

Research Article

Active RFID Attached Object Clustering Method with New Evaluation Criterion for Finding Lost Objects

Masaya Tanbo, Ryoma Nojiri, Yuusuke Kawakita, and Haruhisa Ichikawa

Graduate School of Informatics and Engineering, The University of Electro-Communications, Chofu-shi 182-8585, Japan

Correspondence should be addressed to Masaya Tanbo; tanbo.masaya@ichikawa-lab.org

Received 6 October 2016; Revised 21 December 2016; Accepted 5 January 2017; Published 27 February 2017

Academic Editor: José R. Amazonas

Copyright © 2017 Masaya Tanbo et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

An active radio frequency identification (RFID) tag that can communicate with smartphones using Bluetooth low energy technology has recently received widespread attention. We have studied a novel approach to finding lost objects using active RFID. We hypothesize that users can deduce the location of a lost object from information about surrounding objects in an environment where RFID tags are attached to all personal belongings. To help find lost objects from the proximity between RFID tags, the system calculates the proximity between pairs of RFID tags from the RSSI series and estimates the groups of objects in the neighborhood. We developed a method for calculating the proximity of the lost object to those around it using a distance function between RSSI series and estimating the group by hierarchical clustering. There is no method to evaluate whether a combination is suitable for application purposes directly. Presently, different combinations of distance functions and clustering algorithms yield different clustering results. Thus, we propose the number of nearest neighbor candidates (NNNC) as the criterion to evaluate the clustering results. The simulation results show that the NNNC is an appropriate evaluation criterion for our system because it is able to exhaustively evaluate the combination of distance functions and clustering algorithms.

1. Introduction

Radio frequency identification (RFID), which involves wireless communication of data to identify RFID tags attached to objects, is considered a key technology in the Internet of Things (IoT) field. In recent years, active RFID tags that use Bluetooth low energy (BLE) technology to communicate have attracted increasing attention. BLE is supported by many mobile operating systems (e.g., Android, iOS, and Windows Phone), and many smartphone products for finding lost objects that use BLE tags have been released. Products developed for finding lost objects use the received signal strength indicator (RSSI) to report the location of the object. However, these products cannot provide sufficient information to identify an object's position; that is, users only know that the lost object is within a certain range and whether it is moving closer or further away. The authors have studied a method to support user in finding lost objects more effectively. The authors hypothesize that users can determine the location of lost objects using information about the surrounding

objects. In this paper, we introduce a method to calculate the proximity between active RFID tags using an RSSI series. Our approach enables the estimation of the group to which the lost object belongs from its proximity to surrounding objects using a distance function and hierarchical clustering. There are many combinations of distance functions and hierarchical clustering algorithms, and this method gives different group estimation or clustering results for different combinations, but there is no criterion for evaluating the clustering results. We propose the number of nearest neighbor candidates (NNNC) as the evaluation criterion.

The remainder of this paper is organized as follows. In Section 2, we describe the requirements of the proposed system and the problems faced in existing methods. Section 3 presents the framework of our system. In Section 4, we propose the evaluation criterion. Section 5 presents the results of evaluation of the existing method using the NNNC. In Section 6, we describe application of the NNNC. Finally, Section 7 concludes the paper and identifies areas for future work.

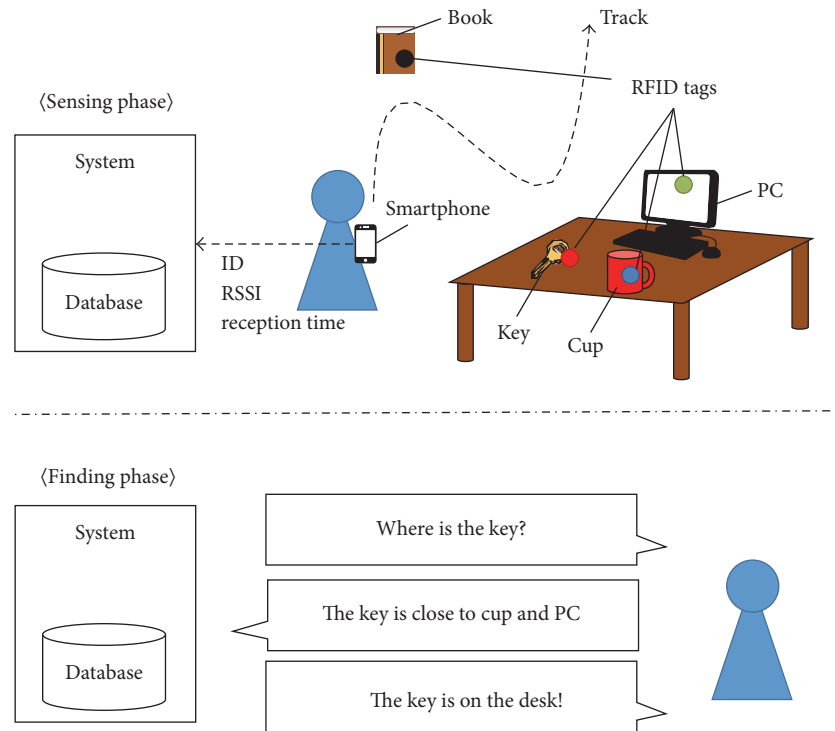


FIGURE 1: A user is moving in a room with a smartphone. The smartphone senses data from RFID tags and sends this data to a support system for finding lost objects. The system records the data in a database and presents information about objects that are close to the lost object. The user can then identify the location of the lost object.

2. Support System for Finding Lost Objects

Finding lost objects is constantly required in peoples' day to day lives. According to published statistical research on finding lost objects [1], common strategies used for finding objects can be classified into five categories: the locus search (33%), exhaustive search (24%), retrace search (19%), memory search (11%), and delegation search (11%). The percentages in parentheses in the preceding list show the fraction of people selecting this technique when finding lost objects. From the locus search, in which the object is normally to be found, the retrace search, which is based on the sequential order of a person's prior physical locations, and the memory search, which is based on a person's recollection of prior interactions with the object, most people can be said to be trying to recall the location of a lost object from memory. We believe that if we are able to present a list of objects that may be located around the lost object, this will aid in the search and thereby compensate for the memory lapses experienced during the locus, retrace, and memory searches.

2.1. System Requirements. As shown in Figure 1, our support system for finding lost objects functions in two phases: sensing data to estimate the group of RFID tags and finding the lost object using information about the proximity of objects. The RFID tags that only transmit beacons are attached to all personal belongings. In the sensing phase, when the user with a smartphone walks around indoors, the smartphone senses data, such as the IDs, measures RSSIs from the RFID tags,

and logs the time of reception. The system records data from the smartphone in a database. In the finding phase, the user inputs the ID of the lost object, and the system estimates the group of objects that are near the lost object from the RSSIs. The user can determine the location of the lost object from the information about its group presented by the system. The basic concept of this system is similar to that of Konishi et al.'s system [2]. Unlike his system, ours uses the RSSIs to estimate object groups and employs a smartphone to collect sensing data. In realizing such a system, we must consider the following requirements.

Input

The system must have access to the ID, RSSI series, and reception time from every RFID tag as input.

Output

The system provides information about groups of objects that are around the lost object as output.

2.2. Problems in Applying Existing Methods to the System. Indoor tracking and localization is a key research issue in indoor applications such as routing and location services. Many studies have been conducted on methods obtaining location information about various objects. However, it is

difficult to apply these existing methods to our system. There are numerous well-known metrics for localization systems, for example, angle of arrival (AOA), time of arrival (TOA), and time difference of arrival (TDOA), but none of these is suitable for smartphones. The AOA [3] measures the relative angle between transmitters from the direction of propagation of a wireless signal using an antenna array. This cannot be employed in common smartphones, as they do not have an antenna array. The TOA and TDOA [4] compute the distance between the transmitter and receiver by using the transmission time. They require accurate time synchronization between the transmitter and receiver for positioning. Therefore, it cannot be applied to our assumed environment where an inexpensive RFID tag is used for the transmitter. In contrast, no special hardware is required to measure the RSSI, and it can be obtained from all transmitters that communicate wirelessly.

Various location estimation methods that use the RSSI have been proposed. For instance, there is a well-known method that computes distance using the RSSI and a channel propagation model that has been created in advance [5–7]. However, multipath fading and interference can cause the RSSI to fluctuate considerably. Accordingly, the computed distance has low accuracy; in addition, users have the burden of creating a channel propagation model for each environment. Location fingerprint methods [8–10] provide accurate position estimates by considering the RSSI from each point as characteristic of that location. Again, users have the burden of creating a characteristic database for each environment. The centroid method [11] and the approximate point-in-triangulation test (APIT) method [12] produce more cost-effective location estimates than the above-mentioned methods. These methods depend on the relative positional relationship between anchor nodes, which have a known position. Users then set up reference nodes in each room, with the estimation accuracy dependent on the number of reference nodes. Overall, therefore, finding lost objects using existing localization methods places a burden on the user. There is no existing method that is suitable for our system.

3. Method Using Distance Function and Hierarchical Clustering

To estimate the group of RFID tags using only RSSIs, we focus on the change in the RSSI values associated with the movement of the receiver. In free space, the RSSIs from RFID tags will decrease with distance according to the Friis equation. This can be expressed as

$$\text{RSSI} = G_T + G_R + P_T - \text{FSPL}(d), \quad (1)$$

where G_T (dB) and G_R (dB) are the gains of the transmit and receive antennas, respectively, in the device and the RFID tag, P_T (dBm) is the transmit power of the RFID tag, and $\text{FSPL}(d) = 20 \log_{10}(4\pi d/\lambda)$ is the free-space path loss at the transmitter-receiver distance d . If the transmit power of the RFID tag is constant, changes in the RSSI with respect to movement in the radial direction will follow the free-space path loss model, because the antenna gains are

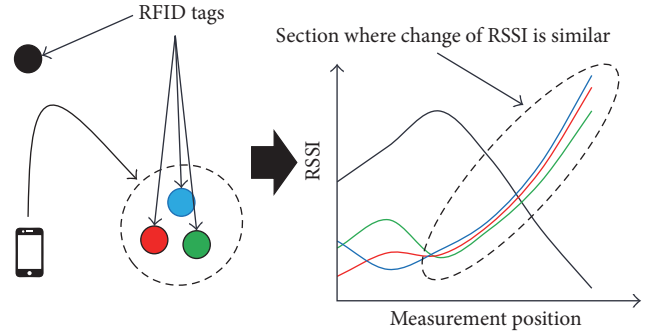


FIGURE 2: Change in RSSI associated with the movement of the smartphone or RSSI sensor.

nearly constant. Accordingly, we consider changes in the RSSI associated with movement in the radial direction of the RFID tags to be similar (Figure 2). From the above, we aim to estimate the nearest neighborhood RFID tag to target the RFID tag attached to the lost object by converting the similarity of RSSI to proximity information.

The present authors have presented a method for calculating the proximity of the lost object to those around it using a distance function between RSSI series and estimating the group by hierarchical clustering in prior work [13]. Figure 3 shows a functional block diagram of the finding phase of the developed support system for finding lost objects. First, a similarity calculation is performed using a distance function. The system uses this to quantify the similarity of RSSIs in the RSSI series. Distance functions define the spatial or temporal difference between two elements in a set. The distances of the multiple elements are given in the form of a matrix, called the distance matrix. These functions are major components used in data mining techniques such as time-series analysis. Therefore, a distance function is appropriate to our challenge, because it has the goal of measuring the similarity among time-series data. Second, groups of objects are estimated using hierarchical clustering. After measuring the relative distance between data in the RSSI series, the system forms clusters of RFID tags in a neighborhood. The hierarchical clustering algorithm exports the clustering result as a matrix called the cophenetic matrix. Finally, the results are displayed as a dendrogram, which is a common method of presenting clustering results. The details about different distance functions and clustering algorithms are shown in Appendix A. Dendrograms display the process of cluster generation and therefore enable the user to intuitively identify objects surrounding the lost object.

For instance, in the locus search, the list of objects around the lost object helps users to find the location of lost object. The information of location in which the surrounding objects are normally to be found makes it easy for users to remind the location of lost object. In addition, in the retrace search and memory search, the lists of time order help users to recall the sequential order of the user's prior physical locations and prior interactions with the object.

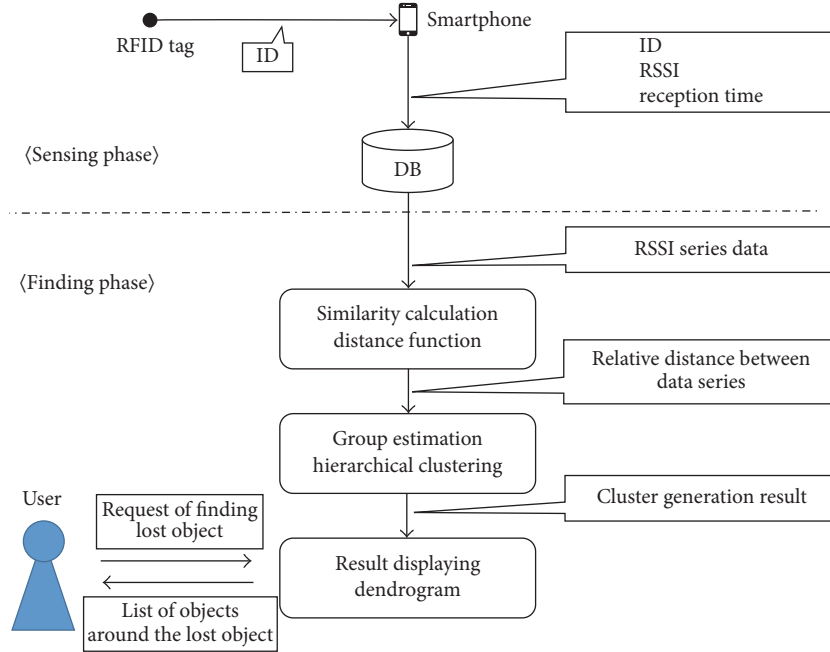


FIGURE 3: Functional block diagram of the support system for finding lost objects.

4. Evaluation Criterion of Group Estimation Accuracy

The methods to search for a nearest neighbor are divided into two categories: hierarchical approach and other. For example, approximate nearest neighbor [14] and locality sensitive hashing [15] are well-known methods for searching a nearest neighbor quickly in a large set of data points in high dimensional space in other than hierarchical approach. In addition, there is a method for attempting to increase the accuracy by combining multiple distance functions [16]. The hierarchical approach uses distance functions and clustering algorithms to search for a nearest neighbor [13]. There are many combinations of distance functions and clustering algorithms. Therefore, the criteria to evaluate clustering results in order to compare the combination of elemental technologies of hierarchical clustering are important.

The cophenetic correlation coefficient is a conventional method to measure the stability of clustering results. It is defined as the Pearson correlation between the distance matrix and the cophenetic matrix. A value of 1.0 means that the concordance between the distance matrix and the clustering result is perfect. With the cophenetic correlation coefficient as a base, we expect that the Pearson correlation between the matrix of the actual distance of the RFID tag and the cophenetic matrix can quantify how well the clustering result reflects the actual position relationship of the RFID tags. However, two problems are encountered while using the Pearson correlation. First, it cannot evaluate whether the combination is suitable for the system directly. Our objective is to estimate the nearest RFID tag to the RFID tag attached to the lost object. The cophenetic correlation coefficient provides information only about linear relationships between the actual distance matrix and cophenetic matrix, and not the

validity of the clustering result directly. Second, it is difficult to determine the threshold to define the goodness. For instance, it is not easy to determine if the calculation result of 0.75 is a good result. Therefore, considerable experimentation is required for defining the threshold to define the goodness. As mentioned earlier, the cophenetic correlation coefficient does not evaluate the correctness of clustering directly. There are some methods, such as Goodman-Kruskal gamma statistic [17] and Mantel test [18] to evaluate the clustering result too. However, they have the same problems as the cophenetic correlation coefficient. From the above, clearly, there is no method to evaluate whether the combination of elemental technologies is suitable for the system using hierarchical clustering.

4.1. Criterion of RFID Clustering Result. As the application of existing methods is not suitable for evaluation in our study, we define the NNNC as a new evaluation criterion. A minimum value of 1.0 means that the clustering algorithm has estimated RFID tags in the nearest neighborhood relationship to be in one cluster firstly and the result is satisfactory. The NNNC may not take a 1.0 even if it is the best clustering result. A purpose of NNNC is to compare the elemental technologies by evaluating a clustering result based on the neighborhood between tags. The NNNC indicates the average number of candidates of the nearest RFID tag to each RFID tag. Hence, NNNC reflects the performance of finding lost objects of the combinations of distance functions and clustering algorithms.

Let T_i ($i = 1, 2, \dots, N$) be N RFID tags. We consider the nearest neighbor matrix P that takes a binary value (0 or 1). If there are N RFID tags, the matrix will have a size of $N \times N$. P_{ij} represents the relationship between the nearest RFID tags by taking a value of 1 when T_i is the nearest RFID tag to T_j .

The correct nearest neighbor matrices X obtained from actual distance matrix A show the correct relationship between the nearest RFID tags. The estimated nearest neighbor matrices Y which are obtained from cophenetic matrix B show the estimated relationship between the nearest RFID tags from clustering

$$X_{ij} = \begin{cases} 1 & (A_{ij} = \min \{A_{i1}, \dots, A_{iN}\}) \\ 0 & (\text{otherwise}), \end{cases} \quad (2)$$

$$Y_{ij} = \begin{cases} 1 & (B_{ij} = \min \{B_{i1}, \dots, B_{iN}\}) \\ 0 & (\text{otherwise}). \end{cases} \quad (3)$$

In hierarchical clustering, element refers to be classified. In this work, the subject is RFID tags. The elements merge in a cluster progressively according to an algorithm, eventually forming one large cluster. In the process, there are cases where an element merges in a cluster that comprises a plurality of elements. If the element is the RFID tag attached to the lost object, it means that the number of candidates to be considered for the nearest neighborhood RFID tag increases to the number of elements in the cluster. Therefore, we multiply the estimated nearest neighbor matrix by the number of elements

$$Y_{ij} = Y_{ij} \times \sum_{k=1}^N Y_{ik}. \quad (4)$$

Next, we multiply the estimated nearest neighbor matrices and the correct nearest neighbor matrices

$$Z = Y \times X. \quad (5)$$

The main diagonal Z_{ii} of Z presents the number of candidates for the nearest neighbor RFID tag to T_i . When $Z_{ii} = 0$, it shows that the nearest neighbor RFID tag to T_i cannot be estimated from the clustering result. Therefore, the value of Z_{ii} is replaced by the number of RFID tags from the estimation.

$$Z_{ii} = N \quad (Z_{ii} = 0). \quad (6)$$

Finally, the NNNC is calculated from the average of Z_{ii}

$$\text{NNNC} = \frac{1}{N} \sum_{i=1}^N Z_{ii}. \quad (7)$$

To support the recall of the location of a lost object, presenting as many objects as possible near the lost object is important. To achieve this, NNNC evaluates the clustering result based on the sequence of the merge cluster. Figure 4 shows an example of the calculation of the NNNC for both good and bad clustering results. The figure in the top right corner shows the actual position of the RFID tag. The figures in the center left and center right show the dendrogram and cophenetic matrix obtained by hierarchical clustering. The clustering result in the center right is a good result that correctly reflects the actual placement of the RFID tags. On

the contrary, the clustering result in the center left does not reflect the actual placement of the RFID tags. In Figure 4, the nearest neighbor RFID tag to T_4 is T_3 . However, a bad clustering result shows that candidates of the nearest neighbor RFID tag to T_4 are T_1, T_2 , and T_3 . In addition, it does not show T_4 as a nearest neighbor RFID tag to T_3 . Therefore, the NNNC of a bad clustering result is increased compared to the good clustering result. We confirmed the validity of NNNC through simulation experiments.

4.2. Indoor Path Loss Model and RSSI Fluctuation. Indoor path loss is necessary for considering fluctuation in addition to the attenuation due to free-space path loss. The shadowing, interference, and multipath fading have been said the main cause of the fluctuation of path loss [19]. First, the shadowing effect has been modeled as a random variable following a zero-mean Gaussian distribution in the log-normal shadowing model [20]. Second, we believe that the effect of the interference is random because we assume that a large number of terminals communicate randomly. Lastly, the fluctuation of the received power due to multipath fading can be modeled as a random value that follows the Nakagami-Rice distribution [21, 22]. Based on the above discussion, we simulated an environment where shadowing, interference, and multipath fading exist by considering χ_σ in the equation:

$$\text{PL}(d) = \text{PL}(d_0) + 10\gamma \log_{10} \left(\frac{d}{d_0} \right) + \chi_\sigma, \quad (8)$$

where d is the transmitter-receiver distance and γ is the path loss exponent. The intercept $\text{PL}(d_0)$ is the path loss in dB at reference distance d_0 and is given by the free-space path loss $\text{PL}(d_0) = 20 \log_{10}(4\pi d_0/\lambda)$. χ_σ (dB) is a zero-mean Gaussian variable with standard deviation σ and represents the shadowing, interference, and multipath fading effect. From (1) and (8), indoor RSSI is calculated as

$$\text{RSSI} = G_T + G_R + P_T - \text{PL}(d_0) + 10\gamma \log_{10} \left(\frac{d}{d_0} \right) + \chi_\sigma. \quad (9)$$

4.3. Verification of the Validity of the Evaluation Criterion. We verified the validity of the evaluation criterion by simulation using MATLAB. We made a virtual room and set three groups of two RFID tags that transmit a radio wave at fixed intervals. The receiver moved straight between two random points at a constant speed and the RSSI was calculated when the RFID tags transmitted the radio waves. Figure 5 shows the placement of the RFID tag, an example of a movement pattern and the calculated RSSI. Table 1 shows the parameters of the simulation. We created 10,000 movement patterns in random and checked whether the NNNC evaluates the clustering result as expected. For evaluation, we defined a score that reflects the number of groups within which RFID tags were placed in their expected group. In this simulation, we placed RFID tags in three groups ((T_1, T_2) , (T_3, T_4) , and (T_5, T_6)). The score was increased by 1 for each clustering result that classified an RFID tag into the correct group. The

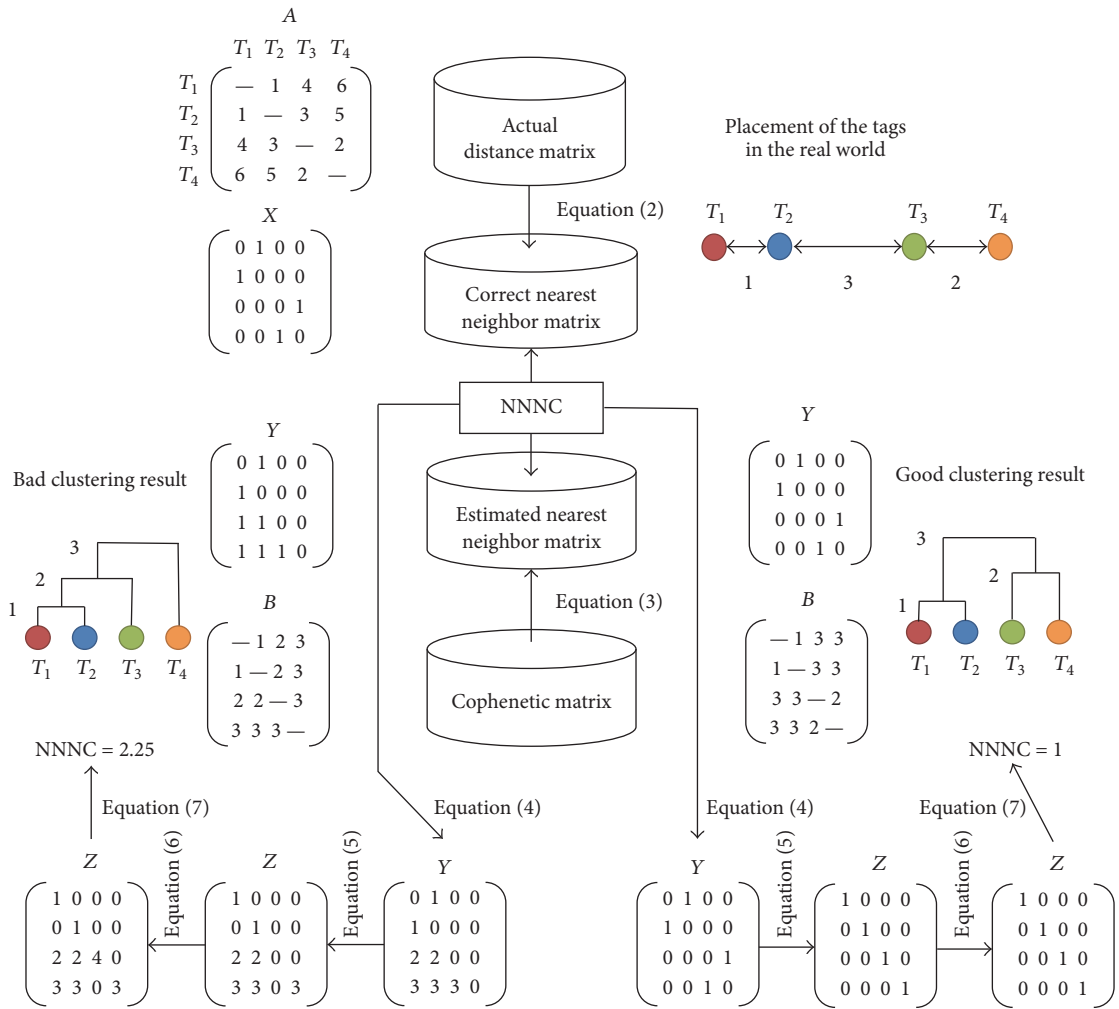


FIGURE 4: The calculation example of the NNNC.

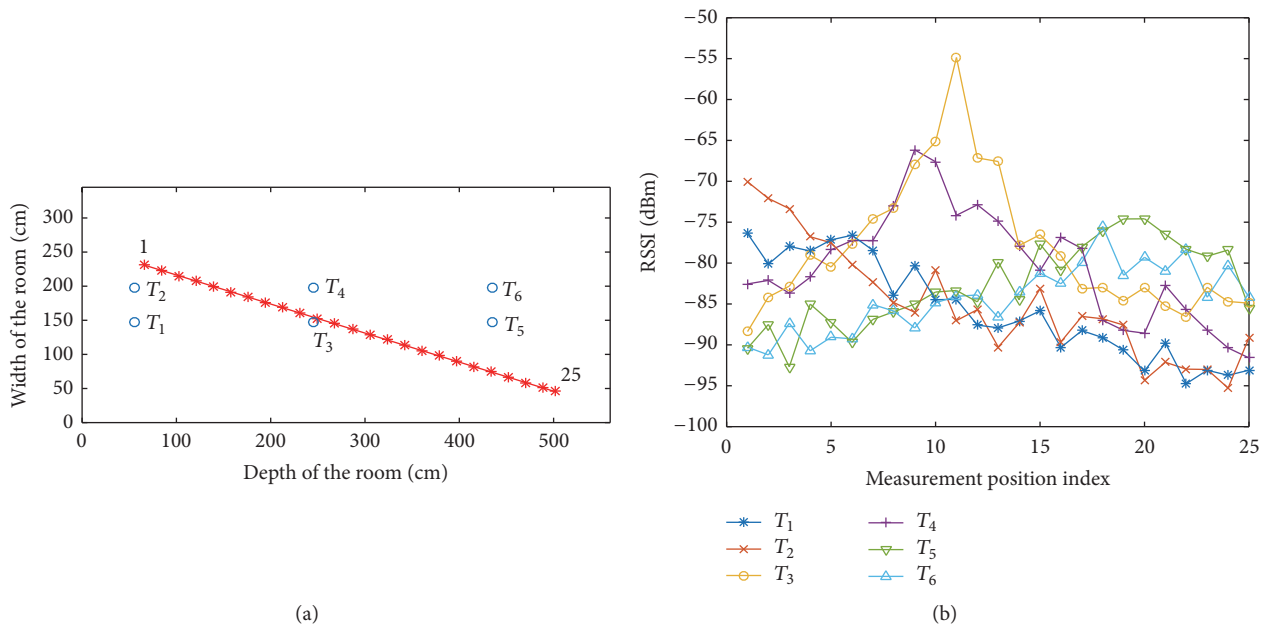


FIGURE 5: (a) Placement of the RFID tag and an example of a movement pattern. (b) Calculated RSSI.

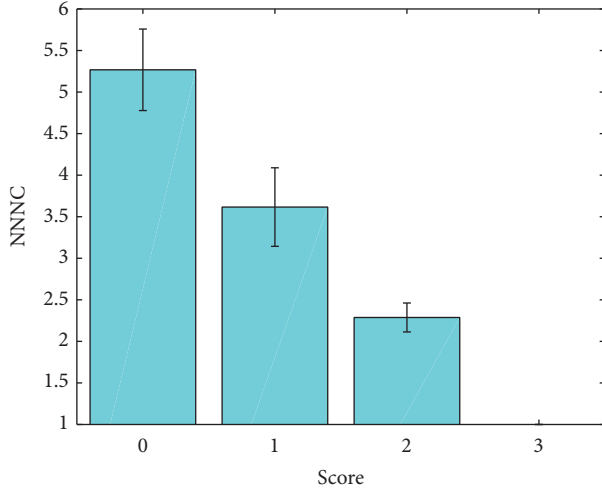


FIGURE 6: NNNCs with different scores in the cosine distance and complete-linkage method.

TABLE 1: Parameters of the simulation used to verify the validity of the evaluation criterion.

Gain of the transmit antenna G_T	0 (dB)
Gain of the receive antenna G_R	0 (dB)
Transmit power P_T	1 (dBm)
Path loss exponent γ	2.0
Standard deviation of zero-mean Gaussian variable σ	2.0
Reference distance d_0	1.0 (m)
Transmission interval	1.0 (s)
Receiver speed	0.2 (m/s)

TABLE 2: NNNCs of the result of the cosine distance and complete-linkage method.

Score	Number of occurrences	Average of NNNC	Standard deviation
0	2067	5.2679	0.49032
1	2859	3.6162	0.47263
2	2108	2.2876	0.17413
3	2966	1	0

maximum score was 3 and the minimum score was 0. Figure 6 and Table 2 show one result of simulation. When clustering classified all RFID tags into the correct group, the NNNC was a minimum. The NNNC increased when the clustering result became unsatisfactory. The result shows that the NNNC is appropriate evaluation criterion to evaluate clustering results.

5. Exhaustive Evaluation of the Distance Function and Clustering Algorithm

In this section, we evaluate the combination of the distance function and the clustering algorithm using the proposed evaluation criterion to determine the group estimation accuracy of the method. Of special interest is verifying whether our system is immune to RSSI fluctuations. If the group

TABLE 3: Parameters of the simulation used for exhaustive evaluation of the distance function and clustering algorithm.

Transmit power P_T	1 (dBm)
Rician K -factor	4.0
Transmission interval	0.5 (s)
Receiver maximum speed	1.0 (m/s)
RSSI series data length	120

estimation accuracy is high in environments where the fluctuation of RSSI is very large because of shadowing, interference, and multipath fading, then our system can be used in various environments such as offices, industrial facilities, and storehouses. In the evaluation experiment for the combination, the evaluation parameters are considered as follows: physical arrangement of the RFID tag, the movement pattern of the receiver, antenna pattern, and effects on the radio wave propagation path such as shadowing, interference, and multipath fading.

5.1. Simulation Result. We simulated the radio wave propagation path, including the shadowing and Nakagami-Rice fading, using QualNet. The room size was 10 m \times 10 m, and 10 parallel RFID tags were placed randomly. The receiver moved in accordance with a random waypoint model; that is, the smartphone started at a random point in the room. Next, a random point was selected as the waypoint and the receiver moved to this waypoint at a random speed ranging between 0 and maximum speed. We used 10 RFID tag placements and 100 movement patterns for each RFID tag placement. Then, we simulated the RSSI series by changing the standard deviation of the RSSI fluctuations from 0 to 8 in increments of 2. The large value of 8.0 of standard deviation of the RSSI fluctuations is typically observed in industrial environments [23]. Figure 7 shows RSSI fluctuations for different sigma values. It can be seen that the trend of change in the RSSI is eliminated by shadowing and the Nakagami-Rice fading. The other simulation parameters are shown in Table 3. After RSSI simulation, hierarchical clustering of the RSSI series was performed for each combination of the distance function and the clustering algorithm, and the NNNC was calculated. Figure 8 shows that the change in the NNNC is associated with increasing RSSI fluctuations, indicated by increasing standard deviation of the RSSI fluctuations. Each plot exhibits an average of 1000 NNNCs. As can be seen from Figure 8, the Euclidean distance and complete-linkage, the unweighted pair group method with arithmetic mean (UPGMA), and Ward's method showed high group estimation accuracies. Overall trends indicate that the NNNC increases linearly with increasing RSSI fluctuation. However, these three combinations restrained the increase of the NNNC. In particular, the combination of the Euclidean distance and Ward's method resulted in the lowest NNNC when the fluctuations were the largest. To evaluate the clustering result from the point of view of finding lost objects, we focus on the value of the NNNC in standard deviation of the RSSI fluctuation being 8. The minimum value of NNNC is approximately 5.5 when using the combination of Euclidean distance and Ward's method.

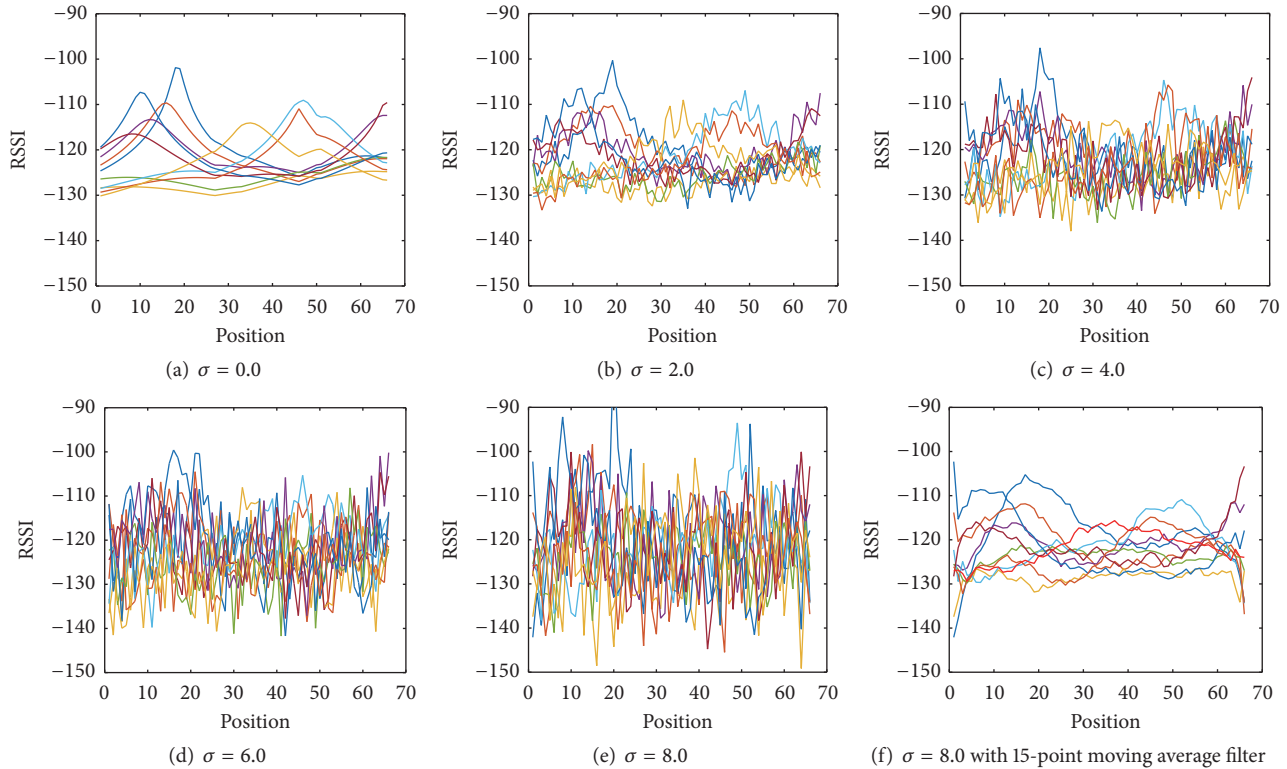


FIGURE 7: Changes in the RSSI when varying the values of σ .

This indicates that the candidate of the nearest neighbor is 5.5. The maximum value of NNNC is approximately 7.0 when using the combination of cosine distance and single-linkage method. This indicates that the ability of the combination of cosine distance and single-linkage method to find lost objects is approximately 1.5 smaller than that of the combination of Euclidean distance and Ward's method. In terms of the distance function, the lowest NNNC was the Euclidean distance as mentioned earlier, followed in order by correlation distances and cosine distances. However, the NNNC did not differ greatly in the case of these distance functions. Compared with other clustering algorithms, the single-linkage method shows particularly unsatisfactory result. There is no big difference between the different algorithms except the single-linkage method. The single-linkage method has a major drawback, known as the chaining phenomenon, whereby one very large cluster is generated as elements are integrated into the cluster one by one. Therefore, the single-linkage method is often used to determine the main cluster owing to its mechanism of combining the closest clusters sequentially. Hence, it follows that the single-linkage method is not suitable for our system.

5.2. Analysis of Influence of Antenna Pattern. We added the antenna pattern to the simulation parameters to analyse its influence on the group estimation accuracy. The antenna pattern of the RFID tag used in the simulation was the pattern measured by the authors (Figure 9). The basic simulation parameters are the same as those shown in Section 5.1.

Figure 10 shows the change in the NNNC when the antenna pattern is as shown in Figure 9. The increasing tendency of the NNNC shown in Figure 10 is similar to the tendency shown in Figure 8. In addition, the difference between the NNNC values in Figures 8 and 10 was very small. Based on the above, we believe that the influence of the antenna pattern of the RFID tag is restrictive.

6. Application of the NNNC

In this section, we describe how users such as system designers use the NNNC. The NNNC is used for selecting elemental technologies of hierarchical clustering for a support system for finding lost objects before the system implementation. The procedure of the selection is as follows.

- (1) The user obtains RSSI series data observed in the environment, where he knows the physical distance between tags and calculates the correct neighbor matrices.
- (2) The user generates clustering results by different elemental technologies.
- (3) The user evaluates the clustering results in terms of NNNC.
- (4) The user selects the technologies indicating the minimum value of the NNNC.

We describe a scenario that applied the NNNC to finding a lost object system as an example of an application and

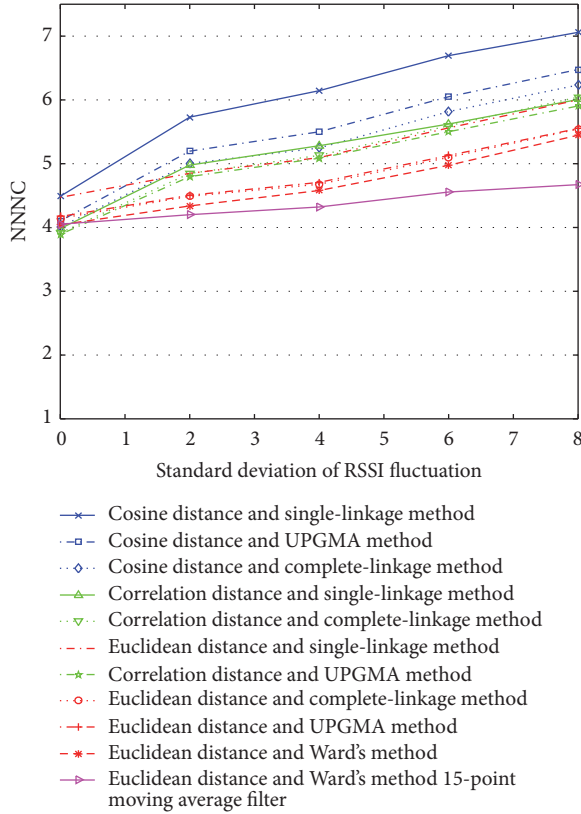


FIGURE 8: The NNCC of each standard deviation of the RSSI fluctuation in each combination. The antenna pattern of the RFID tag is omnidirectional.

discuss how the design process increases the quality of the application. The combination of cosine distance and single-linkage method shows the smallest NNCC, while the combination of Euclidean distance and Ward's method shows the largest NNCC in Figure 10. If the value of the NNCC is large, the number of nearest neighbor candidates is large so that it is difficult to find the lost object. Therefore, the combination of Euclidean distance and Ward's method is concluded as the best method. The result from all combinations shows that the NNCC increase to large values, including the best combination of Euclidean and Ward's method according to the noise increase. This is explained by examining the RSSI series observation data as in Figure 7. That is, it is understood that the correlation between RSSI series becomes difficult to identify on account of the increased noise. As a result, the two groups of nearest neighbors are clustered into the same group even by the best combination. The system designer could conceive the idea of applying a moving average filter to improve the combination. The moving average filter is a common filter for smoothing the time-series data while keeping important patterns and removing unimportant patterns such as noise. For example, Figure 7(f) is obtained by applying a 15-point moving average filter to Figure 7(e), and the combination of Euclidean and Ward's method with the moving average filter shows the smallest NNCC even under the heavy noise in Figure 10. As shown in this example, the NNCC is not

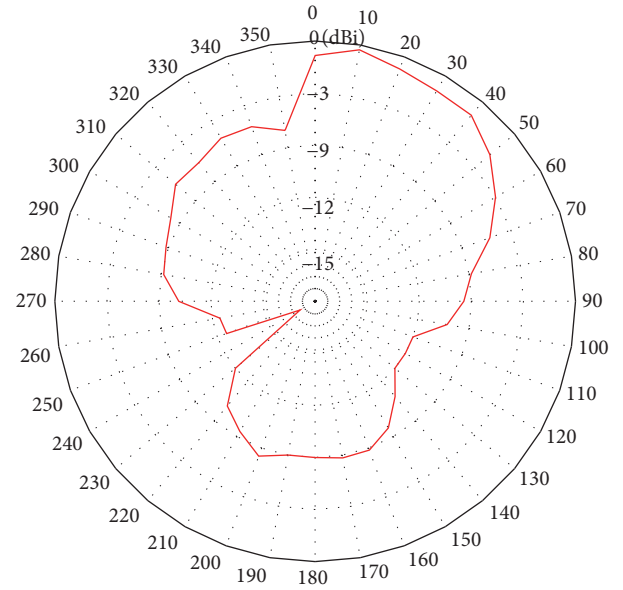


FIGURE 9: The antenna pattern of the RFID tags used in the simulation.

only used for selecting elemental technologies, but is also used to improve technologies for increasing the quality of applications in finding a lost object.

7. Conclusion

In this paper, we introduced a method of finding lost objects in indoor environments using RSSI values and proposed a novel evaluation criterion. We assumed that users can determine the location of lost objects using smartphone applications that determine the proximity between active RFID tags. Our system alerts users regarding the position of a lost object by determining which objects are near the lost object using the RSSI series from the estimated group of RFID tags. The distance function, the clustering algorithm, and the effectiveness of their combination are very important to the successful operation of our system. Hence, there is a need to define a criterion that evaluates the combinations for comparison. The NNCC that we have proposed can evaluate quantitatively the most suitable elemental technologies for systems using the hierarchical clustering. By simulation, we confirmed that the NNCC can suggest the most suitable combination for finding lost objects. When we evaluated the suitability of existing popular distance functions and hierarchical clustering algorithms to our system using the NNCC, we found that the combination of the Euclidean distance and Ward's method yielded the highest group estimation accuracy.

The NNCC could be applied to a nearest neighbor search using hierarchical clustering such as group estimation in a crowd of people in an online-to-offline (O2O) scenario. For example, the movement record and history of visited stores could be considered as feature quantities used in the distance function. It would be possible to use the number of

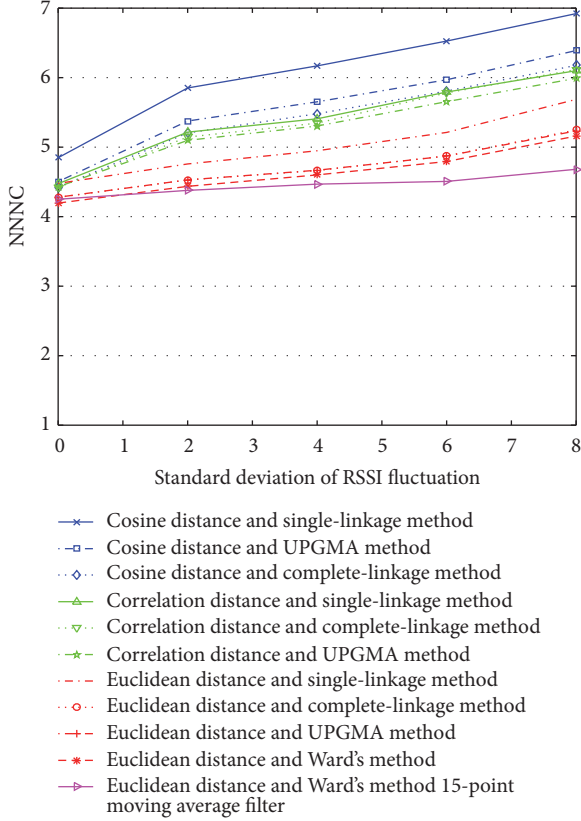


FIGURE 10: The NNCC of each standard deviation of the RSSI fluctuations for each combination. The antenna pattern of the RFID tag is a measured pattern.

people in the estimated group to execute more effective online marketing.

Appendix

A. Details of the Distance Function and Clustering Algorithm

The details of distance functions and clustering algorithms used in this study and in our prior work [13] are as follows.

A.1. Distance Functions. Distance functions define dissimilarity of series data. The magnitude of the value from distance function indicates that the two-time-series data are not similar.

A.1.1. Euclidean Distance. The Euclidean distance is the most popular distance metric. This function measures the distance between two points of a set in Euclidean space. The Euclidean distance d between $P = (p_1, p_2, \dots, p_n)$ and $Q = (q_1, q_2, \dots, q_n)$ is defined as

$$d = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}. \quad (\text{A.1})$$

A.1.2. Cosine Distance. The cosine distance uses the cosine similarity to measure distance. This metric measures the cosine of the angle between two vectors in a series. It takes a maximum value of 1 at 0 degrees, and a minimum of -1 at 180 degrees. The cosine similarity between $P = (p_1, p_2, \dots, p_n)$ and $Q = (q_1, q_2, \dots, q_n)$ is defined as

$$\text{Sim}_{\text{cosine}} = \frac{\sum_{i=1}^n (p_i q_i)}{\sqrt{\sum_{i=1}^n p_i^2} \cdot \sqrt{\sum_{i=1}^n q_i^2}} \quad (\text{A.2})$$

and the cosine distance d is given by

$$d = 1 - \text{Sim}_{\text{cosine}}. \quad (\text{A.3})$$

A.1.3. Correlation Distance. The correlation distance uses the correlation coefficient to measure distance. The correlation coefficient measures the strength of a linear association between two series. It takes a maximum value of 1 when there is a positive association between the two series and a minimum of -1 if there is a negative association. A value of 0 indicates that there is no association. The correlation coefficient between $P = (p_1, p_2, \dots, p_n)$ and $Q = (q_1, q_2, \dots, q_n)$ is defined as

$$\text{Sim}_{\text{correlation}} = \frac{\sum_{i=1}^n (p_i - \bar{p})(q_i - \bar{q})}{\sqrt{\sum_{i=1}^n (p_i - \bar{p})^2} \cdot \sqrt{\sum_{i=1}^n (q_i - \bar{q})^2}}, \quad (\text{A.4})$$

where \bar{p} and \bar{q} denote the average of $P = (p_1, p_2, \dots, p_n)$ and $Q = (q_1, q_2, \dots, q_n)$, respectively, and the correlation distance d is defined as

$$d = 1 - \text{Sim}_{\text{correlation}}. \quad (\text{A.5})$$

A.2. Clustering Algorithms. At the beginning of the hierarchical clustering process, each element is in a cluster of its own. Then, two clusters with the shortest mutual distance are sequentially combined into a larger cluster. This procedure continues until all the elements are included in one cluster. The definitions of mutual distance are different for different clustering algorithms.

A.2.1. Single-Linkage Method. The single-linkage method, also known as the nearest neighbor method, defines the distance between clusters C_A and C_B as

$$d(C_A, C_B) = \min_{a \in C_A, b \in C_B} d(a, b). \quad (\text{A.6})$$

A.2.2. Complete-Linkage Method. In the complete-linkage (or furthest-neighbor) method, the distance between clusters C_A and C_B is defined as

$$d(C_A, C_B) = \max_{a \in C_A, b \in C_B} d(a, b). \quad (\text{A.7})$$

A.2.3. Unweighted Pair Group Method with Arithmetic Mean. The unweighted pair group method with arithmetic mean (UPGMA) combines two clusters with the smallest average

distance between all samples in the clusters. The distance is calculated as

$$d(C_A, C_B) = \frac{1}{|C_A||C_B|} \sum_{a \in C_A} \sum_{b \in C_B} d(a, b), \quad (\text{A.8})$$

where $|C_A|$ and $|C_B|$ are the number of samples in C_A and C_B .

A.2.4. Ward's Method. Ward's method combines two clusters with the minimum change in variance before and after fusion:

$$d(C_A, C_B) = E(C_A \cup C_B) - E(C_A) - E(C_B), \quad (\text{A.9})$$

where $E(P)$ is the variance in cluster P . $E(P)$ is defined as

$$E(P) = \sum_{p \in P} (p - \bar{p})^2, \quad (\text{A.10})$$

where \bar{p} is the centroid of cluster P . Ward's method is known to be a well-balanced clustering algorithm. However, it is computationally expensive and is unsuitable for all distance functions except the Euclidean distance.

Competing Interests

The authors declare that they have no competing interests.

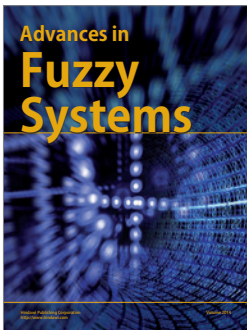
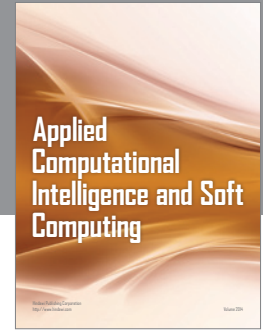
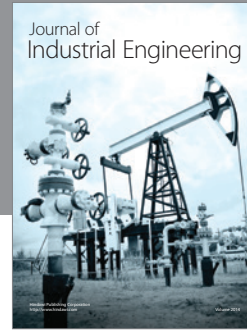
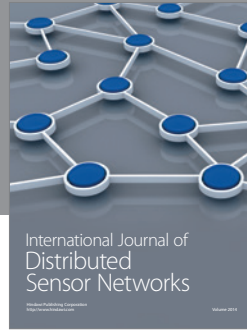
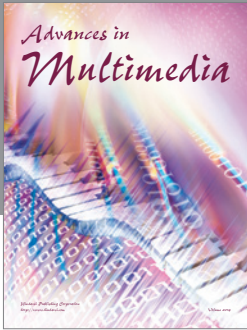
Acknowledgments

This work was supported by JSPS KAKENHI Grant nos. JP25240010 and JP16K16042.

References

- [1] R. Peters, R. Pak, G. Abowd, A. Fisk, and W. Rogers, "Finding lost objects: informing the design of ubiquitous computing services for the home," Tech. Rep. GIT-GVU-04-01, Georgia Institute of Technology, 2004.
- [2] S. Konishi, Y. Kawakita, and H. Ichikawa, "Method for estimation of distance between objects and its application for finding lost objects," in *Proceedings of the IEEE Consumer Communications and Networking Conference (CCNC '12)*, pp. 708–710, January 2012.
- [3] E. Colin, A. Moretto, and M. Hayoz, "Improving indoor localization within corridors by UHF active tags placement analysis," in *Proceedings of the IEEE RFID Technology and Applications Conference (RFID-TA '14)*, pp. 181–186, IEEE, Tampere, Finland, September 2014.
- [4] A. De Angelis, S. Dwivedi, P. Händel, A. Moschitta, and P. Carbone, "Ranging results using a UWB platform in an indoor environment," in *Proceedings of the International Conference on Localization and GNSS (ICL-GNSS '13)*, 5, 1 pages, June 2013.
- [5] F. Tlili, N. Hamdi, and A. Belghith, "Accurate 3D localization scheme based on active RFID tags for indoor environment," in *Proceedings of the IEEE International Conference on RFID-Technologies and Applications (RFID-TA '12)*, pp. 378–382, IEEE, Nice, France, November 2012.
- [6] M. Hasani, E.-S. Lohan, L. Sydänheimo, and L. Ukkonen, "Path-loss model of embroidered passive RFID tag on human body for indoor positioning applications," in *Proceedings of the IEEE RFID Technology and Applications Conference (RFID-TA '14)*, pp. 170–174, September 2014.
- [7] J. Y. Wang, C. P. Chen, T. S. Lin, C. L. Chuang, T. Y. Lai, and J. A. Jiang, "High-precision RSSI-based indoor localization using a transmission power adjustment strategy for wireless sensor networks," in *Proceedings of the 2012 IEEE 14th International Conference on High Performance Computing and Communication & 2012 IEEE 9th International Conference on Embedded Software and Systems (HPCC-ICCESS '12)*, pp. 1634–1638, Liverpool, UK, Jun 2012.
- [8] P. Kriz, F. Maly, and T. Kozel, "Improving indoor localization using bluetooth low energy beacons," *Mobile Information Systems*, vol. 2016, Article ID 2083094, 11 pages, 2016.
- [9] M. Muñoz-Organero and C. Brito-Pacheco, "Improving accuracy and simplifying training in fingerprinting-based indoor location algorithms at room level," *Mobile Information Systems*, vol. 2016, Article ID 2682869, 16 pages, 2016.
- [10] P. Bahl and V. Padmanabhan, "Radar: an in-building RF-based user location and tracking system," in *Proceedings of the 19th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '00)*, pp. 775–784, March 2000.
- [11] T. He, C. Huang, B. M. Blum, J. A. Stankovic, and T. Abdelzaher, "Range-free localization schemes for large scale sensor networks," in *Proceedings of the 9th Annual International Conference on Mobile Computing and Networking (MobiCom '03)*, pp. 81–95, ACM, San Diego, Calif, USA, September 2003.
- [12] 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.
- [13] M. Tanbo, R. Nojiri, Y. Kawakita, and H. Ichikawa, "Active RFID attached object clustering method based on RSSI series for finding lost objects," in *Proceedings of the 2nd IEEE World Forum on Internet of Things (WF-IoT '15)*, pp. 363–368, IEEE, Milan, Italy, December 2015.
- [14] S. Arya, D. M. Mount, N. S. Netanyahu, R. Silverman, and A. Y. Wu, "An optimal algorithm for approximate nearest neighbor searching fixed dimensions," *Journal of the ACM*, vol. 45, no. 6, pp. 891–923, 1998.
- [15] M. Datar, N. Immorlica, P. Indyk, and V. S. Mirrokni, "Locality-sensitive hashing scheme based on p-stable distributions," in *Proceedings of the 20th Annual Symposium on Computational Geometry (SCG '04)*, pp. 253–262, ACM, Brooklyn, NY, USA, 2004.
- [16] X. Wang, Y. Zhang, W. Zhang, X. Lin, and M. A. Cheema, "Optimal spatial dominance: an effective search of nearest neighbor candidates," in *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD '15)*, pp. 923–938, New York, NY, USA, June 2015.
- [17] S. P. Smith and R. Dubes, "Stability of a hierarchical clustering," *Pattern Recognition*, vol. 12, no. 3, pp. 177–187, 1980.
- [18] F. Rodrigues, J. Diniz-Filho, L. Bataus, and R. Bastos, "Hypothesis testing of genetic similarity based on RAPD data using mantel tests and model matrices," *Genetics and Molecular Biology*, vol. 25, no. 4, pp. 435–439, 2002.
- [19] A. Savvides, C.-C. Han, and M. B. Strivastava, "Dynamic fine-grained localization in ad-hoc networks of sensors," in *Proceedings of the 7th Annual International Conference on Mobile Computing and Networking (MobiCom '01)*, pp. 166–179, ACM, Rome, Italy, July 2001.
- [20] V. Erceg, L. J. Greenstein, S. Y. Tjandra et al., "Empirically based path loss model for wireless channels in suburban environments," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 7, pp. 1205–1211, 1999.

- [21] M. Nakagami, "Them-distribution-ageneralformula of intensity distribution of rapid fading," in *Statistical Methods in Radio Wave Propagation*, W. Hoffman, Ed., pp. 3–36, Pergamon, 1960.
- [22] S. O. Rice, "Mathematical analysis of random noise," *The Bell System Technical Journal*, vol. 23, no. 3, pp. 282–332, 1944.
- [23] Y. Ai, M. Cheffena, and Q. Li, "Radio frequency measurements and capacity analysis for industrial indoor environments," in *Proceedings of the 9th European Conference on Antennas and Propagation (EuCAP '15)*, April 2015.



Hindawi

Submit your manuscripts at
<https://www.hindawi.com>

