Real World Sensorization for Observing Human Behavior and Its Application to Behavior-to-Speech

Yoshifumi Nishida
Digital Human Research Center, AIST, &CREST, JST
2-41-6 Aomi Koto Tokyo
135-0064 Japan
y.nishida@aist.go.jp

Koji Kitamura
Tokyo University of Science
2641 Yamazaki Noda-shi
Chiba 278-8510 Japan
k.kitamura@aist.go.jp

Hiroshi Aizawa
Tokyo Denki University
2-2 Kandanishiki-cho Chiyoda
Tokyo 101-8457 Japan
h-aizawa@aist.go.jp

ABSTRACT

This paper describes a method for robustly detecting and efficiently recognizing daily human behavior in the real world. The proposed method involves real-world sensorization using ultrasonic tags to robustly observe behavior, real-world virtualization to create a virtual environment by modeling real objects using a stereovision system, and virtual sensorization of virtualized objects in order to quickly register the handling of objects in the real world and efficiently recognizing specific human behavior. A behavior-to-speech system created based on this recognition method is also presented as a new application of this technology.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

General Terms

Measurement

Keywords

Human Behavior Detection, Ubiquitous Computing, Distributed Sensor, Sensor Fusion

1. INTRODUCTION

The observation of human behavior in the real world makes it possible to input personal information into a computer system without conscious operation of an interface. Human-centered applications based on implicit input of human information require the facility to observe and recognize behavior as a basis. This paper describes a system that realizes this function, robustly and efficiently detecting daily human behavior in the real world, and presents a behavior-to-speech system as a new application based of this technology.

The realization and utilization of a function for recognizing human behavior in the real world requires robust observation of a human behavior pattern and efficient recognition of the meaning of that behavior from the observed pattern. Without solving the first problem, the behavior 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. This paper describes a method for solving these two problems.

The robust observation of human behavior is achieved here by sensorizing objects in the real world using a special ultrasonic tag. The authors have already developed an ultrasonic tagging system that can locate objects in three-dimensional (3D) space. The ultrasonic tagging system is superior to other location techniques such as visual, tactile, and magnetic systems in terms of both cost and robustness against environmental noise.

The efficient recognition of target behavior is provided by a virtual object model in which target objects are virtually sensorized. The creation of a virtual environment, extracting the essential features of the real world, is important in order to eliminate unnecessary processes while maintaining the association with target phenomena in the real world. The proposed method allows a user to quickly register the target behavior to be recognized interactively using a computer. Based on this new recognition technique, a behavior-to-speech system is also presented here as an example of potential applications.

2. SENSORIZATION FOR ROBUST DETECTION OF HUMAN BEHAVIOR

The authors have previously developed an ultrasonic tagging system for tracking objects in 3D space. The system consists of an ultrasonic receiver, ultrasonic transmitter, time-of-flight measurement device, network, and personal computer. The ultrasonic receiver receives ultrasonic pulses emitted from the ultrasonic transmitter and amplifies the received signal. The time-of-flight measurement function records the travel time of the signal from transmission to reception, and the network synchronizes the system and collects the time-of-flight data from the ultrasonic receiver. The positions of objects are calculated based on three or more time-of-flight results. The sampling frequency of the proposed system is 50 Hz, and this sampling frequency can be maintained with four or fewer target transmitters [2].

Figure 1 shows a photograph of various types of ultrasonic transmitters; a tiny version, small version, and tag with a small long-life battery. The ultrasonic tagging system calculates the 3D position of an object by trilateration using three distance measurements. The system can perform multilateration by either a least-squares method using

IUI'04, January 13–16, 2004, Madeira, Funchal, Portugal.
ACM 1-58113-815-6/04/0001.

redundant distance data, or by robust random sample consensus (RANSAC) estimation [1]. The authors have confirmed that the system can track objects more robustly by the RANSAC approach when a human handles the objects. Figure 2 shows an example of the measured trajectory of a cup when a person moves the cup to a chair, the floor, and finally a desk. The figure demonstrates that the system can robustly track the positions of objects in most places of the room regardless of obstruction by a hand or body.

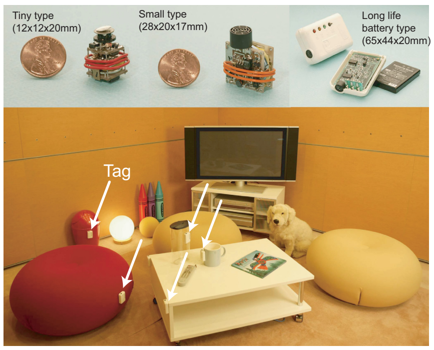

Figure 1: Ultrasonic 3D tag and sensorized environment

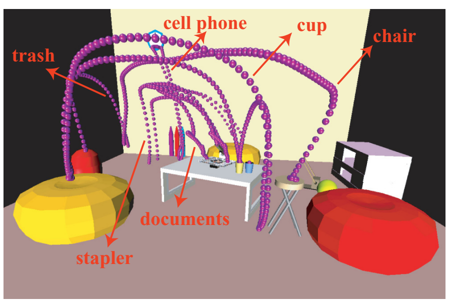

Figure 2: Robust detection of human behavior

3. VIRTUAL SENSORIZATION FOR QUICK REGISTRATION AN DEFFICIENT RECOGNITION OF HUMAN BEHAVIOR

3.1 Virtual Sensorization

The virtualization of real objects efficiently extracts the essential geometric features of real objects by simplifying the 3D shape of the real objects. The 3D shape simplification is performed using a stereoscopic camera and ultrasonic 3D tags (see Fig. 3) via interactive software. The software abstracts the shapes of objects in the real world as simple 3D shape such as lines, circles, or polygons.

The virtual sensorization of virtualized objects allows the essential physical phenomena to be extracted for the target object. In order to describe the real-world events when a person handles an object, the software abstracts the interactions among objects as simplified phenomena such as touch, release, or rotation. The software adopts the concept of virtual sensors and effectors to allow the user to define the function of the objects easily by mouse operations.

3.2 Real object virtualization (Step A)

Step A in Figure 4 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 stereoscopic camera with ultrasonic 3D tags, referred to as an UltraVision module, and photo-modeling software. The multiple ultrasonic 3D tags associated with the camera allows the system to track the position and attitude of the camera. Therefore, it is possible to move the camera freely when the user creates simplified 3D shape models while the system integrates the created 3D shape models into the world coordinate system.

Since the created 3D shape models are simplified, the system does not need complex processing for detecting interaction among the objects. This simplification makes it possible to efficiently describing the real-world phenomena.

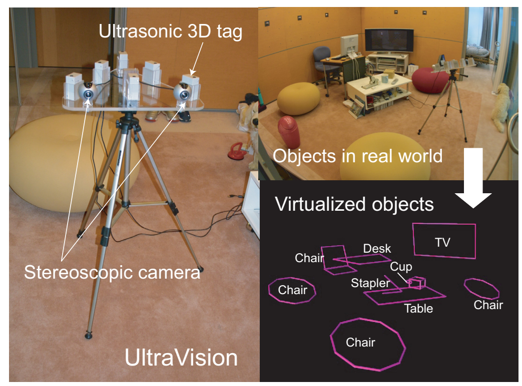

Figure 3: UltraVision module for virtualizing objects and example of virtualized objects

3.3 Virtual sensorization of virtualized objects (Step B)

The software creates the model of an object's motion by attaching virtual sensors and effectors provided by the software to the 3D shape model created in step A. The current system has an angle sensor for detecting rotation, a bar effector to represent touch, and a touch sensor to detect touch. Using this software based on the concept of attaching the virtual sensors and effectors, the user can quickly create software for detecting the interactions among objects without any programming.

3.4 Associating virtual object sensirs with human behavior events (Step C)

Human behavior can be described by the output of the virtual sensors created in Step B. 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. The output can be used for collecting symbolic data of human behavior for various purposes, for example, analyzing human behavior, and developing application based on the human behavior.

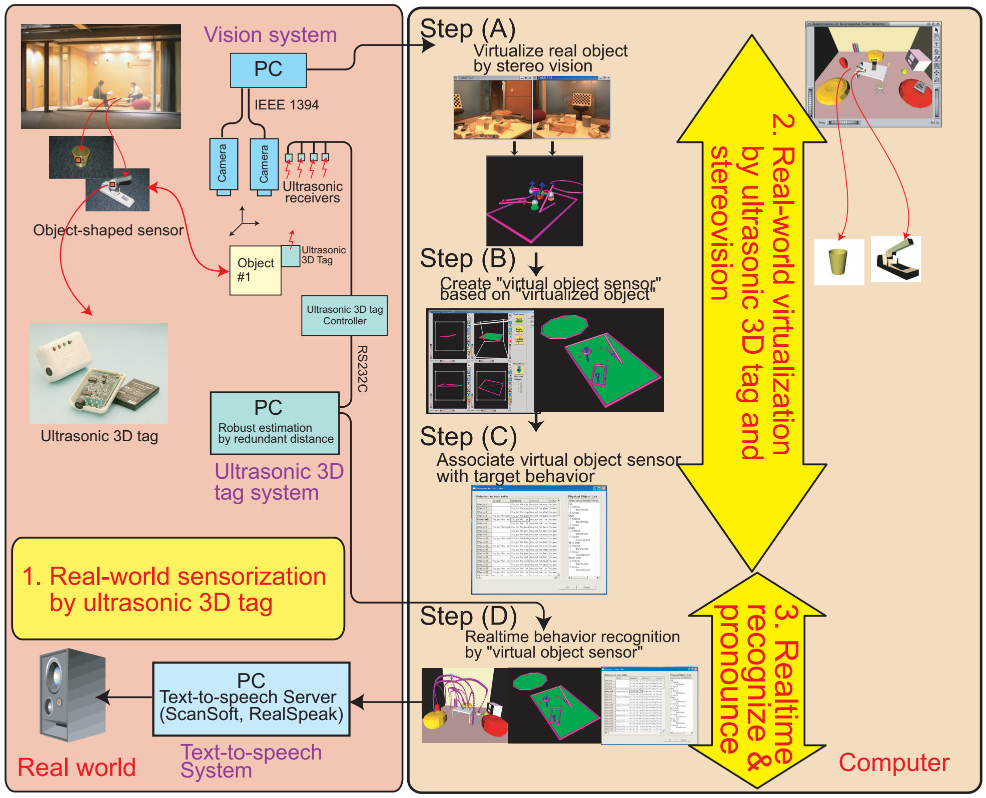

Figure 4: Configuration of behavior-to-speech system

3.5 Real-time detection and recognition of human behavior event (Step D)

By inputting the position data of ultrasonic tags into the software, the software can detect the target events using the virtual sensors and the table defined in Steps A to C as shown in Fig. 4. Since the ultrasonic tag described in Section 2 can robustly track the objects and the virtual sensor created through Step A to C can efficiently recognize the target behavior, the whole system can robustly detect and efficiently recognize daily human behavior in the real world.

4. APPLICATION TO BEHAVIOR-TO-SPEECH

The behavior-to-speech system was created to allow "arning by doing" the real world as an application of the functions of real-world sensorization, real-world virtualization, and virtual sensorization of virtualized objects. The main target of the behavior-to-speech system is to assist a learner of a second language. The configuration of the constructed system is shown in Fig. 4. The system consists of the ultrasonic tagging system for real-world sensorization, the stereoscopic camera with ultrasonic tags (UltraVision) for real-world virtualization and virtual sensorization of virtualized objects, and a text-to-speech server computer for pronouncing the created sentences.

5. CONCLUSION

This paper proposed a method for robustly detecting and efficiently recognizing daily human behavior in the real world. The proposed method involves real-world sensorization to robustly observe behavior, real-world virtualization to simplify the real-world phenomena, and virtual sensorization of virtualized objects to quickly register the handling of objects in the real world. This paper described a system for implementing the proposed method using a ultrasonic tagging system which is a kind of 3D location sensor, and a stereovision system. This paper also presented a behavior-to-speech system created based on the proposed method.

6. ADDITIONAL AUTHORS

Additional authors: Toshio Hori (Digital Human Research Center, AIST, &CREST, JST, email: `t.hori@aist.go.jp`), Makoto Kimura (Digital Human Research Center, AIST, &CREST, JST, email: `makoto.kimura@aist.go.jp`), Takeo Kanade (Carnegie Mellon University, &Digital Human Research Center, AIST, &CREST, JST, email: `Takeo.Kanade@cs.cmu.edu`), Hiroshi Mizoguchi (Tokyo University of Science, email: `hm@rs.noda.tus.ac.jp`).

7. REFERENCES

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

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