2021*

**Abstract:** The development of location-based service (LBS) in outdoor environments relies on Global Positioning System (GPS) technology to determine the location. However, for indoor LBS, GPS is not reliable because it has low accuracy for indoor environments. Therefore, it is necessary to have an indoor positioning system (IPS) for indoor LBS. Wi-Fi-based IPS research is growing along with the many uses and availability of Wi-Fi. For static indoor environments, fingerprinting techniques have better accuracy than distance-based approaches such as trilateration and min-max. The fingerprint technique can also be applied by utilizing simple and straightforward parameters, i.e., received signal strength indicator (RSSI). Some of the fingerprint technique's challenges are the length of time, effort, and high cost of collecting the database. In this study, we apply the interpolation techniques to reduce time and effort in collecting fingerprint data. The interpolation used is the Neville interpolation and Bilinear interpolation. Comparing the positioning results between the classical pattern matching algorithm, minimum Euclidean distance (MED), and modern machine learning-based, the Random Forest algorithm, is discussed in detail. We conducted a measurement campaign in an unideal indoor environment to see how far our proposed method can still handle the fluctuated values of RSSI. From several measurements and scenarios presented in this study, the MED algorithm is still better in accuracy and precision than the random forest algorithm. However, in almost all scenarios, the Random Forest can better perform MED in terms of decreasing the maximum estimated error. The accuracy and precision between MED and random forest are up to 0.5 meters, and the precision difference is up to 20%. Performance improvements due to using database interpolation range from 3% to 30%. The database from the interpolation results is also acceptable in the performance metric of the positioning system. However, using the database from the actual position is better and outperformed the database from interpolation results. The low similarity of actual database and database synthesis from interpolation is due to the non-linear and fluctuated RSSI values in our measurements due to the unstable and time-varying effects.

**Keywords:** Indoor Positioning System, Wi-Fi, Fingerprint Technique, RSSI, Interpolation, Random Forest, MED

## 1. Introduction

At present, communication devices and digital services have become an essential part of human life. The increasing number of mobile and smart devices globally makes it easier for digital service providers to reach users or consumers. These reasons become the primary factors driving the demand and development of location-based services (LBS) [1][2]. The most widely used location determination system is the Global Positioning System (GPS), especially for outdoor environments [3].

As the world is still affected by Coronavirus Disease-2019 (COVID)-19 pandemic, it has been a significant issue in several aspects. Including how people will mostly stay indoors for many activities, most of them also need the LBS for the indoor environment. However, GPS cannot give reliable location determination results for indoor positioning since the GPS signal cannot penetrate walls and other obstacles[4]. Many indoor LBS applications should be developed for location determination systems indoor as the alternative to GPS. Applications of indoor LBS can be but not limited to indoor advertising, information location of products, or persons in supermarkets, airports, and train stations. In the health and rescue sectors, i.e., patient tracking systems in hospitals, tracking firefighters while performing rescue operations indoors, and automatically controlling robot movement in warehouses. With the many potentials for developing LBS in indoor environments, the Indoor Positioning System (IPS) appears essentially for determining the position of indoor objects [5], [6].

This paper proposed the IPS in an unideal environment with low-cost devices based on received signal strength indicator (RSSI) from Wi-Fi devices. One of the

*E-mail: dwi.jokosuroso@ugm.ac.id, alvin.samuel.h@mail.ugm.ac.id, arifin.effendi@mail.ugm.ac.id, singgih@ugm.ac.id*

technologies often used in indoor positioning systems is Wi-Fi because the market potential is getting more prominent, and the broader use of Wi-Fi [7]. Today, almost every mobile device can access Wi-Fi without the need for additional software or hardware [8]. Both of these things make the Wi-Fi infrastructure installed in a room can be used as IPS.

We also observed that people's trend or behaviour because of the COVID-19 pandemic is that people are afraid to stay indoors for some time with other people in condensed public facilities such as restaurants. In some places, especially nearby the authors live, we also found new open business that has the concept, outdoors and semi-outdoor restaurants or dining places. They tend to have the partial portion of indoor and outdoor in a mix of a vibrant environment. In the unideal environment where the room's geometry is not symmetrical, the outdoor portion also appears in some corner of the room, giving the new perspective in the new normal after pandemic COVID-19 will be [9].

Indoor positioning can be difficult since the general model and techniques and technologies are not yet standardized. It triggers the varying of method/techniques and performance metrics which are ranging. Our proposal employed several methods or pattern matching algorithms in the fingerprint techniques because, as mentioned in several previous publications, the fingerprint technique is suitable only applied in the static environment. We want to know how the dynamic environment can be simulated by utilizing an unideal environment yielding unideal indoor positioning properties and tackling it with those several methods?

This paper presents how we tackle the long and exhaustive effort to collect the fingerprint database by applying interpolation methods for data synthesis. In the pattern matching algorithm, we proposed the known Random Forest algorithm to estimate the target's location by comparing the target data parameter to those in the database. Random forest is an algorithm in machine learning that combines bagging and decision tree techniques [10]. We applied the Wi-Fi-based RSSI parameter because of its simple and straightforward implementation for IPS. The significant benefit of using Wi-Fi-based is installing the system to the building or a dedicated indoor environment with the Wi-Fi instalment and globally available.

The other Wi-Fi-based properties are the link quality indicator (LQI), channel-based, i.e., channel state information (CSI), angular-based, and many more. However, these parameters need additional hardware and a specific calibration method that is complex and sophisticated [11],[12]. In the signal point of view, one of the RSSI limitations is how its values are somehow fluctuations and sensitives to the time-varying effect. However, seeing from some publications, an acceptable accuracy threshold that RSSI can still offer mainly depends on the requirements and scenarios. The RSSI-based method seems old-fashioned compares to the latest long-term evolution (LTE) and the next generation of mobile communication, the fifth-generation, 5G, which predicted in these eras of communications, the use of RSSI would be faded and replaced by CSI-based. The widespan of wideband in 5G also gives the new point-of-view on the wideband-based indoor positioning channel. Some preliminary works on the radio channel localization utilizing the 5G's frequency channel sounder also can be found in [13]. Unfortunately, these LTE and 5G approaches are still costly and complicated in measurements and signal processing [14],[15].

To handle the RSSI values' time-varying effect, we applied the commonly used technique in IPS, namely the fingerprint technique. The fingerprint technique is divided into two phases; the offline phase and the online phase [5],[16], [17]. The offline phase is the phase of measuring the signal properties at each reference point in the room. This data is then processed into a database. A pattern-matching algorithm is used in the online phase to compare target or users' data with those in the database to estimate the user's position [17]. Examples of classical pattern matching algorithms are minimum Euclidean distance (MED) and maximum likelihood. The fingerprinting technique's advantages besides reducing RSSI time-varying effects, fingerprint technique can be built by low-cost equipment with yield also high precision [18]. The fingerprint technique's weakness is that it takes a lot of time and effort when collecting data and only best in the static environment [19].

Our approach employed the interpolation technique to answer the challenge or disadvantages of database collection in the fingerprint-based IPS. We considered several scenarios; the sole reason is to ensure that the database values' interpolation approach can work well. The scenarios include

- conducting two measurement campaigns,
- picking the train and validation data from the measurement combination, and
- considering some variables since the interpolation database is also compared to the complete database we collected in every point of our designed measurement campaign quality of the data synthesis results.

The variation of these scenarios may be too excessive at some points. However, we can observe the localization trends results by comparing these variations in our proposal scenarios from the training database variations and the validation data. This comparison of methods can be one point of view, especially in using the positioning technique more efficiently.

Our proposal gives more perspectives on the database point of view and how we can use many data, in which some of them have a degree of randomness and fluctuation. In which, we applied the Random Forest algorithm based on the decision tree to face these issues. We also observed that the unideal environment to validate our proposal and utilize several fingerprint technique methods could contribute, especially in seeing that the Wi-Fi-based RSSI values are the most straightforward parameter for positioning.

In this paper, we employed the ESP-32 devices based on Wi-Fi. We arrange the devices as the station (STA) and access points (APs). We followed the Star's Topology of Wireless Sensor Networks (WSNs) based on indoor localization or positioning. In our approach, we use 4 (four) APs as the reference nodes and 1 (one) STA as the target or user. The data collection starts by validating devices' conditions by measuring the relatively line-of-sight (LOS) condition. Then, we conduct the real measurement in an unideal indoor environment by observing the RSSI values recorded, which follow the general path loss model.

We compared synthesize database with the actual database we collected at each point of the measurement grid. We discussed our IPS system's accuracy by comparing the performance metric, i.e., accuracy, precision, and maximum error, of our proposed method using the random forest and the classical MED. We consider this our research's novelty from the method explanation by applying both interpolation and machine learning-based Random Forest. As there is limited resource discussed, the Neville and Bilinear interpolation for IPS application and the use of the unideal environment as the area of interest for the measurement campaign.

We organize this article as follows: The first section discusses an introduction to our research. We discuss the material and research methods in the second part. The measurement campaign and detail of the analysis method are detailed in the third section. The fourth section presents the results and discussion. Finally, we conclude our findings in the fifth section.

## 2. Materials and Research Methods

Our proposal's main objective is to tackle the disadvantage issues in the fingerprint technique on the exhaustive database collection. We applied the interpolation methods in order to synthesis the data between fingerprints point for the database. We also conducted the field measurement in the unideal environment to see how far our proposal can be acceptable if the indoor environment and positioning parameters are not favorable. Our previous works on the ideal indoor environments and achieve a favorable condition can be accessed in [20]-[23].

In this perspective, our method can be seen as a combination of several known methods. However, we measured that this approach is visible and prospective. Our proposed method's challenge is to make sure the randomness of the Random Forest algorithm's data enables the algorithm to work more precisely and accurately.

### A. Wireless Sensor Networks (WSNs)-based Indoor Localization/IPS

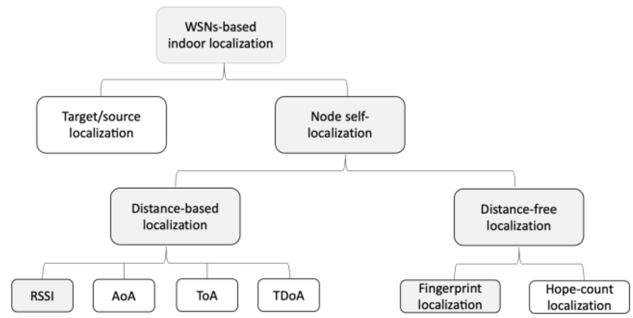



Figure 1. Taxonomy of WSN-based indoor localization[19]

The taxonomy of WSNs-based indoor localization is depicted in Fig. 1. WSN-based indoor localization/IPS has been a popular topic in this two decades. Many proposals have been published, and several new methods have been applied in real-life applications. The author's previous works published related to WSNs based indoor localization by using fingerprint technique can be found in [21],[22],[23]. The advantage of applying WSNs is to be more flexible in the node-self localization if we are discussing the real-life implementation.

WSNs star topology is most used for IPS applications. The topology consists of the reference nodes and one sink node. The sink node can be the target node in the perspectives of IPS. This topology allows the sink to re-send the package information broadcast request from reference nodes. Fig. 2 shows the illustration of this topology which we used in our devices.

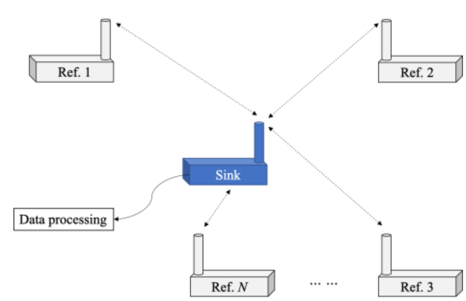



Figure 2. Star topology for nodes in WSN-based IPS [20]

### B. Fingerprint Technique

In WSNs-based IPS, the fingerprint technique is widely popular due to its advantages when the parameter for positioning is not stable, i.e., RSSI. Several methods in fingerprint technique comparisons can be found in [24].

As shown in Fig. 3, the fingerprint technique comprises two fundamental phases; offline and online phase. In the offline phase, the fingerprint database is collected. It can be grids in the area of interest with a specific density. The estimated position can be determined by comparing the database's parameter to the target's present parameter. This process is done in the online phase by a pattern matching algorithm.

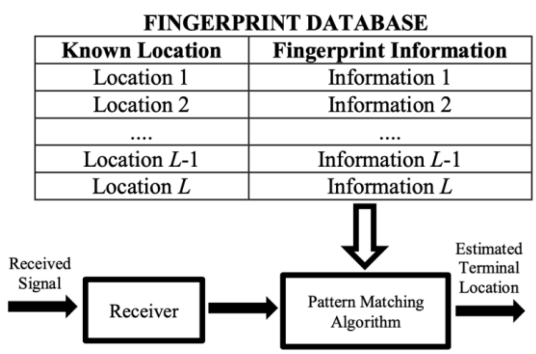



Figure 3. Illustration of fingerprint technique process [21]

*C. Received Signal Strength Indicator (RSSI)*

In a simple indoor localization/IPS, we often find the straightforward use of RSSI. RSSI is an important parameter as it represents the relationship between the received and the transmit power [25].

$$RSSI = \frac{P_r}{P_t} \tag{1}$$

RSSI expressed in dB is

$$RSSI\ (dBm) = P_r(dBm) - P_t(dBm) \tag{2}$$

In our proposed method, we stored the RSSI data from the WSNs topology mentioned earlier to build the unideal environment's fingerprint database. The target/user data is also collected in the same fashion and estimated based on similarity with those in the database by applying a pattern matching algorithm.

*D. Pattern Matching Algorithm: Minimum Euclidean Distance*

In the fingerprint technique, to locate the target, we need a pattern matching algorithm. Several classical pattern matchings have been proposed, including the famous minimum Euclidean distance (MED) [26]. This pattern-matching algorithm calculates the Euclidean distance between the fingerprint data in the database and the fingerprint data that the user/target receives. The data in the database that has the smallest Euclidean distance with the fingerprint data is the data that most closely matches the user's data so that the coordinates that are paired with the fingerprint data are the predictive results of the user's position.

$$E_d = \sqrt{\sum_{i=1}^{n} (RSSI_{ri} - RSSI_{ui})^2} \tag{3}$$

Eq. (3) calculates the Euclidean distance where, $RSSI_{ri}$ is the fingerprint data in the database from the $i$-th transmitter, $RSSI_{ui}$ is the fingerprint data of the user from the ith transmitter, and $n$ is the number of transmitters.

*E. Pattern Matching Algorithm: Random Forest*

1. Decision Tree

The decision tree is an algorithm in machine learning which has the shape like a tree structure. This tree structure is comprised of a root node, where here, containing all training data. This data can be divided into two or more nodes that act like branches in a tree that call the child nodes. The number of child nodes is following certain conditions [10],[27]. Splitting is done to meet the stopping criterion. Nodes that cannot be divided again (terminal nodes) will be labeled with a specific value based on the data contained in that node. In the classification and regression trees (CART) algorithm, each node is divided into two child nodes (binary split). A condition $q$ is chosen if this condition gives the most significant reduction in impurity, $\Delta im(q,t)$.

$$\Delta im(q,t) = im(t) - \frac{N_{t_L}}{N_t} im(t_L) - \frac{N_{t_R}}{N_t} im(t_R) \tag{4}$$

Where $N_t$ is the amount of data at node $t$, $N_{t_L}$ the amount of data on child nodes $t_L$ and $N_{t_R}$ is the amount of data on child node $t_R$ and $im(t)$ is the impurity at node t based on the criteria of Shannon entropy (5) or gini index (6).

$$im_H(t) = -\sum_{k=1}^{n} pr(c_k|t) \log_2\big(pr(c_k|t)\big) \tag{5}$$

$$im_G(t) = \sum_{k=1}^{n} pr(c_k|t) \big(1 - pr(c_k|t)\big) \tag{6}$$

Where $pr(c_k|t)$ is the probability that the value of $c_k$ appears at node $t$. Fig. 4 shows the illustration of a decision tree with two features.

| Target | Feature1 | Feature2 |
|---|---|---|
| A | 1 | 0 |
| A | 1 | 1 |
| A | 1 | 0 |
| A | 0 | 1 |
| A | 0 | 1 |
| A | 1 | 1 |
| B | 1 | 0 |
| B | 0 | 0 |

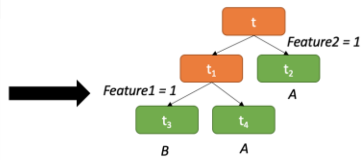



Figure 4. Example of a decision tree [10]

2. Random Forest

Random forest is a learning algorithm that combines several decision trees for classification, regression, or other tasks. The construction of a multitude of decision tree at training phase yield output the class as the mode of the classes or mean/average prediction of the individual trees [10]. Random forest uses bootstrap aggregating (bagging) techniques and randomly selects variables at each node when making each decision tree as seen in Fig. 5 [26] [27]. The bootstrap technique generates several training data by randomly removing some data from the original training data and replacing them with the remaining random data [28].

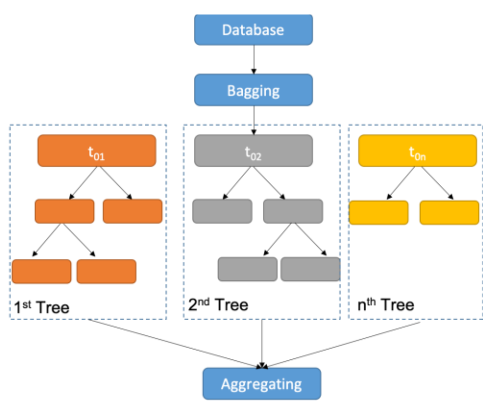



Figure 5. Illustration of random forest [10][27]

Each decision tree built will be combined (aggregating) by taking the most votes for the classifier or the average of each decision tree's predictions for the regressors [10], [26]. The random forest we used in this study was the random forest regressor in Scikit-learn [29].

*F. Interpolation*

1. Neville Interpolation

Suppose there are 4 points $(x_0, y_0)$, $(x_1, y_1)$, $(x_2, y_2)$ and $(x_3, y_3)$, a point $(x, y)$ can be interpolated by following the steps to solve Neville's interpolation using Table 2 (in Appendix) begins from the smallest b. Where b is the degree of the polynomial equation [30]. The general equation for the Neville method and its settlement are shown in the Appedix.

2. Bilinear Interpolation

Suppose there are 4 points, namely points $(x_1, y_1), (x_2, y_1), (x_1, y_2), (x_2, y_2)$ which have values of $f(x_1, y_1), f(x_2, y_1), f(x_1, y_2), f(x_2, y_2)$ which is depicted in Fig. 6. The first step is linear interpolation on the $x$ -axis when $y = y_1$ and $y = y_2$ so that the values of $f(x, y_1)$ and $f(x, y_2)$can be found using Eq. (7) and Eq. (8):

$$f(x, y_1) = \frac{x - x_1}{x_2 - x_1} f(x_2, y_1) + \frac{x_2 - x}{x_2 - x_1} f(x_1, y_1) \quad (7)$$

$$f(x, y_2) = \frac{x - x_1}{x_2 - x_1} f(x_2, y_2) + \frac{x_2 - x}{x_2 - x_1} f(x_1, y_2) \quad (8)$$

The next step is to perform linear interpolation on the $y$-axis using Equation (9).

$$f(x, y) = \frac{y - y_1}{y_2 - y_1} f(x, y_2) + \frac{y_2 - y}{y_2 - y_1} f(x, y_1) \quad (9)$$

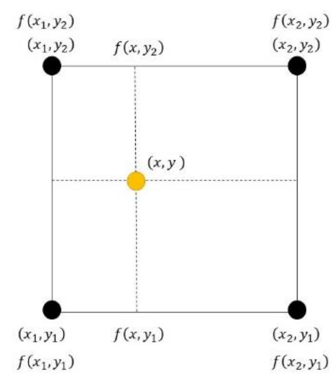



Figure 6. Bilinear interpolation scheme [31]

## 3. MEASUREMENT CAMPAIGN

### 3.1. General Measurement Setup

We conducted a measurement campaign in typical housing for students nearby our university. We collected the fingerprint database and tested our approach in this semi-indoor location, which is an unideal indoor environment with lots of metal objects the unsymmetrical geometry. Fig. 7 is the fingerprint data measurement location, and Fig. 8 and Fig. 9 are the measurement schemes applied in our proposed method. The difference between the two measurement schemes used is the difference in the position of the access point.

The Wi-Fi-based RSSI signal is obtained from the ESP-32 device, which is low cost and portable system. As seen in

Fig. 8 and 9, we employ five units ESP-32, which acts as the target collecting the RSSI information from four access points (APs). The RSSI values fluctuation effects are also first observed by merely placing the transmitter and receiver both using ESP-32 and measured the RSSI values varying by distances. The detailed measurement campaign equipment is listed in Table I.

Figure 7. Real measurement campaign environment

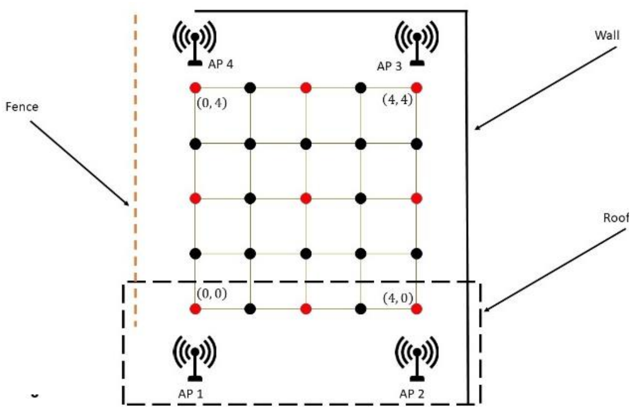



Figure 8. Measurement setup and first measurement campaign scheme

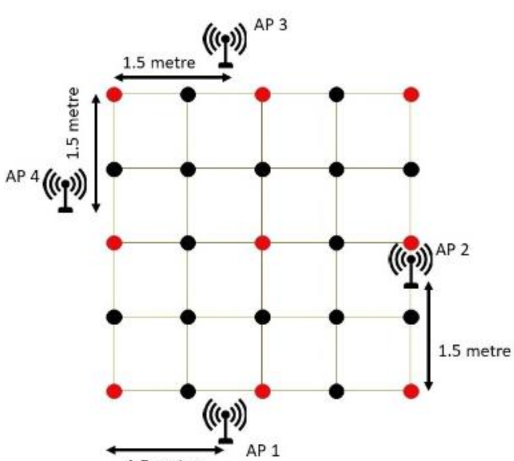



Figure 9. Second measurement campaign scheme

TABLE I. MEASUREMENT EQUIPMENTS AND DETAIL

| No. | Device | Specification | Detail |
|---|---|---|---|
| 1 | ESP32 (5 units) | *Microcontroller* which has Wi-Fi and BLE module, working frequency of 2.4 GHz. | 4 units act as the transmitter (TX) and 1 device for receiver (RX) |
| 2 | Laptop Acer Aspire ES 14 | *Processor* AMD A-8, RAM 4 GB | Recording the parameters of measurement campaign. |
| 3 | Python 3 | 3.7 ver. | To train and embed the pattern matching algorithm and database enhancement. |

### 3.2. The RSSI Values Observation

The RSSI values that we observed before the database collection can be seen in Fig. 10 with the illustration setup. The antenna symbol represents the station (STA) as the transmitter, and red dots in the varying distance will be a receiver (APs), and it moves in 1m distance in each RSSI data collection. Fig. 11 shows the real setup for RSSI vs. distance effects.

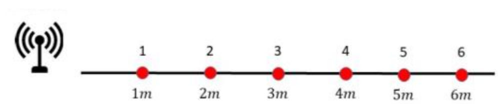
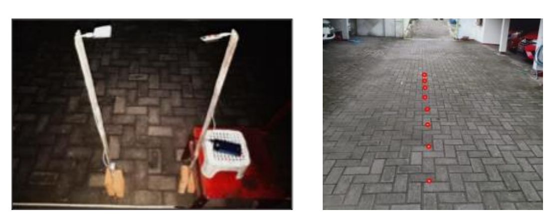



Figure 10. The RSSI values vs distance measurement illustration



(a) (b)

Figure 11. The measurement setup for RSSI values vs distance

Fig. 12 depicts the empirical path loss results vs. the general path loss (PL) model (the theoretical path loss). We used the measured RSSI values for the PL comparison as in [25]. From Fig. 12, the RSSI values from measurement somehow follow the PL model in general. However, we can see some fluctuations in the RSSI values vs. distance in all APs. We also measured various orientations for the effects of orientation of both STA and APs for the RSSI values.

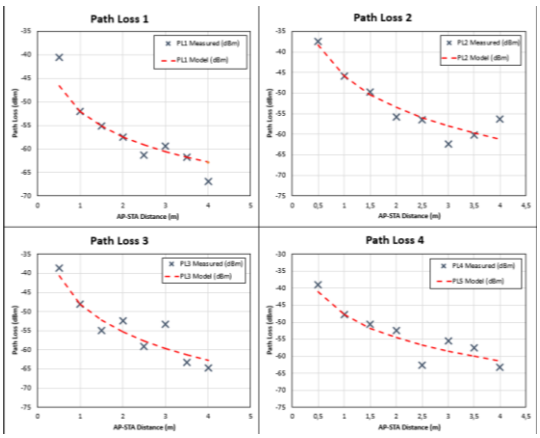



Figure 12. The measurement setup for RSSI values vs distance

The illustration and results for these orientation effects can be seen in Fig. 13 and 14, respectively. In this measurement, we placed the STA and AP at a 1 m distance. As the RSSI vs. distance measurement, we also measured all 4 APs to ensure there is no defect in our hardware.

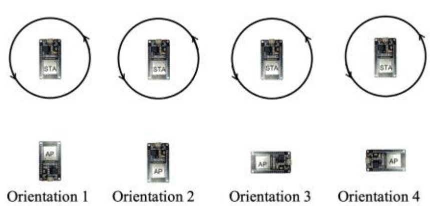



Figure 13. The measurement setup for RSSI values vs orientation

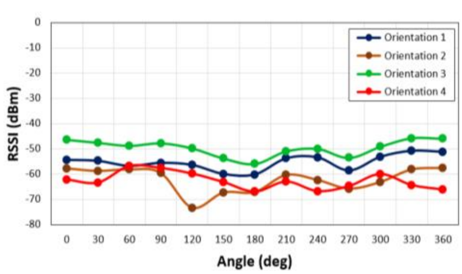



Figure 14. The RSSI values of STA vs orientation

RSSI values observation measurements yield that the RSSI values in a certain distance and orientation follow the theoretical and basic concept of received power behaviour. There is no defect to the equipment used in the measurement campaigns. The fluctuated RSSI values in some distance and orientation points are naturally due to the environment where we tested the device. We conducted the test in the parking lot in the same area of interest for the general measurement campaign.

## 3.3. Measurement Campaign Schemes

### 3.3.1. First Scheme

In the first measurement campaign scheme, at each reference point (red and black dots) shown in Fig. 8 and Fig. 9, the fingerprint data measurement is carried out four times with different transmitting device orientations. In each orientation, the measurements were taken for 2 minutes. Every 3 seconds, there is one RSSI value recorded from each access point; however, the fingerprint data used in the database is the average value of 15 seconds of measurement. The measurement step in the second scheme is the same as the first measurement step. However, each orientation's measurement at a particular reference point is only carried out for 1.5 minutes. In the first measurement scheme, there are seven fingerprint data from each orientation at each reference point. Five of these data are used as databases, and 2 data are used as validation data.

### 3.3.1. Second Scheme

There are five fingerprint data from each orientation at each reference point in the second measurement scheme, as depicted previously in Fig. 8. Three of these fingerprint data were used as a database, and 2 of these were used as validation data. The database is used to train the random forest model, while the validation data is used to determine the random forest model's performance.

### 3.3.2. Databases

There are two types of databases and data validation used. The first type uses the average RSSI value of various orientations at each reference point. The second type uses all the measured fingerprint data and has one additional variable, namely orientation. For example, each reference point has 20 fingerprint data in the database.

For each type of database, there are four databases used. The first database only used fingerprint data from the red-colored reference point shown in Fig. 8, then named Database_2m for the first type of database and DatabaseComplete_2m for the second type of database. The second database uses the red and black-point fingerprint data and is named Database_1m for the first type of database and Database Complete_1m for the second type of database. The third database is the first database engineered with the Neville method's interpolation technique so that all black points become the interpolated fingerprint data. The interpolation scheme with the Neville method can be seen in Fig. 15. This database will be named Database_IntNe for the first type of database and DatabaseComplete_IntNe for the second type of database. The fourth database is the first database engineered by using Bilinear interpolation. The Bilinear interpolation scheme is shown in Fig. 16.

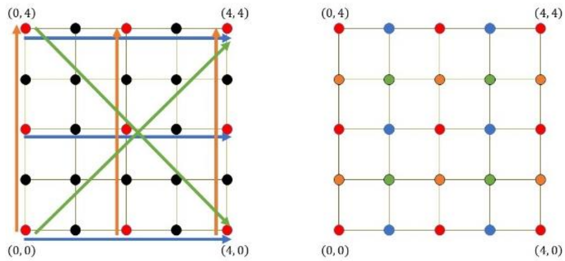



Figure 15. Interpolation scheme in Neville interpolation

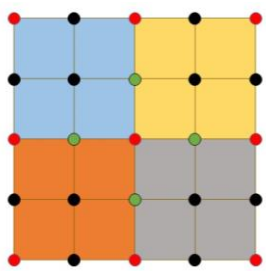


Figure 16. Interpolation scheme in Bilinear interpolation

In Fig. 16, the black dots are only interpolated inside one square while the green dots are interpolated inside two squares so that the result is the average of the interpolated results of the two squares. This database will be named Database_IntBi for the first type of database and DatabaseComplete_IntBi for the second type of database.

There are two types of validation data used, the first type of validation data is called DataValidation, and the second type is called DataValidationComplete. The sharing of these two types of validation data is similarly done as in the database.

### 3.3.2. Positioning Scenarios

We apply three scenarios in our proposed method; Scenario 1 uses the first type of database and the first type of validation data. Scenario 2 uses the first type of database and the second type of validation data. Finally, Scenario 3 uses the second type of database and the second type of validation data. The performance of the pattern matching algorithm that will be studied is the maximum accuracy, precision, and error. Accuracy is the mean error of the predicted results and the actual position [24] [32].

$$Accuracy = \frac{1}{j}\sum_{i=1}^{j} Error_i \quad (9)$$

Precision is the amount of data that has an error less than or equal to the desired error divided by the total number of data [16].

$$Precision = \frac{1}{j}\sum_{i=1}^{j} 1(Error_i \leq Accuracy_d) \quad (10)$$

The maximum error is the difference between the predicted results and the actual position with the greatest value from each scenario

## 4. Results and Discussion

We presented the results as the measurement scheme and positioning scenarios. The three performance metrics, accuracy, precision, and maximum error, are analyzed and compared between the minimum Euclidean distance (MED) as the classical pattern matching algorithm and the proposed method using the Random Forest.

### *A. Scenario 1, the first measurement scheme*

As depicted in Fig. 17, by using the interpolation technique on Database_2m, it makes the MED algorithm's performance increase in maximum accuracy by 32%, increasing maximum precision by 12%, and maximum error does not change. For the random forest algorithm, the maximum accuracy increased by 24.9%, the maximum precision increased by 20%, and the maximum error decreased by a maximum of 18.2%. The Random Forest gives slightly better performance in terms of increase the system's precision and decrease the maximum error.

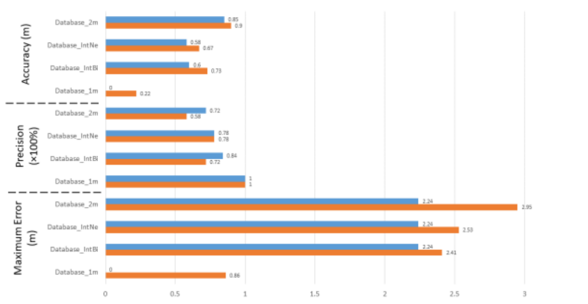



Figure 17. The pattern matching algorithm performance in Scenario 1

We compared the interpolated database and Database_1m (the actual database on the grid of 1m), which yields the differences in accuracy is up to 0.6 meters, the difference in precision can be up to 28%, maximum error difference can be up to 2.24 meters. As the accuracy and precision are still acceptable by observing the discrepancy, the handling of maximum error is poor. We observed that some RSSI outliers are included in the interpolation process.

If the two pattern matching algorithms are compared to their performance when using the same database, the most significant difference in accuracy reaches 0.22 meters, the biggest difference in precision reaches 14%, and the maximum error difference reaches 0.86 meters, where the performance of the Euclidean distance algorithm is better than the random forest algorithm in this case. However, the overall performance metric increasing of the Random

Forest is slightly better than MED in precision and terms of maximum error reduction.

*B. Scenario 2, the first measurement scheme*

In this scenario, we utilized the interpolation technique on Database_2m, resulting in increased performance on the MED of its maximum accuracy by 12%, precision by a maximum of 8%, and the maximum error has decreased 12.5%. For the random forest algorithm, accuracy increased by a maximum of 8.7%, precision increased by a maximum of 8%, and maximum error decreased by 9.8%.

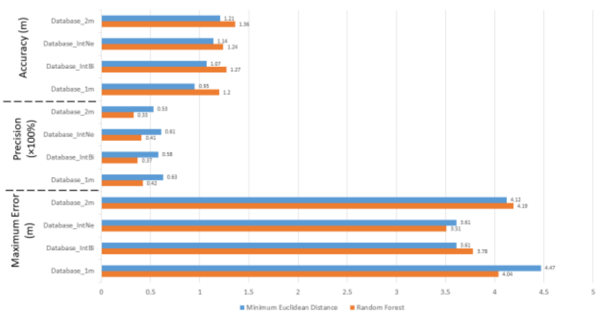



Figure 18. The pattern matching algorithm performance in Scenario 2

The difference in accuracy between the pattern matching algorithm that uses the interpolated database and Database_1m can reach 0.19 meters, the difference in precision can reach 5%, where the performance of the pattern matching algorithm and the maximum error difference can reach 0.86 meters.

If the two pattern matching algorithms' performance is compared when using the same database, the most significant difference in accuracy is 0.25 meters, and the difference in precision reaches 21%. The results show that the minimum Euclidean distance algorithm still performs better. On the other hand, the maximum error difference can reach 0.43 meters, where the random forest algorithm has better performance to reduce the estimated error location.

*C. Scenario 3, the first measurement scheme*

Using the interpolation technique on Database_2m makes the MED performance increasing accuracy by 7.1%, precision decreases by 1%, and maximum error decreases by 12.5%. For the random forest algorithm, the accuracy increased by 3.8%, and the precision increased by 1%, and the maximum error increased by 3.12%.

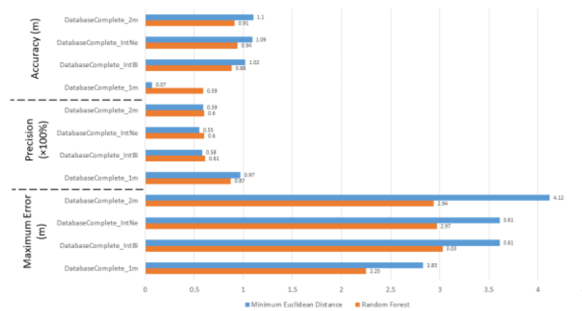



Figure 19. The pattern matching algorithm performance in Scenario 3

The difference in accuracy between the interpolated database and Database_1m reaches 1 meter, the difference in precision reaches 42%, and the maximum error difference is 0.78 meters.

Suppose the performance of the two pattern matching algorithms is compared when using the same database. In that case, the accuracy can reach 0.5 meters, and the difference in precision can reach 10%, where the MED performance has better performance than the Random Forest. In comparison, the maximum error difference between the two algorithms can reach 1.18 meters, where the Random Forest algorithm performs better than MED. In this scenario, in terms of estimated error reduction, the Random Forest performed ahead of MED.

*D. Scenario 3, the second measurement scheme*

Using the interpolation technique on the Complete Database_2m when using the minimum Euclidean distance algorithm increases the accuracy by 13.2%, increases the precision by 3%, and does not affect the maximum error. When using the random forest algorithm, accuracy increased by 8.7%; precision increased by 3%, and the maximum error decreased by 4%.

The difference in accuracy between the Complete Database_1m and the interpolated database is 0.94 meters, the difference in precision reaches 34%, and the maximum error difference reaches 4 meters as shown in Fig. 20.

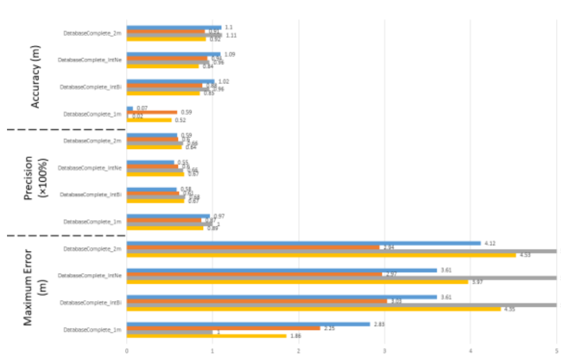



Figure 20. Comparison of the performance of the pattern matching algorithm from the two measurement schemes using scenario 3

When the two algorithms' performance is compared when using the same database, the accuracy reaches 0.5 meters, the difference in precision reaches 11%, where the MED performance is better than random forest. The maximum error difference reaches 1.03 meters, where the random forest algorithm's performance is better than the MED.

The change in the performance of the pattern matching algorithm when using the interpolation technique in DatabaseComplete_2m for the two measurement schemes is as follows, in the first scheme, the increase in accuracy reaches 7.1%, while in the second scheme, the increase in accuracy reaches 13.2%. The precision increases to 1% in the first scheme, while in the second scheme, it reaches 3%. There is an increase of 3.12% in the first scheme for the maximum error, while in the second scheme.

The similarity between the two is that the increase in performance cannot match the pattern matching algorithm's performance when using DatabaseComplete_1m or a database of measurement results at each reference point. The reason is that the results of interpolation at the reference points denoted by black dots are different from the actual measurement results at that point, as shown in Fig. 21 and 22. This similarity could result from the inconsistency of RSSI values received by the STAs as the effects the room geometry, interference objects within the measurement campaign area. Thus, the interpolation techniques give a poor accuracy in recreating the RSSI values in the several grid positions. These reasons yield a high error prediction.

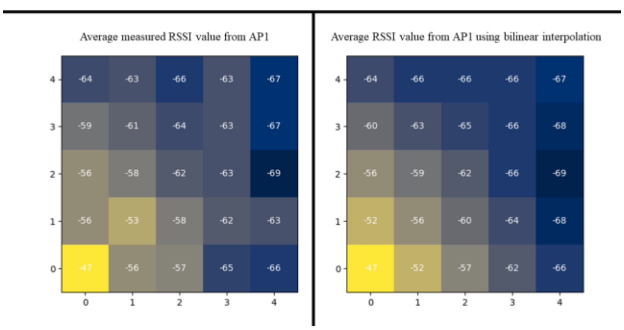



Figure 21. Comparison of the mean RSSI value of measurement results and Neville interpolation

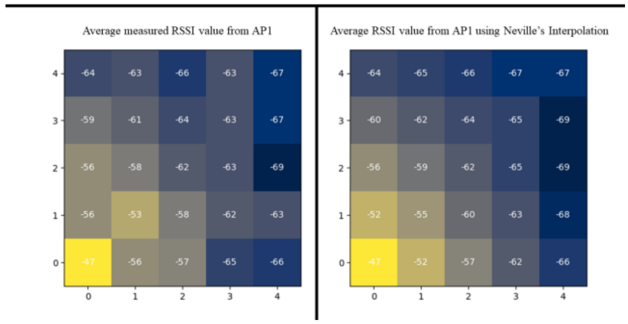



Figure 22. Comparison of the mean RSSI value of measurement results and Bilinear interpolation

## 5. CONCLUSIONS

This paper discusses how we propose the method to reduce the effort in constructing a database by interpolation. We also give a new perspective on the indoor positioning system in an unideal environment. We can see the environment is not favourable and wanted to see how far our proposal is still valid. We also applied the machine learning-based pattern matching algorithm, the Random Forest, to observe whether it can reduce the estimated error.

To validate our proposal, we conduct a measurement campaign, testing the devices to ensure the RSSI data's validity. We divided our findings into several scenarios to know the best scenario for applying both the interpolation technique and the correct pattern matching algorithm. From several scenarios performed, the difference in accuracy between the two algorithms reaches 0.5 meters, and the difference in precision reaches 20%, where the minimum performance of the Euclidean distance is better in most cases. The use of interpolation techniques can improve the pattern matching algorithm's performance by 3% to 30%. The difference in the performance of the pattern matching algorithm that uses an interpolated database of results with Database_1m for accuracy can reach 1 meter, for precision, it can reach 42%, and for the maximum error, it can reach 2.24 meters.

When using various measurement schemes, the pattern matching algorithm's performance improvement using the interpolated database results cannot approach the pattern matching algorithm's performance when using the database from measurement at each reference point. The reason is that the RSSI value of the measurement results and the interpolation results are somehow different as we found that the RSSI values are not linear by their distance due to propagating environment. The performance of MED is still superior in terms of accuracy and precision, at least in our proposal. However, in terms of reducing the maximum error, the random forest performed better in all scenarios.

As the answer to the open question regarding the RSSI values used as the parameter for positioning, we have planned to employ a ray-tracing method for future works to validate the RSSI values from measurements. We are also investigating the optimization algorithm to remove the RSSI data outlier from the measurement campaign to increase our indoor positioning system's performance.


## ACKNOWLEDGMENTS

This research is supported by "*Hibah Penelitian dan Anggaran Tahunan Fakultas Teknik, Universitas Gadjah Mada (UGM) Tahun 2020.*"


## APPENDIX

The Neville interpolation general equation:

$$Nv_b[x_i, x_{i+1}, \dots, x_{i+b}] = \frac{(x-x_{i+b})Nv_{b-1}[x_i,x_{i+1},\dots x_{i+b-1}]+(x_i-x)Nv_{b-1}[x_{i+1},x_{i+2},\dots,x_{i+b}]}{x_i-x_{i+b}}$$

Table II shows the detail Neville settlement method.

TABLE II. THE NEVILLE METHOD SETTLEMENT

| | $b = 0$ | $b = 1$ | $b = 2$ | $b = 3$ |
|---|---|---|---|---|
| $x_0$ | $Nv_0[x_0$ $= y_0$ | $Nv_1[x_0, x_1$ | $Nv_2[x_0, x_1, x$ | $Nv_3[x_0, x_1, x_2, x$ |
| $x_1$ | $Nv_0[x_1$ $= y_1$ | $Nv_1[x_1, x_2$ | $Nv_2[x_1, x_2, x$ | - |
| $x_2$ | $Nv_0[x_2$ $= y_2$ | $Nv_1[x_2, x_3$ | - | - |
| $x_3$ | $Nv_0[x_3$ $= y_3$ | - | - | - |

The parts of the unideal environment for the measurement campaign can be seen in Fig. 23-25 and show many IOs and the non-geometry structure of the buildings. From Fig. 23, we can see that the only part of APs in an indoor environment are AP 1 and AP2. While the other APs are located outdoor (without the roof). Graphical results on how the MED and random forest yield the prediction position can be depicted in Fig. 25 and 26.

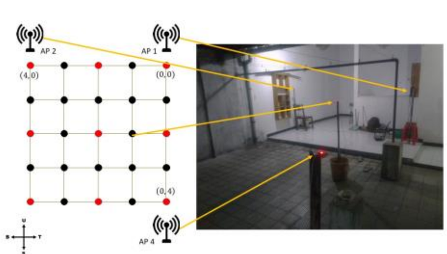



Figure 23. Measurement perspective toward North

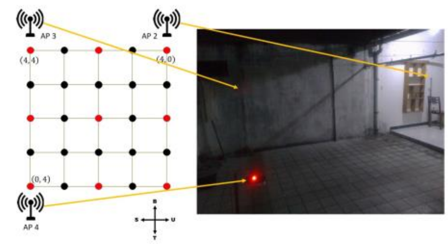



Figure 24. Measurement perspective toward West

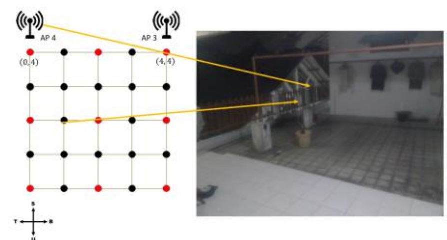



Figure 25. Measurement perspective toward South

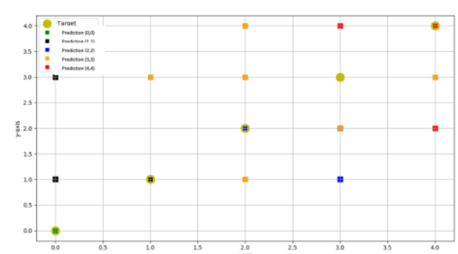


Figure 26. MED prediction visualization

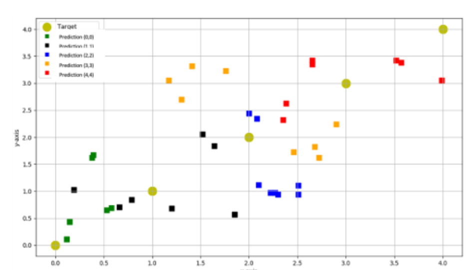


Figure 27. Random forest prediction visualization

## REFERENCES


[1] Yogesh K. Dwivedi, et.al, "Setting the future of digital and social media marketing research: Perspectives and research propositions", International Journal of Information Management, 2020, 102168, ISSN 0268-4012, doi.org/10.1016/j.ijinfomgt.2020.102168.

[2] T. Falkowski, "Feature model for the specification of industrial indoor location- based services," *Procedia Manuf.*, vol. 24, pp. 141–146, 2018, doi: 10.1016/j.promfg.2018.06.048.

[3] I. A. Getting, "Perspective/navigation-The Global Positioning System," in *IEEE Spectrum*, vol. 30, no. 12, pp. 36-38, Dec. 1993, doi: 10.1109/6.272176.

[4] A. Basiri *et al.*, “Indoor location based services challenges , requirements and usability of current solutions,” *Comput. Sci. Rev.*, vol. 24, pp. 1–12, 2017, doi: 10.1016/j.cosrev.2017.03.002.

[5] H. Liu, H. Darabi, P. Banerjee and J. Liu, "Survey of Wireless Indoor Positioning Techniques and Systems," in *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 37, no. 6, pp. 1067-1080, Nov. 2007, doi: 10.1109/TSMCC.2007.905750.

[6] A. Yassin *et al.*, "Recent Advances in Indoor Localization: A Survey on Theoretical Approaches and Applications," in *IEEE Communications Surveys & Tutorials*, vol. 19, no. 2, pp. 1327-1346, Secondquarter 2017, doi: 10.1109/COMST.2016.2632427.

[7] C. Yang and H. Shao, "WiFi-based indoor positioning," in *IEEE Communications Magazine*, vol. 53, no. 3, pp. 150-157, March 2015, doi: 10.1109/MCOM.2015.7060497.

[8] C. Wu, Z. Yang and Y. Liu, "Smartphones Based Crowdsourcing for Indoor Localization," in *IEEE Transactions on Mobile Computing*, vol. 14, no. 2, pp. 444-457, 1 Feb. 2015, doi: 10.1109/TMC.2014.2320254.

[9] I. B. Gurbuz and G. Ozkan, "Transform or Perish: Preparing the Business for a Postpandemic Future," in *IEEE Engineering Management Review*, vol. 48, no. 3, pp. 139-145, 1 thirdquarter,Sept. 2020, doi: 10.1109/EMR.2020.3014693.

[10]G. Louppe, “Understanding Random Forests: From Theory to Practice,” no. July, 2014, [Online]. Available: http://arxiv.org/abs/1407.7502.

[11]X. Wang, L. Gao, S. Mao and S. Pandey, "CSI-Based Fingerprinting for Indoor Localization: A Deep Learning Approach," in *IEEE Transactions on Vehicular Technology*, vol. 66, no. 1, pp. 763-776, Jan. 2017, doi: 10.1109/TVT.2016.2545523.

[12]K. Wu, J. Xiao, Y. Yi, D. Chen, X. Luo and L. M. Ni, "CSI-Based Indoor Localization," in *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 7, pp. 1300-1309, July 2013, doi: 10.1109/TPDS.2012.214.

[13K. Witrisal *et al.*, "High-Accuracy Localization for Assisted Living: 5G systems will turn multipath channels from foe to friend," in *IEEE Signal Processing Magazine*, vol. 33, no. 2, pp. 59-70, March 2016, doi: 10.1109/MSP.2015.2504328.

[14]N. Garcia, H. Wymeersch, E. G. Larsson, A. M. Haimovich and M. Coulon, "Direct Localization for Massive MIMO," in *IEEE Transactions on Signal Processing*, vol. 65, no. 10, pp. 2475-2487, 15 May15, 2017, doi: 10.1109/TSP.2017.2666779.

[15]C. Laoudias, A. Moreira, S. Kim, S. Lee, L. Wirola and C. Fischione, "A Survey of Enabling Technologies for Network Localization, Tracking, and Navigation," in *IEEE Communications Surveys & Tutorials*, vol. 20, no. 4, pp. 3607-3644, Fourthquarter 2018, doi: 10.1109/COMST.2018.2855063.

[16]S. He and S. -. G. Chan, "Wi-Fi Fingerprint-Based Indoor Positioning: Recent Advances and Comparisons," in IEEE Communications Surveys & Tutorials, vol. 18, no. 1, pp. 466-490, Firstquarter 2016, doi: 10.1109/COMST.2015.2464084.

[17]M. A. Al-Ammar *et al.*, “Comparative survey of indoor positioning technologies, techniques, and algorithms,” in *Proceedings - 2014 International Conference on Cyberworlds, CW 2014*, 2014, pp. 245–252, doi: 10.1109/CW.2014.41.

[18]S. Xia, Y. Liu, G. Yuan, M. Zhu, and Z. Wang, “Indoor fingerprint positioning based on Wi-Fi: An overview,” *ISPRS Int. J. Geo-Information*, vol. 6, no. 5, 2017, doi: 10.3390/ijgi6050135.

[19]Z. Farid, R. Nordin, and M. Ismail, “Recent advances in wireless indoor localization techniques and system,” *J. Comput. Networks Commun.*, vol. 2013, 2013, doi: 10.1155/2013/185138.

[20]Suroso, Dwi Joko, Muhammad Arifin, and Panarat Cherntanomwong. "Distance-based Indoor Localization using Empirical Path Loss Model and RSSI in Wireless Sensor Networks." *Journal of Robotics and Control (JRC)* 1.6 (2020): 199-207.

[21] D. J. Suroso, P. Cherntanomwong, P. Sooraksa, and J. I. Takada, “Location fingerprint technique using Fuzzy C-Means clustering algorithm for indoor localization,” in *IEEE Region 10 Annual International Conference, Proceedings/TENCON*, 2011, pp. 88–92, doi: 10.1109/TENCON.2011.6129069.

[22] P. Cherntanomwong and D. J. Suroso, "Indoor localization system using wireless sensor networks for stationary and moving target," 2011 8th International Conference on Information, Communications & Signal Processing, Singapore, 2011, pp. 1-5, doi: 10.1109/ICICS.2011.6173554.

[23] D. J. Suroso, P. Cherntanomwong, P. Sooraksa and J. Takada, "Fingerprint-based technique for indoor localization in wireless sensor networks using Fuzzy C-Means clustering algorithm," 2011 International Symposium on Intelligent Signal Processing and Communications Systems (ISPACS), Chiang Mai, Thailand, 2011, pp. 1-5, doi: 10.1109/ISPACS.2011.6146167.

[24]S. He and S. H. G. Chan, “Wi-Fi fingerprint-based indoor positioning: Recent advances and comparisons,” *IEEE Communications Surveys and Tutorials*, vol. 18, no. 1. pp. 466–490, 2016, doi: 10.1109/COMST.2015.2464084.

[25] K. Kaemarungsi and P. Krishnamurthy, "Modeling of indoor positioning systems based on location fingerprinting," *IEEE INFOCOM 2004*, Hong Kong, China, 2004, pp. 1012-1022 vol.2, doi: 10.1109/INFCOM.2004.1356988.

[26]L. Breiman, “Bagging predictors: Technical Report No. 421,” *Dep. Stat. Univ. Calif.*, no. 2, p. 19, 1994, [Online]. Available: https://www.stat.berkeley.edu/%7B~%7Dbreiman/bagging.pdf.

[27] L. Breiman, “ST4_Method_Random_Forest,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1017/CBO9781107415324.004

[28]B. Efron, “Bootstrap Method: Another Look at the Jackknife,” *Ann. Stat.*, vol. 7, no. 1, pp. 1–26, 1979

[29]. Varoquaux, L. Buitinck, G. Louppe, O. Grisel, F. Pedregosa, and A. Mueller, “Scikit-learn,” *GetMobile Mob. Comput. Commun.*, vol. 19, no. 1, pp. 29–33, 2015, doi: 10.1145/2786984.2786995

[30] D. B. Ninh, J. He, V. T. Trung, and D. P. Huy, “An effective random statistical method for Indoor Positioning System,” *Futur. Gener. Comput. Syst.*, vol. 109, pp. 238–248, 2020, doi: 10.1016/j.future.2020.03.043.

[30] J. Kiusalaas, *Numerical Methods in Engineering with Python 3*. New York: Cambridge University Press, 2013.

[31] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. New York: Cambridge University Press, 1992.

[32] J. Hightower and G. Borriello, “Location systems for ubiquitous computing,” *Computer (Long. Beach. Calif).*, vol. 34, no. 8, pp. 57–66, 2001, doi: 10.1109/2.940014.

**Dwi Joko Suroso** received his B. Eng. in Engineering Physics, Universitas Gadjah Mada (UGM), Indonesia and M. Eng. from King Mongkut’s Institute of Technology Ladkrabang, Thailand on 2010 and 2012, respectively. He is now full lecturer at Dept. of Nuclear Engineering and Engineering Physics, UGM. His research interest is in indoor localization, wireless sensor networks, and radio propagation model.

**Alvin S. H. Rudianto** received Bachelor of Engineering on 2020 majoring Engineering Physics (EP) from Universitas Gadjah Mada (UGM), Indonesia. He is now a research assistant in wireless research laboratory, EP UGM. His research interest is fingerprint-based indoor positioning system (IPS) database enhancement. His side work also on control system and instrumentation.

**Muhammad Arifin** received B. Eng. from Engineering Physics, Universitas Gadjah Mada (UGM), on 2020. He is now an UWB-based Local Positioning System researcher in PT. Delameta Bilano, Jakarta, Indonesia. His research interest including indoor localization, UWB-based indoor and outdoor positioning system with a sub-centimeter accuracy, and multiple devices positioning algorithms.

**Singgih Hawibowo** is a faculty member at the Department of Nuclear Engineering and Engineering Physics at Universitas Gadjah Mada (UGM), Indonesia. He got his bachelor's degree in 1987 in nuclear engineering from Universitas Gadjah Mada and his Ph.D. degree in 1997 in energy engineering from Technical University of Berlin. His research interests include sensors and measurement systems.