3D Ultrasonic Tagging System for Observing Human Activity

Yoshifumi Nishida
Digital Human Research Center
AIST, & CREST, JST
Aomi Koto-ku, Tokyo
135-0064 Japan

Hiroshi Aizawa
Dept. of Electronic Engneering
Tokyo Denki Univ.
Nishiki-cho Chiyoda-ku, Tokyo
101-8457 Japan

Toshio Hori
Digital Human Research Center
AIST, & CREST, JST
Aomi Koto-ku, Tokyo
135-0064 Japan

Nell H. Hoffman
School of Computer Science
Carnegie Mellon University
5000 Forbes Avenue Pittsburgh, PA
15213 USA

Takeo Kanade
Digital Human Research Center
AIST, & CREST, JST
Aomi Koto-ku, Tokyo
135-0064 Japan

Masayoshi Kakikura
Dept. of Electronic Engneering
Tokyo Denki Univ.
Nishiki-cho Chiyoda-ku, Tokyo
101-8457 Japan

***Abstract*— This paper describes an ultrasonic tagging system developed for robustly observing human activity in a living area. Using ultrasonic transmitter tags with unique identifiers, the system is shown through experimental application to be able to track the three-dimensional motion of tagged objects in real time with high accuracy, resolution and robustness to occlusion. The use of an ultrasonic system is desirable because of its low cost and use of commercial components, and the proposed system achieves high accuracy and robustness through the use of many redundant sensors. The system employs multilateration to locate tagged objects using one of two estimation algorithms, a least-squares optimization method or a random sample consensus method.**

I. Introduction

Information processing services centered around human activity in the real world has attracted increased attention recently [1]. Human-centered applications require the facility to observe and recognize activities as a basis, and the present paper describes a method for observing and recognizing behaviors robustly and in real time based on sensorizing objects in the real world.

Generally, the problem of human behavior recognition can be formulated as a kind of pattern recognition problem as follows.

$$P(\hat{W}|Y) = \max_{W_i} \frac{P(Y|W_i)P(W_i)}{P(Y)}, \tag{1}$$

where $P(W_i|Y)$ denotes the posterior probability that the meaning of an observed behavior pattern Y is W_i, $P(Y)$ denotes the probability that a behavior pattern Y will be observed, $P(W_i)$ denotes the probability that the behavior meaning W_i will occur, and $P(Y|W_i)$ denotes the conditional probability. Thus, the problem of human behavior recognition becomes that of searching for the maximum posterior probability $P(\hat{W}|Y)$.

Two problems complicate the recognition of human behavior: the ability to observe a behavior pattern Y robustly, and the efficient recognition of meaning W from the observed pattern. Without solving the first problem, equation (1) cannot be formed. Without tackling the second problem, guaranteeing a solution to the equation within the time frame demanded by the application is impossible.

As a method for efficient recognition of behaviors, the idea of object-based behavior recognition has been proposed [2]. In theory, the behavior of handling objects in an environment such as an office or home can be recognized based on the motion of the objects. However, when applying the method to real environments, it is difficult to even achieve an adequate level of object recognition, which is the basis of the method.

Separating the problems of object recognition and behavior recognition is becoming increasingly realistic with the progress in microcomputers, sensor, and wireless networks technology. It has now become possible to resolve object recognition into the problems of sensorizing objects and tagging the objects with identification codes (IDs), and to address behavior recognition separately through the development of applied technology.

The present authors have developed a three-dimensional ultrasonic location and tagging system for the fundamental function of robustly tracking objects. This system enables a new approach of tag-based behavior recognition. In terms of cost and robustness against environmental noise, the ultrasonic system is superior to other location techniques such as visual, tactile, and magnetic systems. A number of ultrasonic location systems have already been proposed or commercialized [3], [4], [5]. However, the work [3] does not describe a method for improving the robustness, accuracy, and resolution of position, and although Shih [4] proposed a robust estimation method by "direct substitution", the system had difficulty in maintaining accuracy of position and calculation in real time.

The system presented in the present paper is developed specifically to address the issue of robustness and accuracy in real time.

The ultrasonic location system calculates the three-dimensional (3D) position of an object by trilateration based on three distance measurements. Like other location sensing systems such as motion capture, the system requires more than a certain minimum number of receivers to eliminate the effect of occlusion and outliers. The system is comparably inexpensive due to the availability of cheap ultrasonic receivers, which also makes it possible to increase the number of ultrasonic receivers to mitigate undesirable effects. An ultrasonic location system therefore provides significant advantages in terms of robust positioning, high accuracy, and high resolution through the collection of redundant distance data.

This research focuses on the development of a function for estimating the 3D position of objects with high accuracy, high resolution and robustness to occlusion through the use of redundant distance data. This paper describes the 3D position estimation function and the results of experiments conducted in a regular room area. In the next secton, the developed 3D ultrasonic tagging system is first introduced briefly. Section III describes the algorithms for estimating the 3D position of objects in detail. Trilateration or multilateration algorithms have been proposed in the field of aerospace[6], [7]. This paper presents the multilateration algorithms applicable to a more general case that multiple ultrasonic receivers are put on arbitrary positions. The results of experimental application of the system are then presented and discussed.

II. Ultrasonic tagging system

A. System configuration

Figure 1 shows the system configuration for the 3D ultrasonic tagging system. The system consists of an ultrasonic receiving section, an ultrasonic transmitting section, a time-of-flight measuring section, a network section, and a personal computer. The ultrasonic receiving section receives ultrasonic pulses emitted from the ultrasonic transmitter and amplifies the received signal. The time-of-flight measuring section records the travel time of the signal from transmission to reception. The network section synchronizes the system and collects time-of-flight data from the ultrasonic receiving section. The positions of objects are calculated based on more than three time-of-flight results. The sampling frequency of the proposed system is 50 Hz.

Figure 2 shows a photograph of the prototype network and time-of-flight measurement components, which are to be attached to a wall. Figure 3 is a photograph of the tagging unit (transmission unit), which consists of an ultrasonic transmitter, a wireless communication unit, a microcomputer (FLASH PIC) and power (two alkaline AA batteries).

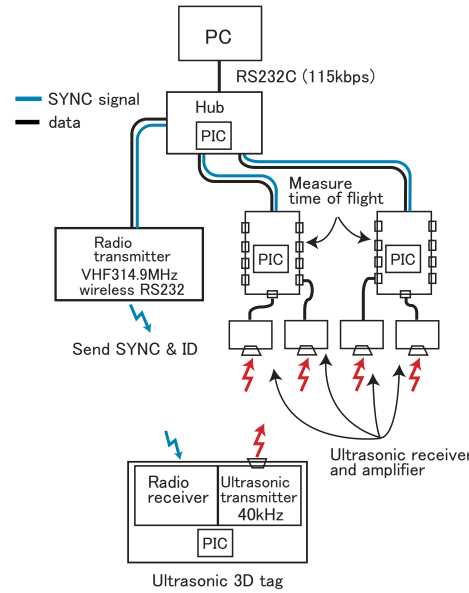

Fig. 1. Configuration of 3D ultrasonic tagging system

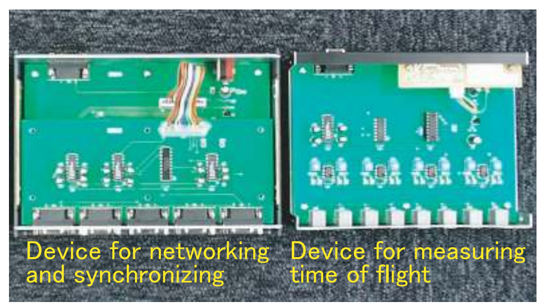

Fig. 2. Network and time-of-flight measurement components

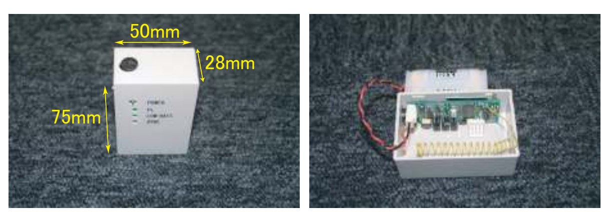

Fig. 3. Ultrasonic tag

The room used to conduct the experiments is shown in Fig. 4. The room was $3.5 \times 3.5 \times 2.7$ m in size, and was fitted with 307 ultrasonic receivers embedded in the wall and ceiling. Tags were attached to various objects, including a cup and a stapler as shown in Fig. 5. Some objects were fitted with two transmitters. The purpose of the experimental room is to clarify the effect of the use of redundant sensors. More than 300 receivers do not mean that the algorithms described in the next section need such a large number of sensors. In actual usage, a smaller number of receivers can be used.

Fig. 4. Room with embedded ultrasonic sensors for prototype development

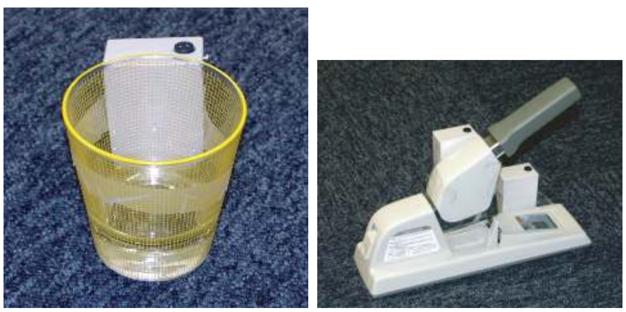

Fig. 5. Example of attaching tags to objects

III. Use of redundant ultrasonic receivers

A. Trilateration

The ultrasonic tagging system calculates the 3D position of an object by trilateration using three distance measurements. Two methods of trilateration are investigated for use with the proposed system: multilateration based on a least-squares method using redundant distance data, and multilateration based on robust estimation.

The basic principle of triangulation can be described by

$$(x_i - x)^2 + (y_i - y)^2 + (z_i - z)^2 = l_i^2, \quad (i = 1,2,3). \qquad (2)$$

where l_i denotes the distance measured by the ith ultrasonic receiver at position (x_i, y_i, z_i) from the ultrasonic transmitter at (x, y, z), as shown in Fig. 2. Thus, the position (x, y, z) of an ultrasonic transmitter can be calculated given three distance measurements $l_i (i = 1,2,3)$ obtained by three receivers that do not lie on the same line.

B. Multilateration: Basics

The estimation error ε_i can be defined by

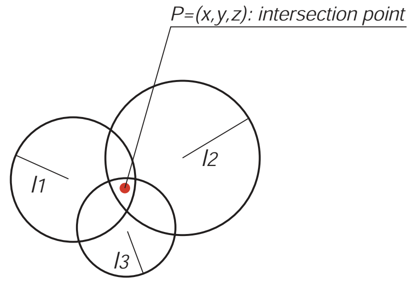

Fig. 6. Intersection point

$$\varepsilon_i = \left| l_i - \sqrt{(x_i - x)^2 + (y_i - y)^2 + (z_i - z)^2} \right|. \qquad (3)$$

By solving the minimization problem

$$(\hat{x}, \hat{y}, \hat{z}) \quad = \quad \min_{(x,y,z)} \sum_i^n \varepsilon_i, \qquad (4)$$

we can estimate the optimal value $(\hat{x}, \hat{y}, \hat{z})$.

The minimization problem of Eq. (4) involves the solution of a non-linear equation and therefore requires repetitive numerical computation. Shih [4] proposed a direct substitution method to solve Eq. (4) that involved substituting random and arbitrary (x, y, z) into Eq. (4) and adopting the coordinate giving the minimum error as the optimal value. This method is a robust estimation (M-estimator), but involves large calculation cost to guarantee the accuracy of the estimated position. For ultimate accuracy, it would therefore be necessary to evaluate all possible coordinates (x, y, z). However, such calculation is not suitable for real-time application.

C. Multilateration method 1: linearization of the minimization problem

To obtain an algorithm suitable for accurate estimation in real time, Eq. (4) may be linearized to allow an analytical solution.

Using distance data l_i, l_j and the receiver positions $(x_i, y_i, z_i), (x_j, y_j, z_j)$, we obtain the following spherical equations for the possible position of the target.

$$(x_i - x)^2 + (y_i - y)^2 + (z_i - z)^2 \quad = \quad l_i^2, \qquad (5)$$
$$(x_j - x)^2 + (y_j - y)^2 + (z_j - z)^2 \quad = \quad l_j^2. \qquad (6)$$

By subtracting Eq. (6) from Eq. (5), we obtain an equation for intersecting planes between the spheres, as shown in Fig. 7.

$$2(x_j - x_i)x + 2(y_j - y_i)y + 2(z_j - z_i)y = \\ l_i^2 - l_j^2 - x_i^2 - y_i^2 - z_i^2 + x_j^2 + y_j^2 + z_j^2 \qquad (7)$$

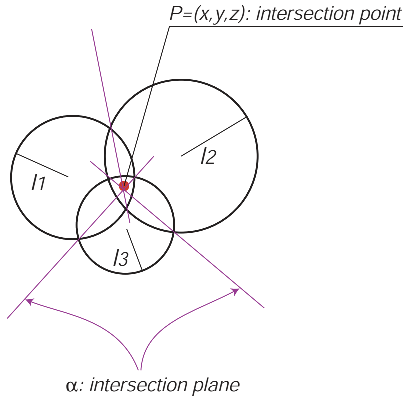

Fig. 7. Planes of intersection between spheres used to give the estimated position

By inputting pairs of (i, j) into the above equation, we obtain simultaneous linear equations, as expressed by

$$AP = B, \tag{8}$$

$$\text{where} \quad P = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \tag{9}$$

$$A = \begin{pmatrix} 2(x_0 - x_1) & 2(y_0 - y_1) & 2(z_0 - z_1) \\ 2(x_0 - x_2) & 2(y_0 - y_2) & 2(z_0 - z_2) \\ 2(x_0 - x_3) & 2(y_0 - y_3) & 2(z_0 - z_3) \end{pmatrix}, \tag{10}$$

$$B = \begin{pmatrix} l_1^2 - l_0^2 - x_1^2 - y_1^2 - z_1^2 + x_0^2 + y_0^2 + z_0^2 \\ l_2^2 - l_0^2 - x_2^2 - y_2^2 - z_2^2 + x_0^2 + y_0^2 + z_0^2 \\ l_3^2 - l_0^2 - x_3^2 - y_3^2 - z_3^2 + x_0^2 + y_0^2 + z_0^2 \\ \vdots \end{pmatrix}. \tag{11}$$

The position $(\hat{x}, \hat{y}, \hat{z})$ can then be calculated by a least-squares method as follows.

$$P = (A^T A)^{-1} A^T B. \tag{12}$$

This method minimizes the square of the distance between the planes expressed by Eq. (7) and the estimated position. The algorithm is described in detail in Fig. 8. In actual usage, the rank of matrix A must be considered.

D. Multilateration method 2: Robust estimation by RANSAC

Data sampled by the ultrasonic tagging system is easily contaminated by outliers due to reflections. Method 1 above is unable to estimate the 3D position with high accuracy if sampled data includes outliers deviating from a normal distribution. In the field of computer vision, robust estimation methods that are effective for sampled data including outliers have already been developed. In this work, the random sample consensus (RANSAC) [8], [9] estimator is adopted to eliminate the undesirable effects of outliers. The procedure is as follows.

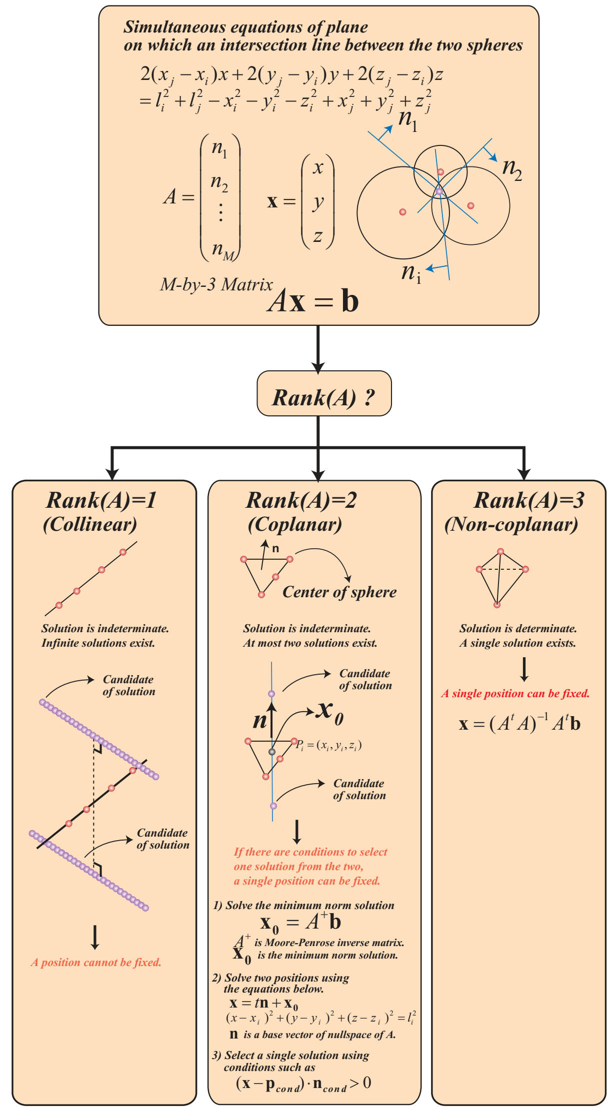

Fig. 8. Algorithm for estimating 3D position by a least-squares method considering the rank of **A**

1) Randomly select three distances measured by three receivers (jth trial).
2) Calculate the position (x_{cj}, y_{cj}, z_{cj}) by trilateration using Eq. (2).
3) Calculate the error ε_{cji} for all receivers ($i = 0, 1, \ldots, n$) by Eq. (13), and find the median ε_{mj} of ε_{cji}.
4) Repeat steps 1 to 3 as necessary to find the combination of measurements giving the minimum error, and adopt the corresponding 3D position.

$$\varepsilon_{cji} = \left| l_i - \sqrt{(x_i - x_{mj})^2 + (y_i - y_{mj})^2 + (z_i - z_{mj})^2} \right| \tag{13}$$

$$\varepsilon_{mj} = med_j \, |\varepsilon_{cji}| \tag{14}$$

$$(\hat{x}, \hat{y}, \hat{z}) = min \; \varepsilon_{mj} \tag{15}$$

IV. Experimental application

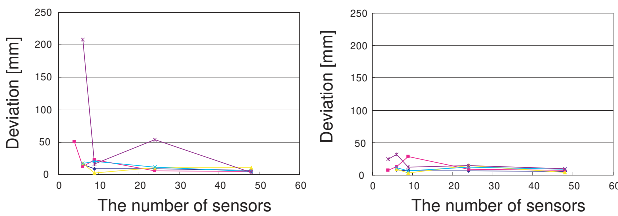

Fig. 9. Relationship between resolution and the number of sensors for the least-squares method (left) and RANSAC (right)

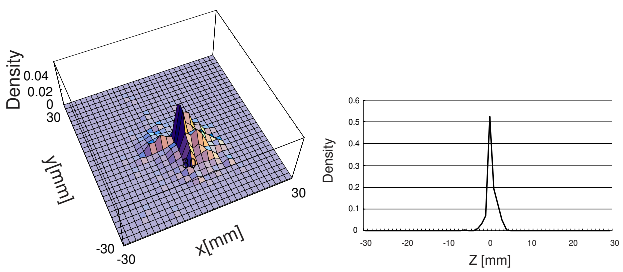

Fig. 10. Resolution in the x and y directions (left) and z direction (right) (grid size: 2×2 mm)

A. Resolution

Figure 9 shows the relationship between the number of receivers and the deviation of the estimated position for 4, 6, 9, 24, and 48 receivers in the ceiling. To compare the effect of the RANSAC method and that of the least-squares method, one receiver is selected randomly and 500[mm] is added to the distance data of the selected receiver as outlier. Each point was derived from 30 estimations of the position. The 5 lines in the figures represent estimation for 5 different locations of the transmitter. The resolution increases with the number of receivers, and the RANSAC method provides a more stable estimation with higher resolution compared to the least-squares method.

The resolution in the x, y, and z directions is illustrated in Fig. 10, which shows the probability density distribution for 1000 estimations using RANSAC. The resolution in x and y directions is about 15 mm, while that in the z direction is about 5 mm.

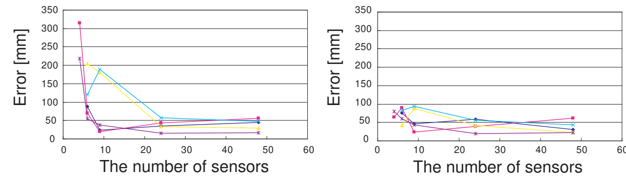

Fig. 11. Relationship between positioning accuracy and the number of receivers for the least-squares method (left) and RANSAC (right)

B. Positioning accuracy

Figure 11 shows the relationship between the number of receivers and the error of the estimated position for 4, 6, 9, 24, and 48 receivers. The error is taken as the distance from the position measured by a visual motion capture system. One receiver is selected randomly and 500[mm] is added to the distance data of the selected receiver as outlier. Each point was derived from 30 estimations of the position. The 5 lines in the figures represent estimation for 5 different locations of the transmitter. The error decreases as the number of receivers is increased, and the RANSAC method is appreciably more accurate with fewer receivers. It is considered that the least-squares method is easily affected by outliers, whereas the RANSAC method is not.

Figure 12 shows the 3D distribution of error for 1400 measured positions in the room. The figures show that the error is lowest (20–80 mm) immediately below the 48 receivers in the ceiling, increasing toward the edges of the room.

The results of experiments for evaluating accuracy and resolution demonstrate that it is possible to improve accuracy and resolution by increasing the number of receivers, and that the undesirable effect of outliers can be mitigated through the use of RANSAC estimation.

C. Robustness to occlusion

As in other measuring techniques such as vision-based methods, it is necessary to increase the number of sensors to solve the problem of sensor occlusion, where the line of sight to the target object is obstructed by other objects such as walls or room occupants. In the present tagging system, the problem of occlusion occurs often when a person moves or operates an object. These situations give rise to two separate problems; a decrease in the number of usable sensors for the target, and an increase in reflections due to obstruction and movement. As one of the most typical situations where occlusion occurs, this section focuses on occlusion due to a hand.

Figure 13 shows how the error increases and the number of usable sensor decreases as a hand approaches an object fitted with an ultrasonic transmitter for the least-squares and RANSAC methods. Although the error increases significantly by both methods when the hand approaches the object, the RANSAC method is much less affected

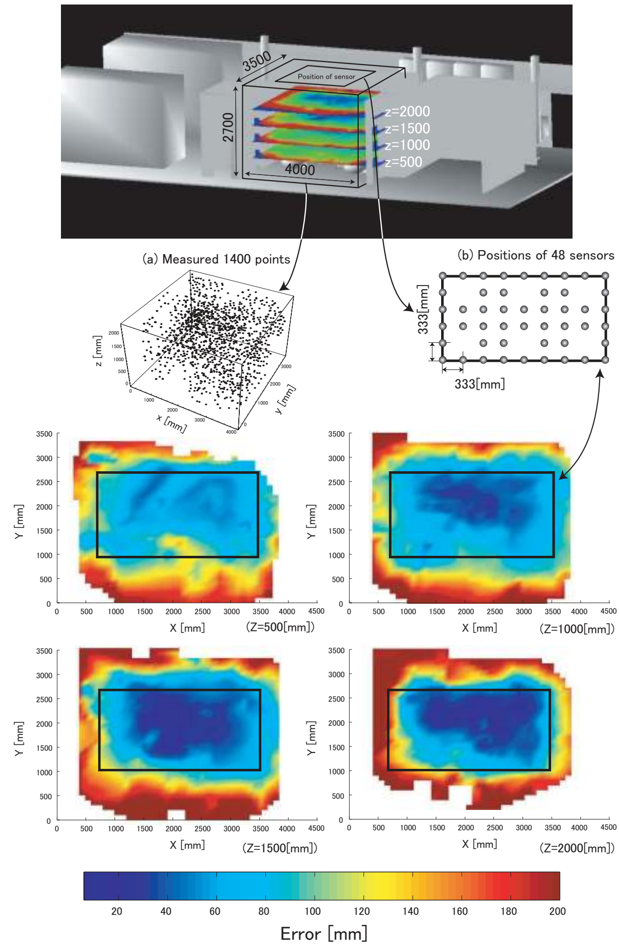

Fig. 12. 3D distribution of error in the experimental room

than the least-squares method. This demonstrates that the proportion of outliers increases when occlusion occurs, and that RANSAC is more robust in this situation because it can mitigate the effect of such outliers.

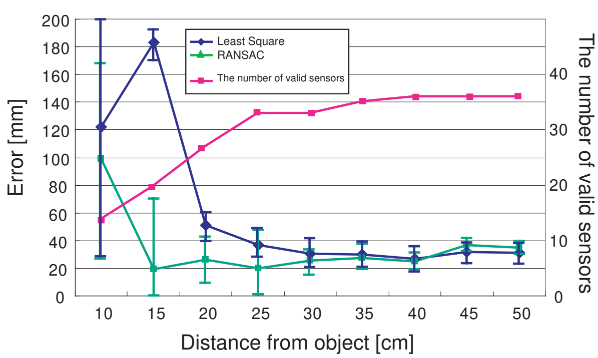

Fig. 13. Accuracy of the ultrasonic tagging system when occlusion due to a hand occurs

D. Real-time position measurement

Figure 14 shows the measured trajectory for a person moving a cup to a chair, the floor, and a desk. The figure demonstrates that the system can robustly measure the positions of the objects in most places of the room regardless of occlusion by a hand or body.

In the current system, the sampling frequency is about 50 Hz. This frequency decreases to $50/n$ Hz when n objects are being monitored. However, it is possible to maintain a high sampling frequency by selecting which transmitters to track dynamically. For example, a transmitter can be attached to a person's wrist, and the system can select transmitters in the vicinity of the wrist to be tracked, thereby reducing the number of transmitters that need to be tracked at one time and maintaining the highest sampling frequency possible. Figure 15 shows the measured trajectory in a dynamic selection mode. The red sphere in the figure shows the position of the hand.

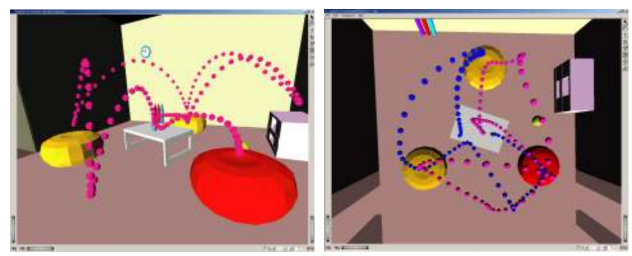

Fig. 14. Measured trajectory for moving a cup around the room

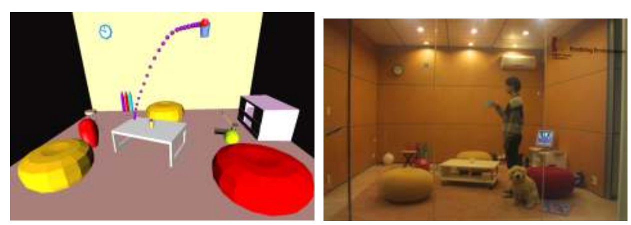

Fig. 15. Dynamic selection of transmitters

E. Recognition of human behavior

Figure 16 shows the measured trajectories when several objects are moved one after another (see video). Behavior recognition is performed by interpreting the change of state using the ultrasonic tag.

Output example .....,
04:03:55 place yellow cup on desk
04:04:05 hold mobile phone
04:04:12 place mobile phone on floor
04:04:19 hold chair
04:04:31 place chair on floor
04:04:34 hold trash
04:04:40 place trash on floor
04:04:46 hold stapler

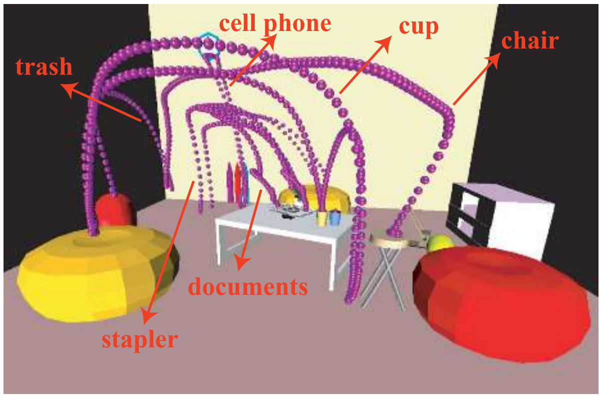

Fig. 16. Trajectories for movement of several objects one after another

04:04:52 place stapler on desk
04:04:59 hold documents
04:05:13 staple documents with stapler .....

V. Conclusion

A 3D ultrasonic tagging system that provides robust observation of human activity was presented. The ultrasonic tagging system consists of an ultrasonic transmitter/receiver, a wireless communication unit and a host computer, and can implemented at low cost. The system measures the 3D position of any object fitted with an ultrasonic transmitter with a unique ID.

In order to estimate the 3D position with high accuracy, high resolution, and robustness to occlusion, the authors propose two estimation methods, one based on a least-squares approach and one based on RANSAC.

The system was tested in an experimental room fitted with 307 ultrasonic receivers; 209 in the walls and 98 in the ceiling. The results of experiments conducted using 48 receivers in the ceiling for a room with dimensions of $3.5 \times 3.5 \times 2.7$ m show that it is possible to improve the accuracy, resolution, and robustness to occlusion by increasing the number of ultrasonic receivers and adopting a robust estimator such as RANSAC to estimate the 3D position based on redundant distance data. The resolution of the system is 15 mm horizontally and 5 mm vertically using sensors in the ceiling, and the total spatially varying position error is 20–80 mm. It was also confirmed that the system can track moving objects in real time, regardless of obstructions.

Further development of the system will include refinement of the method for measuring the 3D position with higher accuracy and resolution, miniaturization of the ultrasonic transmitters, development of a systematic method for defining and recognizing human activities based on the tagging data and data from other systems, and development of new applications based on human activity data.

VI. References

[1] T. Hori, "Overview of Digital Human Modeling," in *Proceedings of 2000 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2000), Workshop Tutorial Note,* pp. 1–14

[2] H. Mizoguchi, T. Sato, T. Ishikawa, "Robotic Office Room to Support Office Work by Human Behavior Understanding Function with Networked Machines," *IEEE/ASME Transactions on Mechatronics*, Vol. 1, No. 3, September 1996, pp. 237–244

[3] A. Hopper, P. Steggles, A. Ward, P. Webster, "The Anatomy of a Context-Aware Application," in *Proceedings of 5th Annual International Conference Mobile Computing and Networking (Mobicom99)*, pp. 59-68

[4] S. Shih, M. Minami, H. Morikawa, T. Aoyama, "An Implementation and Evaluation of Indoor Ultrasonic Tracking System," in *Proceedings of the 2001 IEICE Domestic General Conference*, B-7-20

[5] http://www.isense.com

[6] K.C. Ho "Solution and Performance Analysis of Geolocation by TDOA, " *IEEE Transaction on Aerospace and Electronic Systems*, Vol. 29, No. 4, October 1993, pp. 1311-1322

[7] D.E. Manolakis, "Efficient Solution and Performance Analysis of 3-D Position Estimation by Trilateration, " *IEEE Trans. on Aerospace and Electronic Systems*, Vol. 32, No. 4, October 1996, pp. 1239-1248

[8] P. J. Rousseeuw, A. M. Leroy, "*Robust Regression and Outlier Detection*," Wiley, New York, 1987

[9] M.A. Fishler, R.C. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Application to Image Analysis and Automated Cartography," *Communication of the ACM*, Vol. 24, June 1981, pp. 381–395