Vision-based Gestural Interaction Using Plush Toys

Yoshinori Kawasaki[1], Takeo Igarashi[1, 2], Takashi Ajioka[3] and Ikuji Honda[3]
[1]Graduate School of Information Science and Technology, The University of Tokyo / [2]PRESTO, JST
[3]Graduate School of Integrated Design Engineering, Keio University
E-mail: {kwsk, takeo}@ui.is.s.u-tokyo.ac.jp, mr041433@hc.cc.keio.ac.jp, honda@math.keio.ac.jp

ABSTRACT
As more and more computers are embedded in our living spaces, appropriate techniques for interacting with these devices become increasingly important. We currently lack design principles to make such information appliances simple, entertaining, and attractive for ordinary, non-technology oriented users. We are currently developing gestural interaction techniques using everyday objects like plush toys as a testbed for exploring such design principles. These everyday objects are more familiar and appealing than special purpose devices like remote controllers. We use computer vision techniques to avoid attaching cumbersome sensors to the objects.

KEYWORDS: Plush toy, Everyday objects, Remote Control, Ubiquitous computing, Gesture, Vision-based interface.

INTRODUCTION

Interaction techniques for the remote control of information appliances are an important research topic in ubiquitous computing. IR-based remote controllers for electronic appliances are widely used, allowing people to control TVs, air conditioners, and lamps from the comfort of their sofa or bed. However, having a special controller for each device will be unsuitable in a ubiquitous computing environment where computers and services are embedded everywhere. Current remote control devices are not natural components of our daily environment, so they inevitably make our rooms messy and uncomfortable. Instead, we need methods for controlling our appliances in a natural and pleasing way.

Much of the previous research on remote control interfaces can be divided into two categories: interfaces with a special device for the user, and those without. Gestural Pendant [5], Ubi-Finger [6] and XWand [8] are examples of systems that require the user to wear special purpose devices like pendants or gloves. Although they require the user to carry only one device, it may still be bothersome to attach and detach this device each time the user wants to control appliances; furthermore, the device may be expensive.

An alternative is to embed sensors in the environment and use them to recognize the user's gestures and postures; this approach has been used to control a television through hand gestures [1] and to detect hands on the surfaces of a room [2]. Since these systems observe the user's motions, wearing a special device is not required. However, as in any

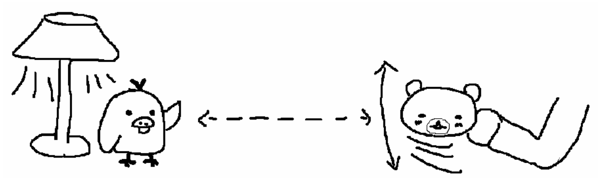

Figure 1: The user shakes a plush toy, and the system recognizes this gesture as a request to turn on the light.

Figure 2: Plush toys vs. remote controllers. Plush toys are more attractive in living rooms.

recognition-based system, the user's intention is not always perfectly understood. A simple gesture like an open hand for turning on the television can be easily misunderstood by the system under real world conditions. We approached this problem of ambiguity and recognition error by designing and testing various interaction protocols in real world settings.

PLUSH TOY GESTURAL INTERACTION

We are currently developing gestural interaction techniques to control computing devices remotely using everyday objects such as plush toys, without any special devices attached to the objects (Figure 1). As one example, we developed a prototype system to control appliances in a living room, equipped with a single camera and processor for each appliance, using computer-vision techniques. We chose everyday objects that exist naturally in a living environment. This approach is appealing to end users since they can use their favorite object as a controller (Figure 2).

In our prototype, we decided to use plush toys since their soft body is easy to grasp and safe to shake, though of course it is technically possible to track other everyday objects. Plush toys tend to have vivid colors, making tracking easier; also, as other researchers have pointed out [3,7], they are especially pleasing to children.

Cameras have become a very inexpensive commodity, with a cost approaching the cost of a keyboard and mouse. We chose to use a single camera to minimize the equipment and setup costs. Furthermore, using only one camera makes it theoretically possible for the camera to be pre-installed with the appliance, instead of installing the camera on the wall or ceiling later on. Naturally, precise tracking is impossible with only one camera, but coarse results are good enough for our purposes.

To develop a practical system for end users, the initial configuration must be simple. Even a knowledgeable user will be frustrated if the system takes a long time to configure, since their goal is simply operating devices. All the user must do before using our system is put an object in front of the camera and push a button to register the object. We believe that this is easy enough for ordinary users, and it allows for rapid re-registration on demand.

HANDLING AMBIGUOUS INPUT

In order to develop a realistic and deployable gesture-based service, the challenge is to interpret ambiguous user actions and avoiding unintentional or accidental recognition. Our main research focus is the design of the interaction protocol between the human and the computer to handle ambiguous input from the user. Here we introduce an example interaction protocol for our current prototype. This protocol requires further investigation; we plan to evaluate it with a user study in the future.

Tracking an object instead of a bare hand can itself reduce accidental recognition. In addition, we use a combination of timing and "unusual" motion as a trigger to begin recognition, so as to avoid unexpected recognition. An example of a trigger action is described later.

Recognition-based interfaces are inherently error-prone, and common strategies for error correction are repetition and choice [4]. For simplicity, we adopt a cancel-and-repeat strategy. The user can cancel recognition by simply hiding the plush toy or showing its back to the camera (the system considers these states equivalent, since it cannot find the toy). The system provides feedback while recognizing so that the user can cancel a command before the command is executed.

EXAMPLE SCENARIO

A user sitting on a sofa grasps a stuffed bear on a table, and holds it up for a moment toward a camera on a TV. This is the trigger action for starting recognition. After a moment, the system makes a small beeping sound to signal the start of recognition. The user shakes the toy vertically, as if it were nodding, to turn on the TV. She then moves it in a clockwise circular motion to adjust the sound volume (Figure 3).

INFORMAL USER FEEDBACK

Although our system is still in an exploratory stage, we asked four volunteers to try the system for some early informal feedback. All the subjects felt that the system was simple and compelling, and reported that they were willing to use in their home. All the participants got used to the system within a few tries. We observed that some users were able to use the system without watching the camera, while they were on the sofa reading a book.

One user critiqued that she might not memorize the gestures if there were more than ten. This restriction on the size of the gesture set is problematic given the number of labeled

Figure 3: Prototype system in use. The user is happily shaking the green stuffed frog.

buttons on a traditional remote controller. This tradeoff between functionality and memorability is a common problem in gesture-based interfaces. Possible approaches might be a guessable gesture set that is well-suited to each application, or a standard gesture set that is applicable for every application; evaluating these approaches will require further exploration.

CONCLUSION AND FUTURE WORK

We are developing gestural interaction techniques that allow users to control appliances remotely using everyday objects. We plan to conduct a user study to help us refine an efficient and usable interaction protocol. We believe that our technique is well-suited to common living environments.

We also plan to exploit the movement of the stuffed animal's hands and legs or the location where the user touches the toy (e.g. hiding teddy's eyes, touching its head, etc), in order to broaden the interaction possibilities.

ACKNOWLEDGMENTS

This work is partly funded by Microsoft.

REFERENCES

1. Freeman, W. T., et al. *Television Control by Hand Gestures.* Int'l Workshop on Automatic Face and Gesture Recognition, 1995.
2. Jerry Alan Fails et al. *Light Widgets: Interacting in Every-day Spaces.* IUI 2002.
3. Johnson, M. P., et al. *Sympathetic Interfaces: Using a Plush Toy to Direct Synthetic Characters.* CHI 1999.
4. Mankoff, J., et al. *Interaction techniques for ambiguity resolution in recognition-based interfaces.* UIST 2000.
5. Starner, T., et al. *The Gesture Pendant: A Self-illuminating, Wearable, Infrared Computer Vision System for Home Automation Control and Medical Monitoring.* ISWC 2000.
6. Tsukada, K., et al. *Ubi-Finger: Gesture Input Device for Mobile Use.* APCHI 2002
7. Vaucelle, C., et al. *Doltalk: A computational toy to enhance children's creativity.* CHI 2002.
8. Wilson, A., et al. *XWand: UI for Intelligent Spaces.* CHI 2003