

CSUIAPDet: Indoor Wi-Fi Dataset for Access Point Location Determination based on Low-Cost Robot*

Di Tian^{1,2}, Dangjun Zhao^{1,2,*}

¹ School of Automation, Central South University, South Lushan Road 932, 410083 Changsha, China

² Hunan Provincial Key Laboratory of Optic-Electronic Intelligent Measurement and Control, South Lushan Road 932, 410083 Changsha, China

Abstract

Locations of indoor Access Points (APs) play a critical role in indoor Wi-Fi localization. The existing manual site survey for AP location determination is relatively accurate but labor-intensive and time-consuming, especially in large-scale scenarios. Hence, the existing state-of-the-art methods prefer to perform inversion methods based on received signal strength (RSS) at Reference Points (RPs) to estimate the locations of APs. However, due to the absence of common datasets, comparing these methods on an equal footing is challenging. Therefore, a dataset that covers multiple scenarios from a single room to the whole wide-range floor is first presented in this paper. The dataset is automatically constructed by a low-cost small indoor mobile robot at Central South University, Changsha, China. The dataset contains diversity elements: the occupancy grid map, the real global locations of the pre-deployed APs in each scenario, and the measurements at each reference point. As a result, it is convenient for scholars to propose new methods and execute comparisons. In addition, dataset baselines for AP location determination under different methods are demonstrated. The dataset will be available at <https://github.com/tcscu/CSUIndoorAPDet>.

Keywords

AP Location determination, Wi-Fi indoor localization, Dataset

1. Introduction

The localization technique takes prime importance in lots of application areas, such as transportation[1], measurement industry[2], factory production[3], and commercial services promotion[4]. The developed GPS (Global Position System)[5] provides sufficient accuracy on outdoor localization tasks. However, it is usually degraded by the obstruction from buildings, leaving accurate indoor localization a problem[6]. Hence, lots of indoor localization methods have been proposed. These methods are generally sensor-based and WLAN-based [7]. The former utilizes data from extra sensors such as cameras or LiDAR (Light Detection and Ranging) to compute location results, which is inconvenient for daily usage. The latter, also named wireless localization, takes the convenience of the widely pre-deployed WLAN to construct an indoor localization system, which greatly reduces the investment in external devices, especially in large-scale indoor scenarios.

There are two typical frameworks in indoor wireless localization: the range-based[8] and the fingerprint-based[9]. The first framework mainly localizes the users through a trilateral localization algorithm. This algorithm heavily relies on the locations of the APs and the estimated distances from the APs to the signal-receiving points. The second framework estimates users' locations based on the received signal strengths coupled with the corresponding locations (fingerprints) and the database. In this case, the prior-known locations of APs help to locate the users in a sub-range in the dataset to expedite the solution process[10]. Therefore, the location determination of APs is critical to the effectiveness of the indoor wireless localization system.

Proceedings of the Work-in-Progress Papers at the 14th International Conference on Indoor Positioning and Indoor Navigation (IPIN-WiP 2024)

* Corresponding author.

✉ tiandi_csu@csu.edu.cn (D. Tian); zhao_dj@csu.edu.cn (D. Zhao)

ORCID [0000-0001-6193-6994](https://orcid.org/0000-0001-6193-6994) (D. Tian); [0000-0001-9286-1999](https://orcid.org/0000-0001-9286-1999) (D. Zhao)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

The existing AP location determination methods until now are the model-free methods[11] and the model-based methods [12]. The series of model-free methods are derived from the centroid methods. The model-based methods are the optimization problems constructed based on the Log-distance Path Loss (LDPL) model[13]. These methods are all data-driven and sensitive to the measurements of RSS and the corresponding location values. Unfortunately, the data tested by these methods are not open-access, which is not conducive for researchers to make objective comparisons of existing methods, which hinders the development of the community. Besides, the manual data collection methods in these works above are time-consuming and labor-intensive. Therefore, inspired by the current situation and to promote the development of the research community, we propose the CSUIndoorAPDet: the first open-source wireless dataset for AP location determination.

The main contribution of this paper is the first comprehensive dataset for indoor AP location determination. The characteristics of this dataset and our work are as follows:

- First, our dataset is almost automatically constructed by robots based on SLAM and navigation rather than through manual site surveys. Hence, the process is labor-saving and convenient.
- Then, the dataset total covers 1060m² space and multiple scenarios including a single room and the entire floor. Therefore, the dataset can be utilized to test the applicability of the algorithm.
- Next, the usage example of our dataset on AP location determination is provided. The attached baseline shows clear comparisons among the existing methods, which helps the scholars' improvement in this field.
- Moreover, the information elements of our dataset are categorized by their attributes. These categories can be used for diversity problems, including not only the AP location determination, but the indoor layout generation, AP selection, or WLAN RSS-based localization.

The rest of this paper is organized as follows: Section 2 reviews the related work. Section 3 describes the collection methodology of our dataset. Section 4 details the components of the dataset. Section 5 presents a usage example of our dataset for AP location determination therefore the baseline of existing methods. Section 6 discusses our work, and Section 7 concludes this paper.

2. Related Work

The AP location determination methods are mainly fed with the RSS measurements and corresponding location values to estimate the locations of interested APs. The existing methods can be divided into the model-free and the model-based.

The model-free methods mainly estimate the AP locations by the centroid method. The pure centroid AP location determination method, proposed by Bulusu et al.[14], only use the location values of detected measurements to predict the location of APs. It is inaccurate because the unevenly distributed RSS levels are not included. According to this, the RSS levels were merged into the centroid method by Blumenthal et al. [15] to develop the weight centroid method, which is better than the pure centroid method. Nevertheless, these methods assume that the measured values are well distributed around the AP location, which is not reliable without prior knowledge. Moreover, these methods don't take the transmission characteristic of the indoor wireless signal into consideration.

The model-based methods utilize the log-distance path loss(LDPL) model to estimate the locations of APs. The LDPL model describes the propagation tendency in 3D space. It is an experienced model that describes the relation between the distances from the locations of APs to the corresponding RPs and the RSS levels at these RPs. Therefore, the AP location determination problem can be transformed into the nonlinear least-square optimization by using LDPL as the observation model,

which was proposed by J. Koo et al [12], but only tested in simulation. Y. Zhuang et al. [16] extended this method and provided a real field experiment.

However, the existing methods related to AP location determination are completely demonstrated on private datasets. Their dataset was constructed from a manual site survey, which is labor-intensive and time-consuming. Consequently, it is inconvenient for scholars to make comprehensive comparisons between their methods and the existing methods, hindering the development of indoor wireless localization technology. Hence, to promote further development in this field, we provide the CSUIndoorAPDet dataset.

3. Collection Methodology

This section demonstrates the methodology of data collection using an automatic wheeled robot. The components of the robot are detailed in 3.1. The data acquisition workflow is then introduced in 3.2.

3.1. System and Equipment

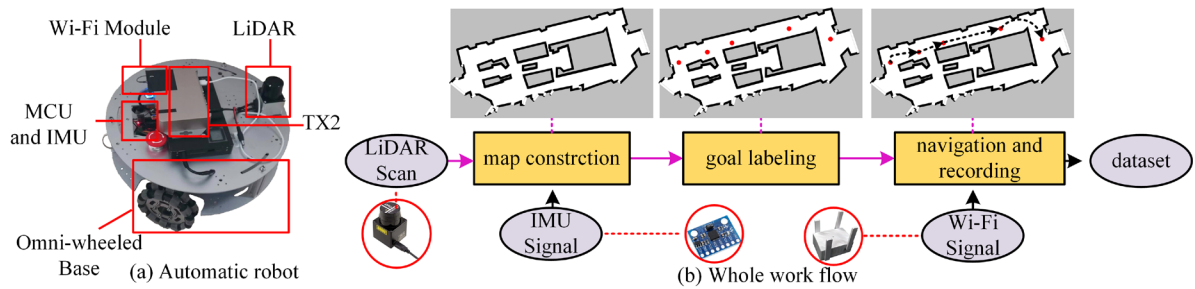


Figure 1: (a) The setup of our automatic robot and (b) the workflow of data collection.

The automatic robot is shown in Fig. 1 (a). Omni-wheeled bases are chosen for their flexibility and suitability in the indoor environment. The STM32F103RCT6 serves as the motor control unit (MCU). And the inertial measurement unit equipped here is MPU6050. It is coupled with the SIMINICS LiDAR to achieve mapping, localization, and navigation based on the NVIDIA Jetson TX2 development board. And the robot operation system (ROS) we use is ROS Melodic on Ubuntu 18.04. In addition, a Wi-Fi module is inserted on the TX2 to sniff the Wi-Fi signal.

3.2. workflow of data collection

The data collection workflow is illustrated in Fig. 1 (b). There are three sequential modules: indoor spatial map construction, goal labeling, and navigation with recording. The indoor occupancy grid map is constructed in the first module. The navigation goals are labeled on the indoor map in the second module. Based on the map and navigation goals, the robot records RSS and location values automatically and sequentially in the third module. The dataset is finally compiled by these measurements.

3.2.1. Indoor Spatial Map Construction

We chose the Cartographer SLAM[17] for indoor map construction based on 2D LiDAR. Constructing maps for indoor space manually is accurate, but it can be time-consuming and labor-intensive, especially in large-scale indoor scenarios. The SLAM technology can also generate accurate indoor layout representation, and it performs nearly real-time even when faced with large indoor areas. In addition, two-dimensional location information is sufficient to meet the most development needs of indoor positioning systems. Therefore, 2D LiDAR SLAM is suitable. The maps constructed are occupancy grid style and stored as PGM format files in our dataset.

3.2.2. Goal Labeling

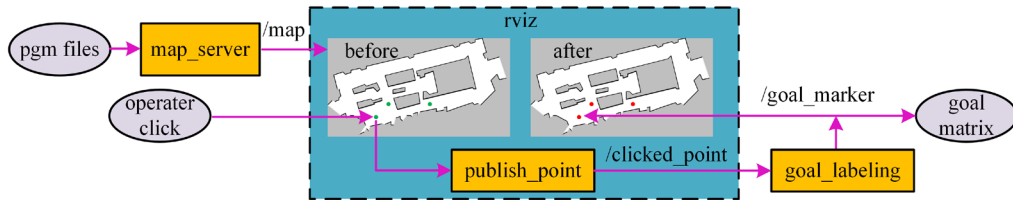


Figure 2: The illustration of goal labeling based on our rosnode

Navigation goals are labeled by the human-computer interaction strategy we developed. The whole operation for goals labeling is shown in Fig. 2. First, the rospkg map_server takes the PGM format maps as input and publish them on the rostopic /map. The map can then be visualized by the software rviz. When the operator uses the mouse to click at the green point in the map, the rosnode publish_point transfers this action into the rostopic /clicked_point. After that, the rosnode goal_labeling translates the messages under the /clicked_point into markers. The location values of these markers are finally saved sequentially as goals in a matrix file. For more information about the ROS, please refer to its official website ROS wiki: <http://wiki.ros.org>.

3.2.3. Navigation with Recording

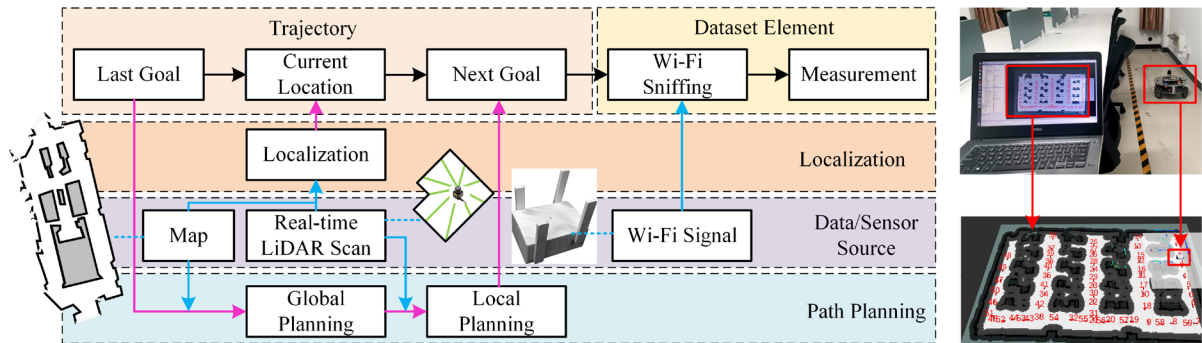


Figure 3: The left: the whole process and data stream of the module navigation and recording. The right: the site experiment of navigation and recording in scene 1, the red boxes indicate the navigation map with goals and the robot, respectively.

This module details the Wi-Fi RSS levels recording in the indoor environment based on the navigation strategy. The navigation strategy is constructed on the algorithms of localization and path planning. The localization method applied is Adaptive Monte Carlo Localization. It is mainly fed with the map and the real-time LiDAR scan to compute the robot's location on the map.

The path planning methods are composed of the global planning algorithm Dijkstra and the local planning algorithm Time Elastic Band (TEB). The global planning algorithm utilizes the obstacle information presented in the map to generate a viable global path from the current location to the desired destination. The local path planning makes adjustments to the global path based on the real-time LiDAR scan.

Based on the methods above, the robot can travel around between near goals. Therefore, every measurement including the RSS level and the corresponding location values can be recorded by our Wi-Fi sniffing rosnode. The whole process and data stream of the navigation and recording is depicted in Fig. 3.

4. Dataset Presentation

In this section, the detailed elements information of the dataset in each scenario are completely presented. The element tree of this dataset is depicted in Fig.4. The abbreviations and meanings for each member are as follows:

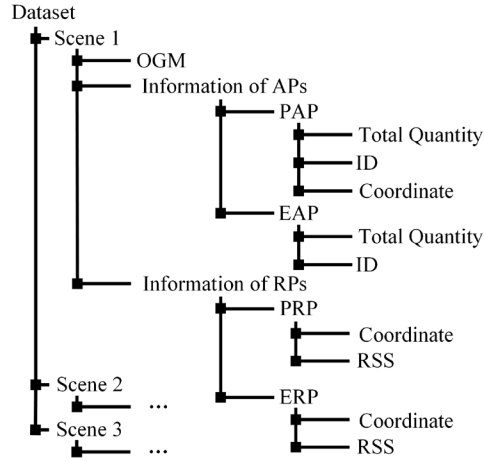


Figure 4: The element tree of this dataset.

- OGM: 2D Occupancy Grid Map of the corresponding experiment site.
- PAP: information of the pre-deployed APs.
- EAP: information of the existing APs.
- PRP: reference points of the pre-deployed APs.
- ERP: reference points of the existing APs.

4.1. Scenarios and OGM

As presented in Fig. 5, the dataset was captured in the Science Education Building and the Shenghua Building in the Main Campus of Central South University(CSU). Three typical indoor scenarios are included: a single room scene 1, an open corridor scene 2, and a whole floor with rooms scene 3. Their corresponding OGM are separately depicted in sub-figure (c), (d), and (e).

4.2. Information of APs

The type of PAP we deployed is TP-Link TL-WDA6332RE, shown in Fig. 5. The number of PAP in each scene (Shown in Table 1) varies due to the layout in the actual scene. Ground truth locations of the PAP in every scene are recorded by the laser ranger in Fig. 5. Ground Truth can be used to quantitatively evaluate the accuracy of AP location estimation methods. Their 2D coordinates values are separately listed in Table 2. Their spatial distributions are marked as red dots in sub-figure (c), (d), and (e) of Fig. 5, respectively.

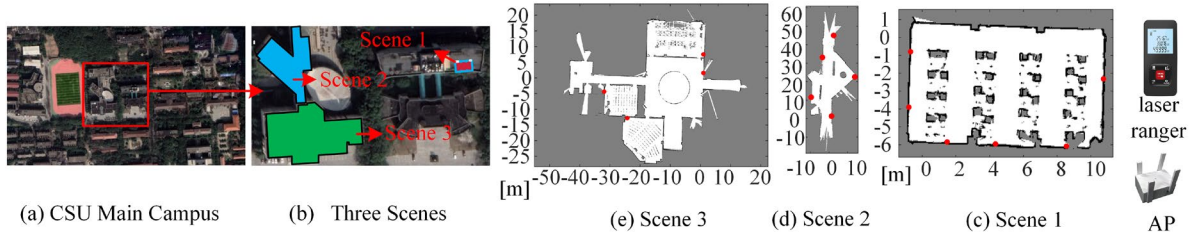


Figure 5: (a), (b) the illustrations of three scenes in the Google Map. (c), (d) and (e) the OGMs of these scenes, red dots in each OGM indicate the locations and the numbers of the PAPs, of which locations are obtained through a laser ranger and listed in Table 1. The unit of OGM is meter.

Table 1

The Number of The PAP and The EPA in Each Scene

Scene ID	Scene 1	Scene 2	Scene 3
PAP Number	6	5	4
EAP Number	122	528	387

Table 2

PAP Location Ground Truth in Each Scene (unit: meter)

Scene ID	Scene 1	Scene 2	Scene 3
AP 1	[-0.685,0.800]	[1.333,1.389]	[0.979,7.500]
AP 2	[10.750,-2.320]	[-9.647,11.620]	[0.979,1.800]
AP 3	[8.558,-6.083]	[13.958,22.621]	[-24.070,-12.730]
AP 4	[4.348,-5.953]	[-3.550,33.171]	[-31.970,-3.580]
AP 5	[1.485,-5.843]	[2.509,45.070]	—
AP 6	[-0.785,-3.900]	—	—

Apart from the pre-configured APs, the existing APs are the most detectable in the indoor environment. Unfortunately, the determination of their locations is extremely difficult. It is not only caused by their large amount but also subjected to personal authorization. Therefore, the ground truth of their locations is not included in the dataset. The main included information consists of two main parts: (1) their number in each scene, which is shown in Table 1; (2) the corresponding RSS at each RP, which is presented in Section 4.3.

4.3. Information of Reference Points

The reference points consist of a series of measurements. The numbers of measurements in each scenario are 2489, 4230, and 4495, respectively, which are listed in Table 3. Each measurement contains an RSS level and a corresponding coordinate in the local map. The RSS levels of all measurements are signals of the PAPs or the EAPs. According to this, the RPs are divided into two categories and named PRP and ERP, respectively. For the coordinates, they are generated by the path planning method on the trajectory or they are directly the goals we set. As previously mentioned in section 3.2.3, the robot records its locations and the RSS levels only once at each coordinate on the path. The robot was set to record 10 times at every navigation goal. However, due to signal propagation issues such as obstruction from obstacles., there may be a difference between the actual measured times and the set value.

Although these RPs can all be fed as input of the mentioned location determination methods, amplitude difference and attenuation during propagation are inevitable. The former is mainly caused by the multipath effect, and the latter is brought by the mixed occlusion from pedestrians, building walls, and the indoor layout. To tackle these, mean values of RSS levels at each RP are recommended.

Table 3

The Number of Measurements in Each Scene

Scene 1	Scene 2	Scene 3
2489	4230	4495

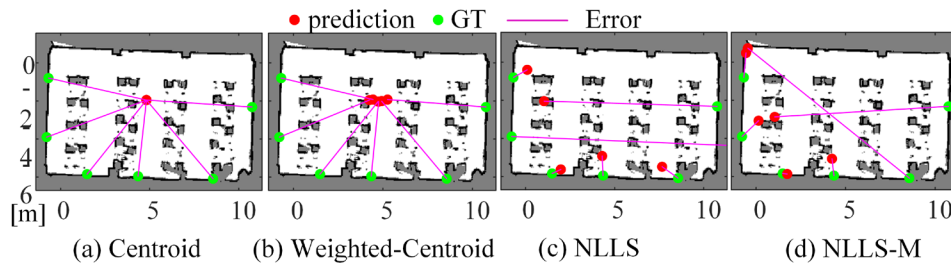
5. Usage Example

This section presents the usage example of our dataset for AP location determination. Some existing methods are tested as the baseline for further comparisons. Note that the objective of this work is not to propose a new indoor AP location determination method, but a common wireless dataset built in a labor-saving strategy, which can be used to test and compare the AP location determination methods.

The example is presented in Scene 1. We have tested the model-free methods including the centroid method[14], the weight-centroid method[15], and the model-based method containing the NNLS[12] and the NNLS-f[16]. Here the information of the pre-deployed AP (PAP) is used to estimate the location value of the corresponding AP. The RSS levels are first filtered by a gaussian filter before solving the problem.

The qualitative results are illustrated in Fig. 6. In each sub-figure, the ground truth of AP locations (green points) and the estimations (red points) of them are drawn in the 2D map. For model-free methods, the results in (a) and (b) indicate that although the weight-centroid method produces better results than the centroid method, the results with these two methods are unsatisfactory. The reason is the measurements are usually not very evenly distributed around the AP without prior knowledge of field applications. The results of model-based methods including (c) Non-Linear Least Squares (NLLS), (d) Non-Linear Least Squares, and Multi-Level Quality Control (NLLS-M) are much better than the model-free methods. However, the results of some APs are unacceptable. Besides, the comparisons with these methods indicate there should be much improvement under new methods provided by researchers in the near future.

To describe more precisely, the quantitative results on AP location determination errors are stated based on Table 4. The metric selected here is the Euclidean distance between the ground truth AP locations and the estimated values from existing methods. In Table 4, as described before, the centroid method (C) and the weighted centroid (C-W) method perform worse than other methods for most AP. The NLLS and NLLS-M can produce better results than the former two for most AP. But the solution is not stable. For example, the location error with NLLS for AP-5 is 0.54m but 121.79m for AP-6. The qualitative findings provide a more precise depiction of the strengths and weaknesses inherent in the existing methodologies, thereby offering preliminary results for subsequent research by scholars.

**Figure 6:** The qualitative results of the tested baselines.**Table 4**

The errors(unit: meter) for AP location determination of existing methods on Euclidean Distance

AP	C[14]	C-W[15]	NLLS[12]	NNLS-M[16]
AP 1	5.62	5.25	0.87	1.29

AP 2	5.94	5.51	9.72	9.73
AP 3	5.57	5.46	1.08	11.3
AP 4	4.02	3.94	1.30	0.89
AP 5	5.12	5.01	0.54	0.26
AP 6	5.93	5.39	121.79	1.24
Mean	5.37	5.09	22.50	4.12

6. Conclusion

This paper provides a new dataset, CSUIndoorAPDet, mainly for indoor AP location determination. First, the collection methodology and workflow of the dataset have been detailed, including the system, equipment, and three sequential modules. After that, the elements of the dataset and the corresponding main information are presented. Next, the usage example and comparisons are applied among the existing methods on one of the scenarios in the dataset, which provides the research community with a baseline of the dataset for further comparisons. In addition, the whole process for RSS level recording is automatic and labor-saving to prevent most path loss caused by the human body. Therefore, researchers who major in the related fields can use this dataset to test their newly provided methods, including AP location determination, AP selection and fingerprint dimension reduction, fingerprint-based Wi-Fi localization, and range-based Wi-Fi localization.

References

- [1] J. Santa, B. Ubeda, R. Toledo and A. F. G. Skarmeta, "Monitoring the Position Integrity in Road Transport Localization Based Services," IEEE Conference on Vehicular Technology (VTC), Montreal, QC, Canada, 2006, pp. 1-5, doi: 10.1109/VTCF.2006.575.
- [2] Y. Long, Y. Gong, Z. Xiao, and Q. Liu, "Accurate Object Localization in Remote Sensing Images Based on Convolutional Neural Networks," IEEE Transactions on Geoscience and Remote Sensing, vol. 55, no. 5, pp. 2486-2498, May 2017, doi: 10.1109/TGRS.2016.2645610.
- [3] T. H. Chiang, Z. H. Sun, H. R. Shiu, K. C. J. Lin and Y. C. Tseng, "Magnetic Field-Based Localization in Factories Using Neural Network With Robotic Sampling," in IEEE Sensors Journal, vol. 20, no. 21, pp. 13110-13118, 1 Nov.1, 2020, doi: 10.1109/JSEN.2020.3003404.
- [4] Yin, C., Ding, S. and Wang, J. "Mobile marketing recommendation method based on user location feedback," Human-centric Computing and Information Sciences, vol. 9, no. 1, 2019, doi: 10.1186/s13673-019-0177-6.
- [5] A. El-Rabbany, Introduction to GPS: The Global Positioning System. Norwood, MA, USA: Artech House, 2002.
- [6] M. Zhou, Y. Li, M. J. Tahir, X. Geng, Y. Wang and W. He, "Integrated Statistical Test of Signal Distributions and Access Point Contributions for Wi-Fi Indoor Localization," IEEE Transactions on Vehicular Technology, 70, no. 5, pp. 5057-5070, May 2021, doi: 10.1109/TVT.2021.3076269.
- [7] N. El-Sheimy and Y. Li, "Indoor navigation: State of the art and future trends," Satellite Navigation, vol. 2, no. 1, p. 7, 2021.
- [8] B. T. Fang, "Trilateration and extension to Global Positioning System navigation," Journal of Guidance, Control, and Dynamics, vol. 9, no. 6, pp. 715-717, 1986, doi: 10.2514/3.20169.
- [9] N. Singh, S. Choe and R. Punmiya, "Machine Learning Based Indoor Localization Using Wi-Fi RSSI Fingerprints: An Overview," IEEE Access, vol. 9, pp. 127150-127174, 2021, doi: 10.1109/ACCESS.2021.3111083.
- [10] D. Quezada-Gaibor, J. Torres-Sospedra, J. Nurmi, Y. Koucheryavy and J. Huerta, "Lightweight Wi-Fi Fingerprinting with a Novel RSS Clustering Algorithm," 2020 International Conference Indoor Positioning Indoor Navigation (IPIN), Lloret de Mar, Spain, 2021, pp. 1-8, doi: 10.1109/IPIN51156.2021.9662612.

- [11] Henri Nurminen, Marzieh Dashti, Robert Piché, "A Survey on Wireless Transmitter Localization Using Signal Strength Measurements", *Wireless Communications and Mobile Computing*. 2017, Article ID 2569645, 12 pages, 2017, doi:10.1155/2017/2569645.
- [12] J. Koo and H. Cha, "Localizing WiFi Access Points Using Signal Strength," in *IEEE Commun. Lett.*, vol. 15, no. 2, pp. 187-189, February 2011, doi: 10.1109/LCOMM.2011.121410.101379.
- [13] K. N. R. S. V. Prasad and V. K. Bhargava, "RSS Localization Under Gaussian Distributed Path Loss Exponent Model," *IEEE Wireless Communications Letters*, vol. 10, no. 1, pp. 111-115, Jan. 2021, doi: 10.1109/LWC.2020.3021991.
- [14] N. Bulusu, J. Heidemann, and D. Estrin, "GPS-less low-cost outdoor localization for very small devices," *IEEE Personal Communications*, vol. 7, no. 5, pp. 28-34, 2000.
- [15] J. Blumenthal, R. Grossmann, F. Golatowski, and D. Timmermann, "IEEE International Symposium on Intelligent Signal Processing," *Proc of IEEE Int. Symp. Intelligent Signal Process.*, Alcalá de Henares, Spain, October 2007.
- [16] Y. Zhuang, Y. Li, H. Lan, Z. Syed and N. El-Sheimy, "Wireless Access Point Localization Using Nonlinear Least Squares and Multi-Level Quality Control," *IEEE Wireless Communications Letters*, vol. 4, no. 6, pp. 693-696, Dec. 2015, doi: 10.1109/LWC.2015.2483509.
- [17] W. Hess, D. Kohler, H. Rapp, and D. Andor, "Real-time loop closure in 2D LIDAR SLAM," 2016 *IEEE International Conference on Robotics and Automation (ICRA)*, Stockholm, Sweden, 2016, pp. 1271-1278, doi: 10.1109/ICRA.2016.7487258.