Updating Method Using Mixture Database in Area Estimation by Finger Printing

Yuhi Tanaka¹, Satoru Aikawa¹,a), and Shinichiro Yamamoto¹
¹ Graduate School of Engineering, University of Hyogo,
2167 shosha, Himeji city, Hyogo 671-2280, Japan
a) aikawa@eng.u-hyogo.ac.jp

Abstract: In our laboratory, we study “area estimation” by Finger Printing for indoor commercial facilities and buildings. We study a method for updating DBs with user-acquired AP information. There is a problem of lowering the estimation accuracy due to the bias of the updated DBs. In order to solve this problem, we proposed a method to create a large distribution by mixing multiple databases.

Keywords: Finger Print, indoor position estimation, wireless LAN

Classification: Navigation, Guidance and Control Systems

References

[1] S.Fujita, K.Kaji, N.Kawaguchi “A Wireless LAN Location Estimation Method based on Gaussian Mixture Model”
[2] D.Taniuchi, T.Maekawa “A Wireless LAN Location Estimation Method based on Gaussian Mixture Model”
[3] T.Matsui, S.Aikawa, S.Yamamoto “Fingerprint positioning method in inter-coordinate estimation using weighted average of existence probability distribution output of 3-layer NN” IEICE Communications Express, Vol.9, No.7, 318–323
[4] Y.Tanaka, S.Aikawa “Application of Bayesian Estimation in Finger Print Area Estimation Using User Information to Update Database”
1 Introduction

In this study, we use Finger Printing for area estimation using wireless LAN access points (APs). Area estimation is different from ordinary coordinate estimation. Each store in an indoor commercial facility or each room in a building is considered as one area, and it is estimated in which area the person is in. Finger printing location estimation has the problem of database update. Then, we update the database with the data which the user got. We call this data UUD (Updating User Data) which is linked to the area to be updated. In this paper, we proposed a method called Mixture DB to solve the problem of UUD bias [1-4].

2 Updating DB in Finger Printing position estimation

In Finger Print location estimation, the operator measures AP information in each coordinate in advance and constructs a database (DB). As the DB deteriorates with the passage of time after constructing the DB, the DB needs to be updated due to the movement of the fixtures. Therefore, we update the database using UD (User Data), which is AP information obtained by the user. The AP information is a pair of identifiers (SSID) and radio wave strength (RSSI) obtained from the AP. The information of the coordinates to be updated to the UD is collected as UUD (Updating User Data).

3 Area estimation

3.1 Spatial distribution of UD

In the coordinate estimation, the DB in which the AP information measured by the operator in advance at each coordinate corresponds to one coordinate is used. On the other hand, in the area estimation, the AP information measured by the operator at multiple locations within an area corresponds to one area in advance. In addition, when updating the DB using UD, users acquire data at various locations in the area and do not always measure at the locations measured by the operator beforehand. Therefore, the updated DB will be a DB of an area with AP information measured at spatially distributed points.

3.2 Distribution of UD over time

It is known that the RSSI is dispersed over time because of multipath fading and its distribution shows a normal distribution. Figure 1 shows the data measured at the University of Hyogo on the same day and at the same coordinates for one hour each. The horizontal axis is the RSSI and the vertical axis is the number of observations. As it can be seen from the figure, the RSSI is dispersed in the data at each time. In addition, we can see that the distribution of the RSSI is significantly different between daytime and nighttime, even at the same point on the same day, due to the influence of changes in the radio environment caused by people and doors. When the respective data are combined, we can see that they show a mixed normal distribution.
3.3 Area estimation method
As described in 3.1 and 3.2, the variance of RSSI in a DB updated with a spatially and temporally distributed RSSI as a UUD is high. When the mean square error (MSE) is used for position estimation, we determine the RSSI for each AP in the coordinates and compare this value with the UD to obtain the similarity. Here, the average value is used for RSSI in DB, so the error becomes large in DB with high distribution. To solve this problem, we use the method of estimation from the similarity of the RSSI distribution. We use the method based on the similarity of the distributions for area estimation where the influence of spatial dispersion is significant.

3.4 Definition of area estimation
Area estimation is the estimation of which area one is in by taking each store or room in a building in an indoor commercial facility as one area. The radio wave environment differs from room to room, which is separated by walls and doors. We also aim to estimate in which stores the services we are targeting in this laboratory are located. For these two reasons, we use the area as a store or room unit.

3.5 Location Estimation Method Using Distributions
The similarity Pa of DBa (DB in area a) to UD is shown in equation (1) [4]. In the DB for location estimation, there are multiple DBs with AP information for each area.

\[ P_a = \sum_{i=1}^{n} N_i \sum_{r=-\infty}^{\infty} \min\{P(\text{UD}, r, AP_i), P(DB_a, r, AP_i)\} \]  

(DBa: DB in area a
i: Number of APs common to DB and UD
r: Class of RSSI
Ni: Number of observations at UD for APi
P: Probability density of the obtained RSSI)
Smaller probability of DB or UD for each discrete RSSI class \( r \), are summed together. This is added to the number of APs common to both DB and UD. At this time, multiply the value obtained for each AP by the number of times the AP was observed, \( N_i \), as the weight. This is to reduce the impact of APs with less \( N_i \) on the estimation results. This is calculated for each area, and the maximum area is the estimated result.

### 4 Proposal

Figure 2 is a schematic diagram of the proposed method. The horizontal axis represents time, the cylindrical figure represents DB, and the square figure represents UD. In the conventional method, the DB is updated in anticipation of long-term changes in the radio wave environment, such as the movement of fixtures, etc. Therefore, a new DB is created every time the DB is updated. However, as shown in Fig. 1, the accuracy of the conventional method is expected to be lower because the distribution is different even in the short term, as shown in Fig. 1. This corresponds to the case where daytime data is DB and nighttime data is UUD as shown in Figure 1. In addition, since the UUD is used to update the database, the data may not be uniform in terms of user location and time period. As a result, the time and space of the UUD may be unevenly distributed in the conventional method. If we update the UUDs with biased UUDs, the estimation accuracy will decrease. For example, if we use a database updated in the daytime and perform location estimation at night, the estimation accuracy will decrease. It would be ideal to build a database for each condition such as time of day and day of the week, but it is not realistic due to the cost. Therefore, we propose a method to create a new DB with different DBs by mixing DBs with different time zones and measurement environments. We call the DB created by this method Mixture DB (MDB). This MDB shows the gray distribution (All) in Figure 1. We believe that this method will improve the estimation accuracy.

![Fig. 2. Outline of the proposed method map](image)
5 Verification

In this chapter, we verify (1) the decrease in position estimation accuracy due to the difference in radio wave environment and (2) the improvement of position estimation accuracy by MDB of the proposed method. The verification experiment was conducted in this university. The target area is six areas. Each room is one area. The measurement was performed by changing the opening/closing pattern of the door in three rooms. This makes the difference of the radio wave environment. We measured the door opening/closing patterns of 12 patterns in the 6 rooms, with 100 coordinates in total, and 20 measurements in each coordinate.

The similarity between the DB and the UD was evaluated using the equation (1). The accuracy is evaluated by estimating in which area of the six areas the data measured at a given coordinate is located. Here, accuracy is the rate of correct area estimation.

(1) As described in Section 3, the distribution of RSSI differs depending on the radio wave environment such as the time of day and the surrounding environment even at the same location. The accuracies were compared for UD with different numbers of doors opened and closed. The vertical axis in Fig. 3a is the accuracy and the horizontal axis is the number of doors closed. As you can see in Fig. 3a, the accuracy decreases as the difference in the radio environment increases. This is assumed to occur when the UUDs are updated with UUDs measured only under biased conditions, such as time of day and user distribution.

(2) As described in Section 3, the bias of UUDs causes the accuracy of MDBs to deteriorate. In this paper, we verify the proposed method to make MDBs. The data is the same as in verification (1). We update the DB and the UD, which have the largest difference in radio wave environment, and compare the accuracy by the type of data used for updating and the rate of updating the DB.

The result is shown in Fig. 3b. The vertical axis shows the accuracy and the horizontal axis shows the ratio of coordinates to update the DB. We can see that the accuracy improves as the update rate increases. As for the change in the data to be updated according to the door opening/closing pattern, the DB updated under the same conditions as the UD data (with all doors closed) was the highest at 99%, followed by the MDB data at 97%. MDB shows that the distribution of DB is a mixture of 12 patterns of measurement data and one distribution.

This result shows that the accuracy of the DB updated with the data of B is 99% as shown in Fig. 2, but the accuracy decreases to 93% when the condition is different, as shown in A. The proposed MDB has a high accuracy of 99% for the data of B. On the other hand, the proposed method, MDB, can estimate with 97% accuracy for both A and B data, but under different conditions, such as A, the accuracy drops to 93%.
6 Conclusion

We proposed a method for creating MDBs to solve the problem that the estimation accuracy is reduced by the bias of UUDs. We confirmed from the experimental data that the accuracy of position estimation is reduced by the difference of the radio wave environment, and it was confirmed that the MDB is able to estimate the position with 97% accuracy for UUD measured under various conditions.