IEICE Communications Express, Vol.13, No.6, 172–175

ComEX

Special Cluster in Antennas and Propagation Technologies 2023

LETTER

Machine learning-based area estimation using data measured under walking conditions

Shota Nakayama[1], **Satoru Aikawa**[1, a)], **and Shinichiro Yamamoto**[1]

Abstract This study examines the accuracy and measurement costs associated with room-level indoor-area estimation using a wireless LAN. Utilizing fingerprinting, a method that compares user-measured access point (AP) information with pre-existing AP data from service providers, this study introduces a cost-effective approach. Our proposed machine learning (ML)-based method leverages data collected by users while traversing different locations within an area, thereby significantly reducing the measurement time. Furthermore, this study contrasts the effectiveness of convolutional neural networks (CNN) and support vector machines (SVM) in area estimation using this novel measurement technique. Both CNN and SVM demonstrated comparable accuracy, with SVM exhibiting a shorter processing time.

Keywords: fingerprinting, indoor location estimation, CNN, SVM, area estimation

Classification: Navigation, guidance and control systems

1. Introduction

Recent advancements have led to the widespread use of satellite-based positioning, such as GPS, for outdoor location estimation. However, these methods are inadequate in indoor, underground, or radio wave-obstructed environments. Consequently, there has been increasing focus on indoor location estimation using various wireless signals [1]. Our study explores room-level indoor location estimation (indoor-area estimation) employing Wi-Fi-based fingerprinting, capitalizing on the existing infrastructure of access points (APs) and eliminating the need for additional hardware installation.

We employ two types of AP information: DB (database), AP information premeasured by service providers, and UD (user data), AP information measured by users. This fingerprinting method estimates areas by comparing DB and UD related to the area [2]. In scenarios such as rooms or stores separated by walls or doors, exact point estimation every few meters within the area is unnecessary. Thus, area estimation is more cost-effective than point estimation, which requires estimating the location of the user at every reference point.

Our goal is to minimize the time and cost of service providers in location estimation. We propose a method in which the data necessary for location estimation are measured while the measurer is moving within an area. In addition, room-level area estimation using fingerprinting integrated with ML has been previously explored and has demonstrated high accuracy [3, 4, 5]. Our study investigated the application of our proposed measurement method in conjunction with ML to achieve cost-effective and accurate location estimation.

In our previous work, we implemented a CNN-based area estimation using this measurement method [6]. Our current focus is on the SVM, a less complex alternative to deep learning methods [7], for room-level location estimation. We aim to assess the potential time savings of SVM while maintaining accuracy, and determine which ML method is most effective in indoor environments.

2. Indoor location estimation

2.1 Fingerprinting

Fingerprinting involves comparing DB (database) with UD, which includes AP information, such as the received signal strength indicator and the MAC address of each AP. Service providers initially establish reference points, measure the AP information at these points, and compile this information into a DB before launching their service. Users then measure the AP information at their current location (UD) and compare it with the DB for each point. The point with the highest similarity is considered the estimated point. Various methods, including the mean squared error and ML, are used for this comparison. In our study, CNN and SVM were employed to compare the DB and UD.

Fingerprinting benefits from APs already installed in many locations, such as shopping malls, thereby reducing the cost of implementing location-estimation systems. Another advantage is its applicability even when AP locations are unknown.

2.2 Point and area estimation

Fingerprinting encompasses two location-estimation methods: point and area estimation. As illustrated in Fig. 1(a), point estimation estimates the user's location among predefined points in the DB. The result is the point closest to the user's actual location.

Conversely, area estimation, as shown in Fig. 1(b), determines the users location within a defined area, not a specific point. This method is applied to locations with clear boundaries, such as rooms in buildings or stores in indoor commercial facilities, such as underground shopping malls. Area estimation, which is less precise than point estimation,

[1] Graduate School of Engineering, University of Hyogo, 2167 Shosha, Himeji-city, Hyogo 671-2280, Japan

[a)] aikawa@eng.u-hyogo.ac.jp

DOI: 10.23919/comex.2024SPL0012
Received December 20, 2023
Accepted February 21, 2024
Publicized March 12, 2024
Copyedited June 1, 2024

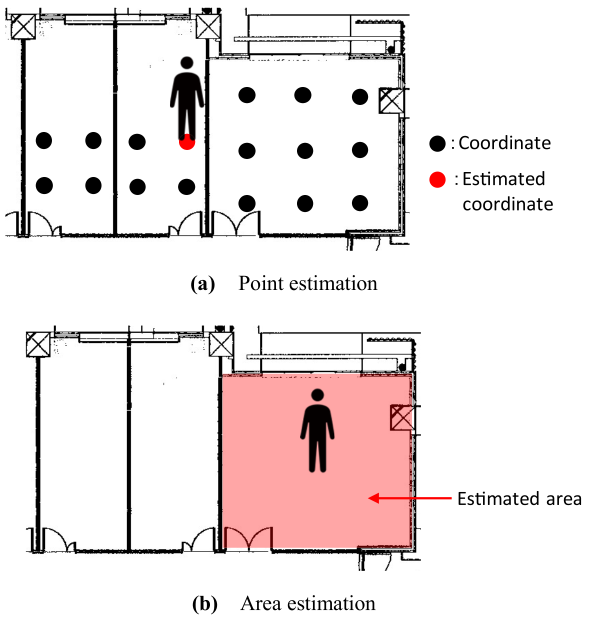


Fig. 1 Location-estimation methods

is expected to reduce the DB creation costs.

The premeasured AP information at multiple reference points is combined into a single area as a measurement method for area estimation. First, the service provider moves to set points and records their positions. Next, they must remain at these points for an extended duration to measure the AP information. Thus, this method increases the measurement time cost.

2.3 ML

Our study utilized a CNN and an SVM for location estimation. CNN are adept at extracting features from 2D data, such as images. For location estimation, the AP information is transformed into a 2D array to facilitate CNN application [8]. Although this method is computationally intensive and time consuming, it yields highly accurate results.

In contrast, the SVM learns input parameters from the training data, aiming to maximize the margin to the closest data point to the hyperplane. In location estimation, SVM uses AP information from each reference point or area as input parameters [9]. SVM is simpler and faster than CNN but may be less accurate depending on the specific task.

3. Proposed method

The proposed method uses walking measurements to realize ML-based area estimation to reduce measurement costs while maintaining a high estimation accuracy.

3.1 Measurement in the walking state

Prior studies on area estimation involved data collection at multiple fixed points within an area, leading to increased measurement costs. To address this issue, our method measures AP information while in motion. The individual conducting the measurement, either a service provider or a user, holds the device at the chest level and walks throughout the target area. Unlike previous methods that involved stationary measurement at fixed points, our approach captures AP information throughout the entire area, including between set points. This strategy not only reduces measurement costs, but also gathers data from various locations and orientations, providing a richer dataset for training ML models.

3.2 ML area estimation

Previous research has demonstrated the effectiveness of ML in achieving high point estimation accuracy [9, 10]. Accordingly, we initially applied a CNN for accurate area estimation [6]. However, considering the relative simplicity of area estimation tasks, we hypothesized that sufficient accuracy could be achieved with reduced processing time using a simpler ML technique. Therefore, our proposed method employs SVM, which is known for its simplicity and shorter processing time compared with CNN.

4. Verification

4.1 Verification method

The verification was conducted in six rooms at the University of Hyogo, as illustrated in Fig. 2. Two measurement approaches were utilized: "Moving" (measurement in the walking state) as depicted in Fig. 2(a), and "Multipoint" (measurement at multiple fixed points) as shown in Fig. 2(b). Both CNN and SVM were used to compare the DB and UD in these scenarios. Four validation tests were conducted using the combinations of DB and UD.

During measurement, the mobile device was held at the chest level. For the Moving measurement, the data collection followed the path indicated by the arrow in Fig. 2(a). In contrast, for the Multipoint measurement, data were collected facing north in each circle marked in Fig. 2(b).

Each area underwent 20 AP scans for both DB (training data) and UD (test data), with each scan taking approximately 3 s, using the terminal in our test environment. To ensure consistency across methods, the total number of AP scans per area was calculated. For instance, in area A with 20 DB measurements, the Moving method involved 20 walking measurements, whereas the Multipoint method entailed 5 measurements per point across 4 coordinates.

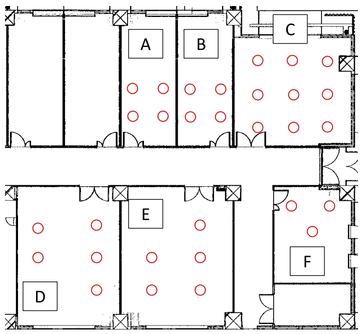
We conducted the tests under two conditions: with all doors open and closed. The open-door scenario simulates environments such as underground malls, whereas the closed-door setting represents enclosed spaces such as rooms in a building. This study also compared the accuracy and processing time of area estimation using CNN and SVM techniques.

4.2 Estimated results

Figure 3 presents the validation outcomes. The accuracy of the estimation method was evaluated based on the correctness rate, which is defined as the percentage of accurate results among all the estimations.

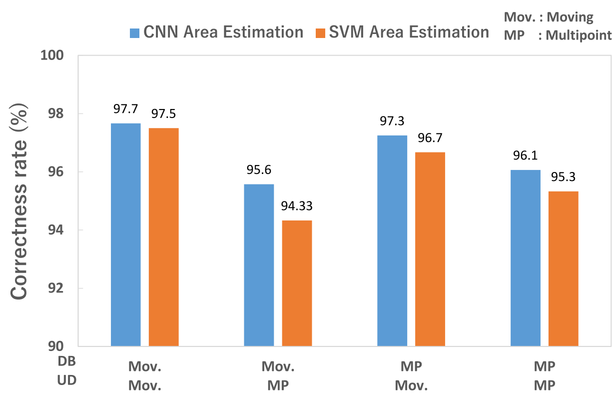
As depicted in Fig. 3, there was no notable difference in the correctness rate between the Moving and Multipoint measurements when applied to the DB. This finding indicates that both the methods yield highly accurate estimates. Crucially, the Moving measurement, our proposed method,

(a) Point estimation

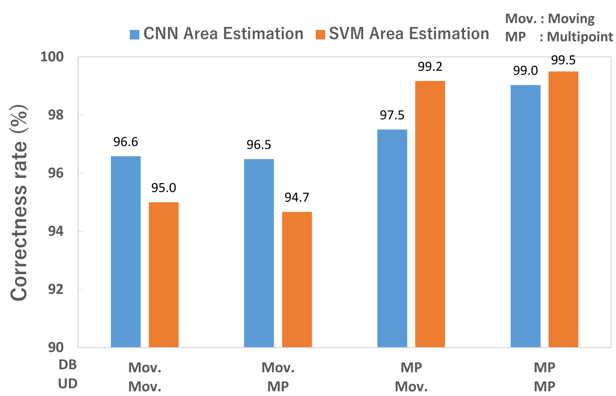


(b) Multipoint

Fig. 2 Experimental environment



(a) Open



(b) Close

Fig. 3 Estimation results for each measurement method

maintains accuracy while significantly reducing the measurement efforts.

In terms of the comparison methods between DB and UD, CNN, and SVM, there was no significant difference in the correctness rate. This suggests that the SVM, despite its simplicity, is capable of delivering high accuracy in room-level area estimation.

Furthermore, Fig. 3(a) reveals that the correctness rate in an "Open" environment is lower due to the complexity of the radio environment. In contrast, as shown in Fig. 3(b), the "Close" environment exhibits a higher correctness rate, which is attributable to well-separated areas and distinct radio environment characteristics in each area. Additionally, the higher accuracy of the CNN method in complex environments, as shown in Fig. 3(a), suggests that CNN-based area estimation is more effective than the SVM in such settings.

4.3 Measurement and processing times

This section compares the measurement and processing times of the two methods (Moving and Multipoint) and the two ML techniques (CNN and SVM).

As indicated in Section 4.2, both the Moving measurement and ML methods demonstrated high accuracy in ML area estimation. We evaluated these methods based on the measurement time. Recall that the device used in this experiment required approximately 3 s for a single AP scan. This implies that 20 AP scans across the six areas can be completed in 360 s. However, the Multipoint method, which requires the measurer to move and confirm the positions at each point, requires a considerably longer duration.

Our findings show that the Multipoint method takes about 1.6 times longer than the Moving method in our experimental environment. The time disparity increased with an increase in the number of AP scans, areas, and points.

Comparing the training times of the two ML methods, the CNN method required 15.0 s, while the SVM method took only 6.68 s in our verification environment. Moreover, in terms of processing time (the time required to produce estimation results using the trained model), CNN required 0.291 s and SVM required only 0.0312 s.

Consequently, combining the Moving measurement technique with the SVM method for comparing UD and DB enables efficient and accurate area estimation, while substantially reducing time costs.

5. Conclusion

This study introduced the Moving measurement method as a means to minimize the premeasurement costs for service providers in indoor location estimation using fingerprinting. Our findings indicate that this method achieves accuracy comparable to that of the traditional Multipoint measurement approach. Thus, the Moving measurement method effectively reduces the measurement costs while preserving the accuracy.

Additionally, we evaluated CNN and SVM as ML methods for area estimation. The accuracy of the SVM was found to be comparable to that of CNN, with the added advantage of reduced processing time. This suggested that SVM, when used with Moving measurement data, is an efficient approach for service providers seeking to reduce time costs.

However, in more complex wireless environments, CNN demonstrated a higher rate of correct responses than SVM. Consequently, future research should focus on identifying environments in which CNN is most effective, allowing for more tailored and efficient indoor location-estimation solutions.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number 21K04065.

References

[1] S. Shang and L. Wang, "Overview of WiFi fingerprinting-based indoor positioning," *IET Communications*, vol. 16, no. 7, pp. 725–733, April 2022. DOI: 10.1049/CMU2.12386

[2] Y. Tanaka, S. Aikawa, and S. Yamamoto, "Updating method using mixture database in area estimation by finger printing," *IEICE Communications Express*, vol. 10, no. 3, pp. 161–166, March 2021. DOI: 10.1587/comex.2020XBL0172

[3] X. Song, X. Fan, C. Xiang, Q. Ye, L. Liu, Z. Wang, X. He, N. Yang, and G. Fang, "A novel convolutional neural network based indoor localization framework with WiFi fingerprinting," *IEEE Access*, vol. 7, pp. 110698–110709, Aug. 2019. DOI: 10.1109/ACCESS.2019.2933921

[4] K.S. Kim, S. Lee, K. Huang, "A scalable deep neural network architecture for multi-building and multi-floor indoor localization based on Wi-Fi fingerprinting," *Big Data Analytics*, vol. 3, no. 4, pp. 1–17, April, 2018. DOI: 10.1186/s41044-018-0031-2

[5] B.A. Akram and A.H. Akbar, "Wi-Fi fingerprinting based room level indoor localization framework using ensemble classifiers," *Mehran University Research Journal of Engineering & Technology*, vol. 38, no. 1, pp. 151–174, 2019. DOI: 10.22581/MUET1982.1901.13

[6] S. Nakayama, S. Aikawa, S. Yamamoto, "CNN area estimation using data measured in walking conditions," ICETC, no. O3-4 Dec. 2022. DOI: 10.34385/proc.72.O3-4

[7] Y. Rezgui, L. Pei, X. Chen, F. Wen, and C. Han, "An efficient normalized rank based SVM for room level indoor WiFi localization with diverse devices," *Mobile Information Systems*, vol. 2017, pp. 1–19, 2017. DOI: 10.1155/2017/6268797

[8] S. Aikawa, S. Yamamoto and T. Muramatsu, "CNN localization using AP inverse position estimation," 2019 IEEE Conference Antenna Measurements and Applications (CAMA2019), pp.1–3, Oct. 2019. DOI: 10.1109/CAMA47423.2019.8959663

[9] L. Xiao, A. Behboodi and R. Mathar, "A deep learning approach to fingerprinting indoor localization solutions," 2017 27th International Telecommunication Networks and Applications Conference (ITNAC), Melbourne, VIC, Australia, pp. 1-7, Nov. 2017. DOI: 10.1109/ATNAC.2017.8215428

[10] A. Mittal, S. Tiku, and S. Pasricha, "Adapting convolutional neural networks for indoor localization with smart mobile devices," GLSVLSI, pp. 117–122, May 2018. DOI: 10.1145/3194554.3194594