Location Determination Method for Wireless Systems Based on Learning Vector Quantization

Tomoaki Ogawa†*, Shuichi Yoshino, and Masashi Shimizu*

Abstract

In future ubiquitous services, object location determination will be a key technology. We describe a location determination method based on signal intensity using signals sent from base stations located at known positions to a terminal located at an unknown position. Instead of using numerical coordinates to define the object's location, it determines the object's location using a name that describes its locality. This method is based on learning vector quantization (LVQ) algorithms. We tested it using a 802.11 wireless LAN (local area network) and RFID (radio frequency identification) tag systems. The results show that it provides more accurate location data than the conventional least-squares method in areas far from the base stations.

1. Introduction

The spread of cellular phones and personal mobile terminals is creating an environment of ubiquitous wireless networks. In the future ubiquitous services, location determination will be a key technology. In the ubiquitous era, wireless devices will be attached to every object and all objects will be connected to the network. It will be possible to determine an object's location from data from these wireless devices. Therefore, applications using location determination systems have begun to attract a great deal of interest [1] and the importance of being able to correctly determine the location of an object or person has grown. In most cases, numerical coordinates are used to represent the location of objects or persons. In this paper, we suppose that an object or a person with a wireless interface represents an object, and that fixed base stations (BSs) can communicate with that object. Since the location determination techniques are to be used for various purposes and under various conditions, both outdoors and indoors, each technique must be well-matched to the system in which it is to be used. These techniques use multiple distance measurements between the object and BS (for example, a global positioning system (GPS) satellite or the BS of a cellular system). The most popular technique for outdoor location determination uses GPS and is commonly employed in car navigation systems [2]. This technique is based on the time of arrival (TOA) of a signal from the GPS satellite and calculates the distance between the user terminal and the satellite to find the coordinates of the user station. A network-assisted GPS technique has been proposed by SnapTrack, Inc. as a means of evaluating the accuracy of GPS-based detection [3]. The user station of this system communicates with a stationary server on the ground and revises the arrival time from the GPS. In this manner, the user station can obtain more accurate coordinates and the usage area can be expanded to include areas between tall buildings, or even indoors if the user is next to a window. The Cricket Location Support System is an indoor location system envisioned by MIT's Project Oxygen [1]. This system uses a combination of RF (radio frequency) and ultrasound technologies to provide a location-support service to users and applications using an RF signal operating in the 418-MHz amplitude modulation band. With each RF advertisement, the beacon transmits a concurrent ultrasonic pulse. Japanese PHS (1.9-GHz-band personal handy-phone system) carriers provide indoor location services to user stations by using the received intensity of signals from BSs on the premises [4]. Furthermore, techniques that have

† NTT Network Innovation Laboratories
Yokosuka-shi, 239-0847 Japan
E-mail: ogawa.tomoaki@lab.ntt.co.jp

been tuned for high accuracy have been reported. These include a positioning method that searches within a limited area for a user station assumed to be moving at walking speed [5] and one that calculates the user position by using only high-confidence data while ignoring low-confidence data from distant BSs [6]. TOA requires fast time resolution of less than a nanosecond to calculate the distance with 1-m accuracy indoors. We focused on improving the accuracy of indoor location determination based on the intensity of a signal received by an object from a BS, where the object cannot receive a GPS signal.

The conventional methods using signal intensity calculate the numerical coordinates of the object using an equation that predicts the distance based on the intensity of the signal from the BS received by the object. However, it is usually impossible to find the parameters of the equation for the correct propagation environment because of the presence of objects that shield or reflect radio waves.

In this paper, we propose a method that resolves this problem by combining the following two schemes. First, we propose a new paradigm for representing the location of an object. For this study, we must first define the difference between ‘positioning’ and ‘locating’. Our method is a ‘locating’ method and represents the location of the object by using the name or designation of the locality, and not by using the coordinates of either the object or BS like conventional ‘positioning’ methods. Second, we developed a new locating method that uses a learning function in conjunction with data obtained in advance. The conventional method that obtains data in advance was presented in [9]. However, this method must assign coordinate positions to all the obtained data. Therefore, gathering data was troublesome. Our method resolves this problem by combining the representation method using the location name with a location determination method employing a fuzzy learning strategy. A field test showed that it can determine the location of the object more accurately than conventional methods, such as the least-squares method (LSM), which is a positioning algorithm.

2. Location representations

2.1 Location representation method based on coordinates

We suppose that the object to be located is in a building and that the building has many BSs for wireless communications such as wireless LAN access points, PHS BSs, or RFID (radio frequency identification) tag receivers. We also assume that the object can link to a BS.

In most location representation methods, numerical coordinates are used to represent the location of the object. The coordinates are usually two-dimensional (horizontal plane), but may sometimes be three-dimensional (including height). In practice, the position of the object is calculated using a positioning algorithm and visually represented on a digital map. In response to requests from applications, the results are matched against a database that stores the names and coordinates of landmarks. When the object is a user’s terminal, this database information is then conveniently provided to the terminal, for example, “there is an entrance nearby”. Thus, it requires two steps for the user to receive the information he or she needs: one process for calculating the coordinates and another for matching the location to a name in the database.

For services utilizing the location determination capability, the location name (e.g., the building’s room number) is more useful to the user than the exact numerical coordinates. Moreover, even if the exact coordinates are obtained, this information is worthless if a mistake is made in converting them and you end up with the wrong room number.

2.2 Proposed representation method using locality name

We propose a new method for representing an object’s location, which uses the designation of the localized area (locality name). It treats the signal intensities from the BSs to the object as inputs and treats the locality name as the output. It applies the learning algorithm described in Section 3.2 to training data gathered in advance by, for example, walking around in the entire target area with a measuring terminal to obtain the signal intensities. The signal intensities are allocated as input. At the same time, the locality name is entered with a keyboard and allocated as output. Then the learning algorithm learns all the obtained input/output pairs in the entire target area directly. In the operation of the location determination system, the input data (signal intensities) are obtained and the filter generated by learning calculates the locality name. This representation method has two merits. One is that it does not need to convert the coordinates into a locality name, so it takes only one step to obtain the location. The other is that nobody needs to know the coordinates of the BSs. Namely the person who installs them does not need to input accurate coordinates for them. This method is suitable for navigation systems using small user terminals that have no screen and that provide only audio navigation instructions, because it outputs only location names.

3. Location determination algorithms

Here, we describe the conventional positioning algorithm and the proposed algorithm, assuming that these algorithms are used in a system comprising wireless BSs and user stations. We suppose that a BS broadcasts a signal such as a beacon periodically and a user station can measure the intensities of signals received from the BSs.

3.1 Conventional position determination algorithm using LSM

LSM is often used for position determination using the intensities of signals received from multiple wireless BSs. In this method, the distance from each BS to the user station is estimated by assuming that the attenuation of signal intensity is inversely proportional to the distance. Assuming that the user station is in a typical office space and uses a 2.4-GHz wireless interface, the function used to calculate the received intensity is expressed by Eq. (1) in logarithmic form (see ITU-R Recommendation P-1238-1 [7]), where the propagation constant is 3.0.

$$L_k = 40.9 + 30 \cdot \log(d_k), \qquad (1)$$

where L_k is the propagation loss, d_k is the propagation distance, and 40.9 is the coefficient.

Position determination works on the principle of triangulation. The position of the user station can be calculated using three or more distances on a two-dimensional plane. First, d_k is acquired using Eq. (1) and the known positions of the BSs (x_k, y_k). Then, LSM is applied. This is a practical method for deriving the optimum parameters; the minimum value of the square error is found by changing the parameters. In position determination, the position (X, Y) of the user station is given by the minimum value of Eq. (2).

$$F(X,Y) = \sum_{k=1}^{n} \left(\sqrt{(X - x_k)^2 + (Y - y_k)^2} - d_k \right)^2 \qquad (2)$$

The LSM method and similar methods that determine the user position by using distances calculated from received intensities have two problems.

1. They require information concerning the propagation environment in the target areas, but the correct propagation environment is usually impossible to ascertain from one parameter, because there are usually objects that shield or reflect radio waves in the area, especially indoors (**Fig. 1**).
2. The error in the distance estimated using Eq. (1) is large in areas far from the BSs, because the distance attenuation expressed in logarithmic form is small in such an area.

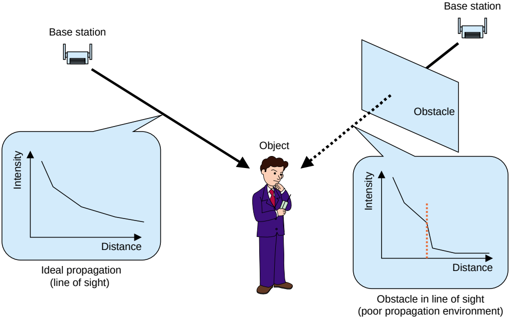


Fig. 1. Relationship between signal intensity and distance from base station in an office.

3.2 Proposed location determination method using a learning function

In this paper, for indoor location determination, we propose a fuzzy strategy using a learning function as a solution to the above problems. First, the whole area is divided into several small zones or localized areas (e.g., Room A and Room B), which are used to indicate the location of the user.

(1) Acquiring input and output data

The proposed method uses the following data sets as training data.

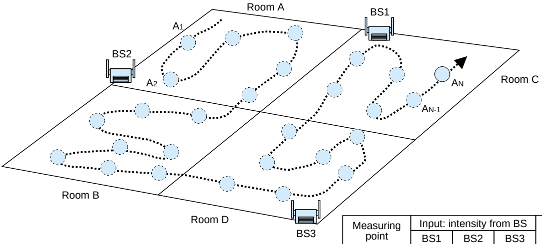
a) Input: Intensities of signals received from BSs at a particular position.
b) Output: Designation of the localized area for a particular set of input data.

The input/output pairs are obtained as training data for the entire target area. The proposed method learns these pairs and makes a rule for deciding the output for new input data. The rule gives the designation of the localized area as output data. **Figure 2** shows an example of obtaining data sets for learning in the four localized areas (Rooms A–D), including three BSs. The data for N measuring points are obtained, and a room designation is assigned to each point's data as output.

(2) Learning process using LVQ algorithm

The proposed method uses the LVQ algorithm to learn the above training data sets. LVQ is a supervised learning algorithm for neural networks and it is used commonly in the field of image analysis. In this algorithm, a set of input elements is considered to be a vector. When S BSs and N measuring points are used, N input vectors are plotted in the feature space, and the value of each vector element is used as an axis of S-dimensional feature space. Reference vectors are plotted in the feature space for quantizing all measuring points. **Figure 3** shows an example of plotting the input vectors and the reference vectors in a 3-dimensional feature space. Each reference vector represents a room in the feature space, and the locations of the reference vectors are moved within the feature space according to the learning rule. This rule compares the input vectors with reference vectors and moves the reference vectors to more suitable positions to represent the rooms. **Figure 4** shows an example of the reference vector activity using this learning rule. It can be proved that LVQ can establish optimal borders for logically discriminating each localized area (rooms in this case), where these lines are called Bayes discrimination [8]. The flow diagram in **Fig. 5** describes the LVQ learning process in the proposed method.

After the process of learning, the generated rule is used to determine the location. When the user station receives new signals from BSs, the received signal intensities are used to construct the input vector in the feature space. This new input vector searches for the reference vector that minimizes the square norm between the input vector and the reference vector. Then, the output of the reference vector, which is the locality name (e.g., room name), is linked to the new



Measuring point	Input: intensity from BS			Output: room designation
	BS1	BS2	BS3	
A_1	50	84	10	Room A
A_2	46	92	14	Room A
......	......	......	......	......
A_{N-1}	85	40	54	Room C
A_N	88	38	48	Room C

Fig. 2. Example of obtaining data sets for learning.

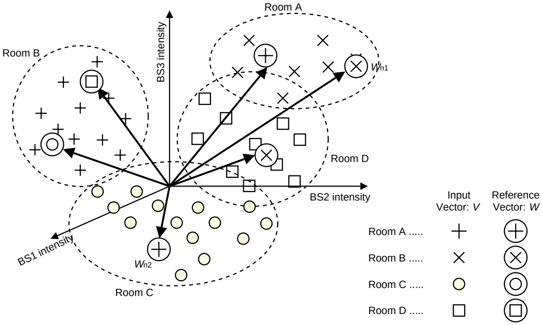


Fig. 3. Example of plotting input vectors and reference vectors in 3-dimensional feature space.

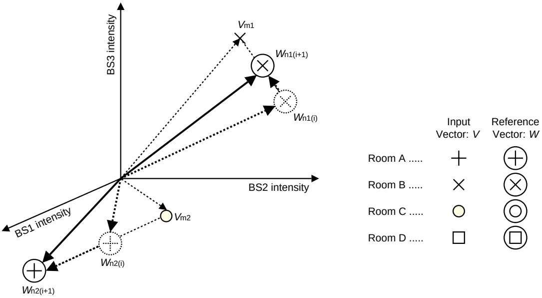


Fig. 4. Example of reference vector activity using the proposed learning rule.

set of received intensities.

In the conventional non-linear optimizing method, it is necessary to measure numerical coordinates of positions throughout the entire area using a pulse counter or optical camera in advance. In contrast, in the proposed method, the BS coordinates do not need to be recorded (**Fig. 6**). Furthermore, information about the propagation environment is automatically projected in the information of the reference vectors.

This algorithm can also be used for other architectures. For example, in an active RFID tag system, the user (or object) has a tag that broadcasts its ID code, and receivers that can measure the intensity from the tag are established as base points. The server collects the data and calculates the location of the tag. Experimental results using active RFID tags are given in [10].

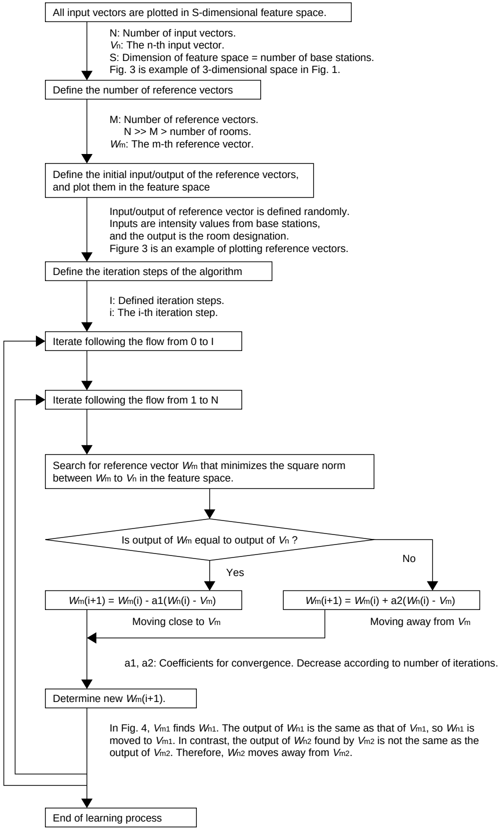


Fig. 5. Flow of learning process by LVQ.

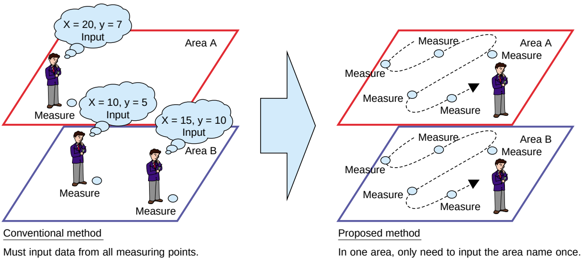


Fig. 6. Difference between conventional and proposed methods of gathering data.

4. Experiment

4.1 Obtaining input (signal intensity) and output (location name) data

We obtained experimental data in an actual office, and compared the locations calculated using the proposed method against those obtained using the conventional LSM method. **Table 1** shows the conditions for obtaining the data, and **Fig. 7** shows the layout of the office.

A frequency hopping wireless LAN conforming to IEEE 802.11 was used as the wireless interface. Signals with the station ID were broadcast from each BS, and the received signal intensities were obtained at all measuring points. A time period averaged over 0.5 to 20 s was used for the intensities. The room was divided into five zones (areas A to E in Fig. 7). We obtained 1000 data sets per measuring point from 300 measurement points. The data were obtained over several days because there were fluctuations in the signal intensity.

Table 1. Experimental conditions.

Room conditions	25.2 × 13.5 m
Number of base stations	5
Signal from base station	Broadcast signal including ID of base station
Signal interval	200 ms (average)
Wireless interface	IEEE 802.11 frequency hopping wireless LAN
Number of localized areas	5
Number of measuring points	300
Measuring point interval	1 m
Averaging time for measuring data	0.5 s to 20 s
Test data	1000 per measuring point

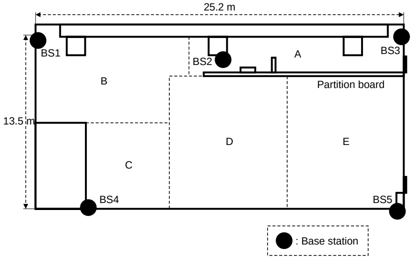


Fig. 7. Layout of office used in the experiment.

4.2 Learning process specifications

Table 2 gives the specifications of the learning process of the proposed method. The training data for learning contained not only the intensities, but also the actual localized areas. We obtained 10 training data sets at 300 measuring points for a total of 3000

Table 2. Specifications of proposed learning process.

Training data sets	3000
Reference vectors	40
Iteration steps	100,000

points. The number of reference vectors was 40, and the reference vectors scattered in the feature space filled in 3000 input vectors of the training data. The learning involved 100,000 iteration steps.

4.3 Conditions for the conventional method

The conditions of the LSM method were as follows. The propagation constant of Eq. (1) was 1.5. This coefficient was defined using the received intensity measured in advance in the experimental room. We searched for F(X, Y) in Eq. (2) at intervals of 0.5 m throughout the experimental room. For comparison with the proposed method, the coordinate positions of the LSM results were changed to the designations of the localized areas.

5. Experimental results

We compared the results for the proposed method with those for the conventional method using the success rate to determine whether or not the methods could correctly indicate the correct localized area. The answers provided by the LSM method were converted from coordinates into locality names (areas A to E in Fig. 7) and judged.

Figure 8 shows the success rate versus the averaging time of the received intensities for a wireless LAN. For both methods, as time elapsed, the success rate increased until the averaging time was longer than 5 s. This indicates that it takes 5 s to obtain the correct answer with either method. The success rate of the proposed method was higher than that for the conventional method for all averaging times. The upper limit of the success rate for the proposed method was approximately 70%, whereas the upper limit of the conventional method was only approximately 60%.

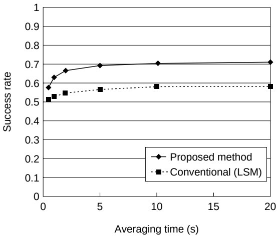


Fig. 8. Success rate versus averaging time of received intensity.

The distributions of the success rate for the experimental room are shown for the conventional and proposed methods in **Figs. 9** and **10**, respectively. The areas where the success rate was lower than 70% are shown in red. In Areas A, B, C, and E, the success rates of both methods were almost equal. On the other hand, in Area D, there was a large difference between them. In the whole of Area D, the success rate of the

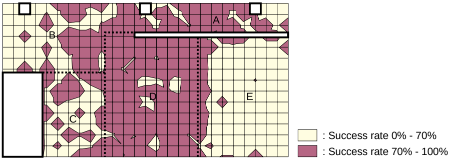


Fig. 9. Distribution of success rate using conventional method in experimental room.

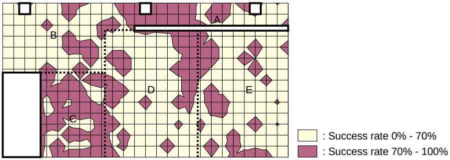


Fig. 10. Distribution of success rate using proposed method in experimental room.

proposed method was higher than 70%, whereas that of the conventional method was lower than 70%. This is because, in areas that contain BSs, the dispersions of the received intensity are small and the intensity attenuation with distance is large, so the propagation environment can be easily estimated. Thus, the conventional method can easily detect a user station near a BS. However, in areas that do not include a BS and in ones that are far from a BS, such as Area D, Eq. (1) in the conventional method does not accurately model the actual propagation behavior, so the success rate is lower. Our method can usually detect the user correctly in Area D, because it characterizes Area D using the reference vectors as "an area without a BS" in the feature space. **Figure 11** shows the success rate versus the averaging time of the received intensities in Area D. We can see that the upper limits are 75% for the proposed method and 25% for the conventional method.

We obtained similar results using active RFID tags. **Figures 12** and **13** show the distribution of the success rate for the conventional and proposed methods using RFID tags. The experimental conditions were as follows: a 12×9 m^2 room was logically divided into four zones, five tag receivers were attached to the ceiling, and the broadcasting interval for tags was 2 s. The results show that the success rate of the proposed method was approximately 90% and that of the conventional method was approximately 40%. The conventional method used a propagation constant in Areas C and D, but the constant was not suitable for Area A or B, so the success rate was low in these areas. On the other hand, the success rate of the proposed method was high for all areas, because the reference vectors expressed the characteristics of the propagation environment. The attenuation of the signal intensity with distance when using the RFID tags was greater than that using the 802.11 wireless LAN, and there was little signal reflection from RFID tags near

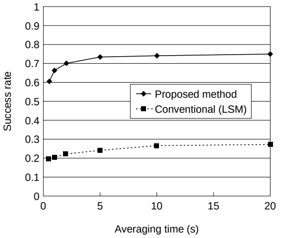


Fig. 11. Success rate versus averaging time of received intensities in Area D.

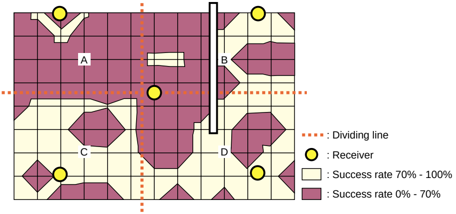


Fig. 12. Distribution of success rate using conventional method with RFID tags.

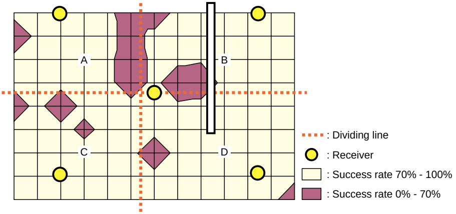


Fig. 13. Distribution of success rate using proposed method with RFID tags.

the walls. Therefore, the success rate of the proposed method using RFID tags was higher than that using the 802.11 wireless LAN.

6. Conclusion

We described a location determination method based on Learning Vector Quantization using the signal intensities from base stations. The method employs a unique concept combining the following two schemes: 1) a location representation method using the locality name and 2) a learning function based on Learning Vector Quantization algorithms. An experiment conducted in an office showed the superior success rate of this method, 70% for a wireless LAN and 90% for wireless active tags, compared with 60% and 40%, respectively, for the conventional method. This new location determination method located the target object more accurately than the conventional method throughout the office. It was especially better in areas that had no base station. This method is effective for indoor human navigation or object detection systems. In the future, we plan to improve the method of obtaining data and investigate user-friendly location determination systems.

References

[1] J. Hightower and G. Borriello, "A Survey and Taxonomy of Location Systems for Ubiquitous Computing," Technical Report UW-CSE 01-08-03, pp. 57-66, IEEE, Aug. 2001.

[2] E. D. Kaplan, "Understanding GPS Principles and Applications," ARETECH HOUSE, 1996.

[3] SnapTrack, Inc., "An Introduction to SnapTrack Server-Aided GPS Technology," http://www.snaptrack.com/, 1999.

[4] H. Koshima and J. Hoshen, "Personal Locator Services Emerge," IEEE Spectrum, 37(2), pp. 41-48, Feb. 2000.

[5] A. Shibuya, M. Nakatsugawa, and S. Kubota, "High Accuracy Detection Method on the Positioning Information System, P-POINTS," Proc. of the Coms. Soc. Conf., IEICE, Chiba, Japan, A-17-4, Sep. 1999 (in Japanese).

[6] A. Shibuya, D. Uchida, M. Nakatsugawa, E. Kudoh, and S. Kubota, "Far-marker Exclusion Algorithm for the Local Information and Positioning System," Proc. of the Coms. Soc. Conf., IEICE, Nagoya, Japan, B-5-54, 2000 (in Japanese).

[7] Rec. ITU-R P.1238-1, "Propagation data and prediction methods for the planning of indoor radio communication systems and radio local area networks in the frequency range 900 MHz to 100 GHz," 1997-1999.

[8] T. Kohonen, G. Barna, and R. Chrisley, "Statistical pattern recognition with neural networks: benchmarking studies," Proc. ICNN, Tokyo, Japan, Vol. I, pp. 62-68, Aug. 1988.

[9] P. Bahl and V. N. Padmanabhan, "RADAR: An In-Building RF-Based User Location and Tracking System," Proceedings of IEEE Infocom 2000, Tel-Aviv, Israel, Mar. 2000.

[10] T. Ogawa, S. Yoshino, and M. Shimizu, "Estimation of the Location Detection Method Adopting Learning Algorithms," Proc. of the Coms. Soc. Conf., IEICE, Niigata, Japan, SB-4-7, Sep. 2003 (in Japanese).

Tomoaki Ogawa
Engineer, Wireless Systems Innovation Laboratory, NTT Network Innovation Laboratories.

He received the B.A. degree in environmental information and Master degree in media and Governance from Keio University, Kanagawa in 1996 and 1998, respectively. After working at NTT Wireless Systems Laboratories from 1998 to 1999, he joined NTT Network Innovation Laboratories where he has been working on wireless *ad-hoc* networking and local positioning systems. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE).

Shuichi Yoshino
Senior Research Engineer, Wireless Systems Innovation Laboratory, NTT Network Innovation Laboratories.

He received the B.E. and M.E. degrees in mechanical engineering from Kanazawa University, Ishikawa, in 1990 and 1992, respectively. He joined NTT Laboratories in 1992 and worked on satellite Internet system development. He is currently engaged in R&D on wireless networking technology for ubiquitous services.

Masashi Shimizu
Senior Research Engineer, Supervisor, Wireless Systems Innovation Laboratory, NTT Network Innovation Laboratories.

He received the B.E. and M.E. degrees in mechanical engineering from Keio University, Yokohama, in 1986 and 1988, respectively. In 1988, he joined NTT Wireless Systems Laboratories, Yokosuka, Japan. Since then, he has been engaged in research on the pointing control for deployable space antennas and surface error compensation through feed distributions control. His recent interest focuses on active RFID and its applications. He is a member of IEICE.