REAL WORLD SENSORIZATION AND VIRTUALIZATION FOR OBSERVING HUMAN ACTIVITIES

Koji Kitamura
Tokyo University of Science
2641, Yamazaki Noda-shi Chiba 278-8510 JAPAN

Yoshifumi Nishida, Makoto Kimura
Digital Human Research Center, National Insititute of Advanced Industrial Science and Technology (AIST)
2-41-6, Aomi Koto Tokyo 135-0064 JAPAN
CREST, JST (Japan Science and Technology Agency)

Hiroshi Mizoguchi
Tokyo University of Science
2641, Yamazaki Noda-shi Chiba 278-8510 JAPAN

Keywords: Human Behavior Detection, Ubiquitous Computing, Sensorization, Distributed Sensor.

Abstract: This paper describes a method for robustly detecting and efficiently recognizing daily human behavior in real world. The proposed method involves real world sensorization for robustly observing his or her behavior using ultrasonic 3D tags, which is a kind of an ultrasonic location system, real world virtualization for creating a virtual environment through modeling 3D shape of real objects by a stereovision system, and virtual sensorization of the virtualized objects for quickly registering human activities handling objects in real world and efficiently recognizing target human activities. As for real world sensorization, this paper describes algorithms for robustly estimating 3D positions of objects that a human handles. This paper also describes a method for real world virtualization and virtual sensorization using the ultrasonic 3D tag system and a stereovision system.

1 INTRODUCTION

The observation of human activities in the real world makes it possible to input personal information into a computer without any conscious operation of an interface. Human-centered applications based on implicit input of human information require the facility to observe and recognize activities as a basis. This paper describes a method for realizing a function for robustly and efficiently detecting daily human activity events in the real world.

There are two problems in realizing and utilizing a function for recognizing human activity in the real world: the robust observation of a human activity pattern, and the efficient recognition of meaning of activity from the observed pattern. Without solving the first problem, a human activity pattern to be analyzed cannot be obtained. 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 activity, the idea of object-based activity recognition has been proposed (Mizoguchi et al., 1996). In theory, the activity 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 activity recognition is becoming increasingly realistic with the progress in ubiquitous computing technology such as 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 activity recognition separately through the development of applied technology.

As for robust observation of human activity, this paper describes a method for ”sensorizing objects in

Kitamura K., Nishida Y., Kimura M. and Mizoguchi H. (2004).
REAL WORLD SENSORIZATION AND VIRTUALIZATION FOR OBSERVING HUMAN ACTIVITIES.
In *Proceedings of the Sixth International Conference on Enterprise Information Systems*, pages 15-20
DOI: 10.5220/0002625400150020

real world" using a special device. The present authors have developed a three-dimensional ultrasonic location and tagging system, an ultrasonic 3D tagging system, for that purpose. 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 (Hopper et al., 1999; Shih et al., 2001). The system presented in the present paper is developed specifically to address the issue of robustness and accuracy in real time when a person handles objects having ultrasonic 3D tags.

As for efficient recognition of target activity, this paper describes a method for "creating virtual objects" and "virtually sensorizing the virtualized objects" for recognizing target activity. It is important to create virtual environment extracting essential features of the real world so that the created virtual environment can eliminate unnecessary process but can maintain association with target phenomena of the real world. The method enables a user to quickly register target activity to be recognized interactively on a computer.

This paper is organized as follows. The next section describes the method for real world sensorization using the ultrasonic 3D tagging system. The developed ultrasonic 3D tagging system is introduced briefly. Algorithms for robustly measuring 3D positions of the objects handled by a person and experimental results are shown. Section 3 describes the method for creating virtual objects and virtually sensorizing the virtual objects using the ultrasonic 3D tagging system and a stereovision system.

2 REAL WORLD SENSORIZATION FOR ROBUST DETECTION OF HUMAN ACTIVITY

2.1 Ultrasonic 3D tag

The ultrasonic 3D tagging system developed by the authors(Nishida et al., 2003) 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. The system can keep the sampling frequency as high as 50 Hz when the number of the target transmitters are less than three or four(Hori et al., 2003). A user of the system can attach ultrasonic receivers on arbitrary positions of ceilings or walls and can easily calibrate the receivers' positions using a portable calibration device.

Figure 1 shows the experimental systems for evaluating a function for robust detection of human activity. The experimental results are shown later. The upper part of the figure shows a tiny, a small, and a long life battery type of ultrasonic 3D tag and objects with ultrasonic 3D tags.

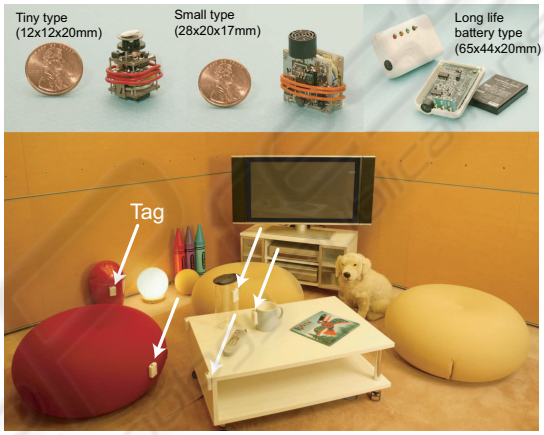

Figure 1: Ultrasonic 3D tag and sensorized environment

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. Some objects were fitted with two transmitters.

2.2 Multilateration method 1: linearization of the minimization problem

Trilateration or multilateration algorithms have been proposed in the field of aerospace(Ho, 1993; Manolakis, 1996). This paper presents the multilateration algorithms applicable to a more general case that multiple ultrasonic receivers are put on arbitrary positions. 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.

$$(x_i - x)^2 + (y_i - y)^2 + (z_i - z)^2 = l_i^2, \quad (1)$$
$$(x_j - x)^2 + (y_j - y)^2 + (z_j - z)^2 = l_j^2. \quad (2)$$

By subtracting Eq. (2) from Eq. (1), we obtain an equation for intersecting planes between the spheres.

$$2(x_j - x_i)x + 2(y_j - y_i)y + 2(z_j - z_i)z = l_i^2 - l_j^2 - x_i^2 - y_i^2 - z_i^2 + x_j^2 + y_j^2 + z_j^2 \quad (3)$$

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

$$\boldsymbol{AP} = \boldsymbol{B}, \quad (4)$$

$$\text{where} \quad \boldsymbol{P} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad (5)$$

$$\boldsymbol{A} = \begin{pmatrix} 2(x_0 - x_1) & 2(y_0 - y_1) & 2(z_0 - z_1) \\ 2(x_0 - x_2) & 2(y_0 - y_2) & 2(z_0 - z_2) \\ 2(x_0 - x_3) & 2(y_0 - y_3) & 2(z_0 - z_3) \end{pmatrix} \quad (6)$$

$$\boldsymbol{B} = \begin{pmatrix} l_1^2 - l_0^2 - x_1^2 - y_1^2 - z_1^2 + x_0^2 + y_0^2 + z_0^2 \\ l_2^2 - l_0^2 - x_2^2 - y_2^2 - z_2^2 + x_0^2 + y_0^2 + z_0^2 \\ l_3^2 - l_0^2 - x_3^2 - y_3^2 - z_3^2 + x_0^2 + y_0^2 + z_0^2 \\ \vdots \end{pmatrix}. \quad (7)$$

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

$$\boldsymbol{P} = (\boldsymbol{A^T A})^{-1} \boldsymbol{A^T B}. \quad (8)$$

This method minimizes the square of the distance between the planes expressed by Eq. (3) and the estimated position. In actual usage, the rank of matrix $\boldsymbol{A}$ must be considered.

2.3 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) (Rousseeuw and Leroy, 1987; Fishler and Bolles, 1981) estimator is adopted to eliminate the undesirable effects of outliers. The procedure is as follows.

1. Randomly select three distances measured by three receivers (jth trial).
2. Calculate the position (x_{cj}, y_{cj}, z_{cj}) by trilateration.
3. Calculate the error ε_{cji} for all receivers ($i = 0, 1, ..., n$) by Eq. (9), 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.

$$\varepsilon_{cji} = \left| l_i - \sqrt{(x_i - x_{mj})^2 + (y_i - y_{mj})^2 + (z_i - z_{mj})^2} \right| \quad (9)$$

$$\varepsilon_{mj} = med_j \, |\varepsilon_{cji}| \quad (10)$$

$$(\hat{x}, \hat{y}, \hat{z}) = min \, \varepsilon_{mj} \quad (11)$$

2.4 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 2 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 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.

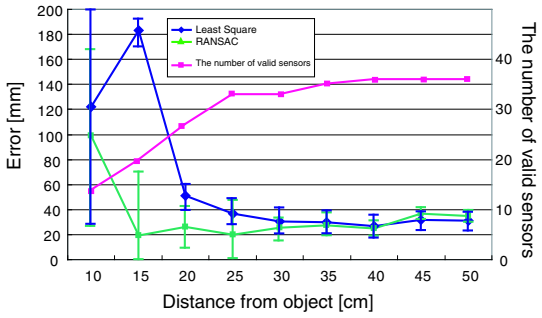

Figure 2: Accuracy of the ultrasonic tagging system when occlusion due to a hand occurs

2.5 Experimental results: robust detection of human activity

Figure 3 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. Basically this frequency decreases to $50/n$ Hz when n objects are being monitored although the system can keep the sampling frequency as high as 50 Hz when the number of the target transmitters is less than three or four(Hori et al., 2003). 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.

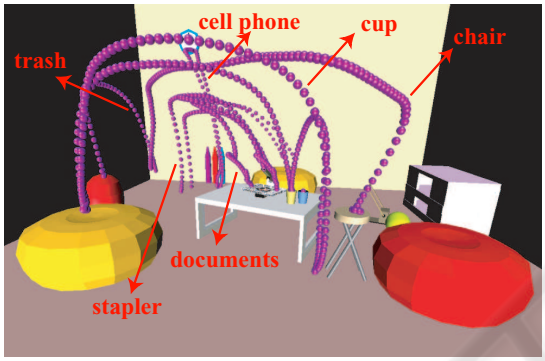

Figure 3: Robust detection of human activity

3 VIRTUAL SENSORIZATION FOR QUICK REGISTRATION AND EFFICIENT RECOGNITION OF HUMAN ACTIVITY

3.1 Virtual Sensorizaztion

This section describes a method for virtualizing real objects and virtually sensorizing the virtualized objects for efficiently recognizing human acitivities.

The real objects virtualization enables to extract essential geometric features of real objects by simplifying 3D shape of real objects. The 3D shape simplification is performed using a stereovision fitted with ultrasonic 3D tags in combination with interactive software. The software abstracts the shapes of objects in real world as simple two-dimensional shapes such as lines, circles, or polygons.

The virtual sensorization of virtualized objects enables to extract essential physical phenomena among the real objects relating to target activity events. In order to describe the real world events when a person handles the objects, the software abstracts the function of objects as simple phenomena such as touch, detouch, or rotation. The software adopts the concept of virtual sensors and effectors to allow the user to define the function of the objects easily through simple mouse operations. For example, to define the activity "put a cup on the desk", the user simplifies the cup and the desk as simple two-dimensional models of a circle and a rectangle using the photo-modeling function of the software. Using a function for editting virtual sensors, the user then adds a "touch" virtual sensor to the model of the desk, and adds a "bar" effector to the model of the cup. Details of real object virtualization, virtual sensorization of virtualized objects, registration of target activity, and real time detection and recognition of the target activity are described in the following.

3.2 Virtual Sensorizaztion Procedure

Step A: Real object virtualization Figure 7 shows examples of simplified 3D shape models of objects such as a tissue, a cup, a desk and a stapler. The cup is expressed as a circle and the desk as a rectangle. The simplification is performed using a stereovision in combination with photo-modeling function (Fig. 6) of the software.

There is a problem with photo-modeling function of stereovision. It is difficult to have a target object to be modeled in stereovision's sights.To solve the problem, the authors developed the stereovision system fitted with multiple ultrasonic 3D tags. We call the system an "UltraVision". Since the UltraVision can track its position and posture, it is possible to move the UltraVision freely when the user creates simplified 3D shape models and the system can integrate the created models into the world coordinate system. Concrete process for integrating models is described in the following.

We assume that the UltraVision is placed at position P1 initially and moves from position P1 to position P2. The UltraVision has sterevision system and the ultrasonic 3D tags. There are two coordinate systems, U1 and C1 as shown in Fig. 4. U1 indicates the local coordinate system whose origin is the position of a tag attached on the UltraVision placed at position P1, and C1 indicates the local coordinate system of stereovision. Coordinate systems U2 and C2 are defined similarly to the case of U1 and U2.

Since the relative location between the stereovision and tags doen't change even if the Ultravision moves, the transformation matrices M_{c1u1} and M_{c2u2} are constant as follows.

$$M_{c1u1} = M_{c2u2} = M_{cu} \tag{12}$$

If M_{cu} is known, we can transform the local coordinate value P_{c1} and P_{c2} to the world coordinate value

P_w using the following equation.

$$P_w = M_{u1w} \cdot M_{cu} \cdot P_{c1} \quad (13)$$

$$P_w = M_{u2w} \cdot M_{cu} \cdot P_{c2} \quad (14)$$

Note that M_{u1w} and M_{u2w} can be calculated using the positions of multiple tags attached on the UltraVision after the UltraVision moves.

Example of moldeling large room actualy using UltraVision based on this process is Fig. 5.

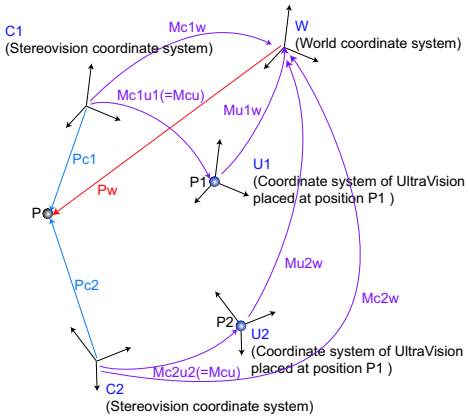

Figure 4: Coordinate Conversion in UltraVision system

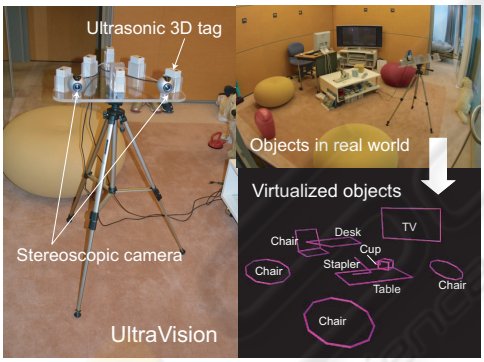

Figure 5: UltraVision for virtualizing objects and example of virtualized objects

Step B: Virtual sensorization of virtualized objects The software creates a model of an object's function by attaching virtual sensors and effectors to the model created in step A. Virtual sensors and effectors are prepared in advance by the software and function as sensors and effectors affecting the sensors on computer. The current system has an "angle sensor" for detecting rotation, a "bar effector" to represent touch, and a "touch sensor" for detecting touch. In the right part of Fig. 8, red indicate a virtual bar effector, and green indicates a virtual touch sensor. Using simple mouse operations, it is possible to add virtual sensors/effectors to the 3D shape model.

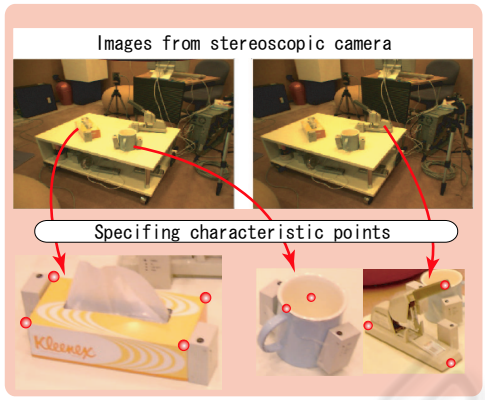

Figure 6: Photo-modeling by stereovision system

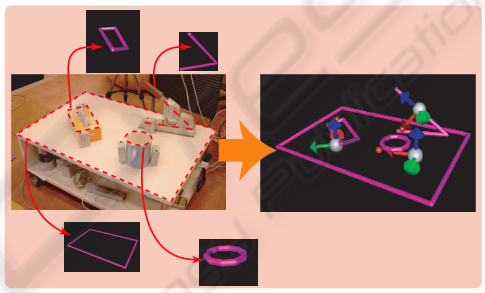

Figure 7: Real object virtualization

Step C: Associating virtual object sensor with human activity event Human activity can be described using the output of the virtual sensors created in Step B. In Fig. 9, red indicates that the cup touches the desk, and blue indicates that the cup does not. By creating a table describing the relationship between the output of the virtual sensors and the target events, the system can output symbolic information such as "put a cup on the desk" when the states of the virtual sensors change.

Step D: Real time detection and recognition of human activity event When the software inputs the

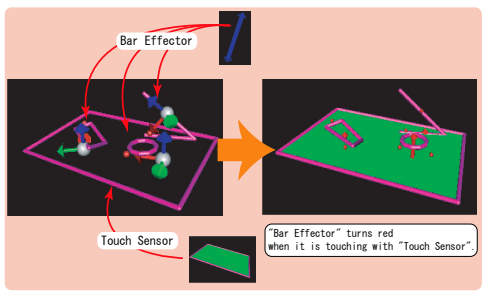

Figure 8: Create model of physical object's function using virtual sensors/effectors

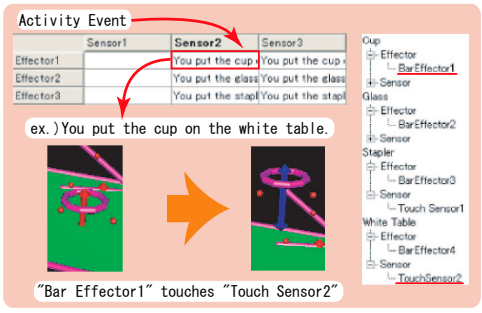

Figure 9: Associate output of virtual sensors with target activity event

position data of the ultrasonic 3D tag, the software can detect the target events using the virtual sensors and the table defined in Step A to C, as shown in Fig. 10

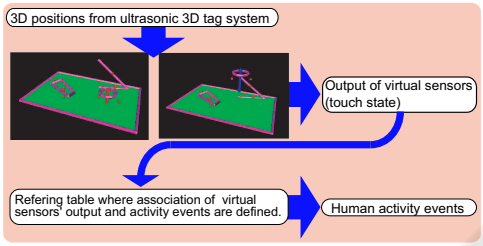

Figure 10: Real time detection and recognition of human activity by virtual object sensor

4 CONCLUSION

This paper described a method for robustly detecting human activity in real world and a method for quickly registering and efficiently recognizing target activity.

The robust detection of human activity is performed by sensorizing objects in real world using an ultrasonic 3D tagging system, which is a kind of an ultrasonic location sensor. In order to estimate the 3D position with high accuracy and robustness to occlusion, the authors propose two estimation methods, one based on a least-squares approach and one based on RANSAC. 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 and robustness to occlusion by increasing the number of ultrasonic receivers and by adopting a robust estimator such as RANSAC to estimate the 3D position based on redundant distance data.

The efficient recognition of human activity involves a method for creating virtual objects using the ultrasonic 3D tagging system and a stereovision and a method for virtually sensorizing the created virtual objects interactively on a computer. To verify the effectiveness of the function, using a stereovision with ultrasonic 3D tags and interactive software, the authors registered activity such as "put a cup on the desk" and "staple document" through creating the simplified 3D shape models of ten objects such as a TV, a desk, a cup, a chair, a box, and a stapler.

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

REFERENCES

Fishler, M. and Bolles, R. (1981). Random sample consensus: A paradigm for model fitting with application to image analysis and automated cartography. *Communication of the ACM*, 24:381–395.

Ho, K. (1993). Solution and performance analysis of geolocation by tdoa. *IEEE Transaction on Aerospace and Electronic Systems*, 29(4):1311–1322.

Hopper, A., Steggles, P., Ward, A., and Webster, P. (1999). The anatomy of a context-aware application. In *Proceedings of 5th Annual International Conference Mobile Computing and Networking (Mobicom99)*, pages 59–68.

Hori, T., Nishida, Y., Kanade, T., and Akiyama, K. (2003). Improving sampling rate with multiplexed ultrasonic emitters. In *Proceedings of 2003 IEEE International Conference on Systems, Man and Cybernetics*, pages 4522–4527.

Manolakis, D. (1996). Efficient solution and performance analysis of 3-d position estimation by trilateration. *IEEE Trans. on Aerospace and Electronic Systems*, 32(4):1239–1248.

Mizoguchi, H., Sato, T., and Ishikawa, T. (1996). Robotic office room to support office work by human behavior understanding function with networked machines. *IEEE/ASME Transactions on Mechatronics*, 1(3):237–244.

Nishida, Y., Aizawa, H., Hori, T., Hoffman, N., Kanade, T., and Kakikura, M. (2003). 3-d ultrasonic tagging system for observing human activity. In *Proceedings of IEEE International Conference on Intelligent Robots and Systems*, pages 785–791.

Rousseeuw, P. and Leroy, A. (1987). *Robust Regression and Outlier Detection*. Wiley, New York.

Shih, S., Minami, M., Morikawa, H., and Aoyama, T. (2001). An implementation and evaluation of indoor ultrasonic tracking system. In *Proceedings of the 2001 IEICE Domestic General Conference*.