Narrative Hand: Applying a fast finger-tracking system for media art

Yasuto Nakanishi*, Kenji Oka**, Masayuki Kuramochi***
Shohei Matsukawa***, Yoichi Sato**, Hideki Koike**
* Graduate School of Information Systems, Univ. of Electro-Communications
** Institute of Industrial Science, Univ. of Tokyo
*** 000STUDIO
{naka, koike}@vogue.is.uec.ac.jp, {oka, ysato}@iis.u-tokyo.ac.jp, {kura, sho}@000studio.com

Abstract

This paper describes the development of a vision-based finger tracking system for media art; the system utilizes an infrared camera and a Kalman filter to accomplish this finger recognition image processing technology. The system is stable and fast, and is able to recognize not only gestures but also the speeds of gestures. Our work, called the "Narrative Hand", switches movies to show the speed at which users close their hands. Our system was exhibited for six days at ACM SIGGRAPH2001 Emergent Technologies, and many visitors experienced it.

1. Introduction

We have developed an augmented desk interface system called the EnhancedDesk which allows users to interact with both physical objects and digital information projected onto a desk simultaneously based on direct manipulation with their bare hands [1][2]. EnhancedDesk is equipped with an infrared camera, a color video camera, and two LCD projectors. Real-time tracking of a user's hands and fingertips is an essential part of EnhancedDesk; a user can simultaneously manipulate both physical objects and electrically projected objects on a desk by using natural hand gestures. An infrared camera is used for reliable detection of the user's hands even when various kinds of information are displayed on the tabletop. Once image regions corresponding to the user's hands are identified, fingertip positions are determined by using several image processing procedures, including morphological operations and template matching [3]. During the past summer, EnhancedDesk was selected as one of the exhibits at Emerging Technologies of SIGGRAPH2001, and more than 2000 visitors tried our system during a period of six days. We developed four demonstration applications; Narrative Hand, introduced in this paper, is one of them. We developed our demonstrations so that they emphasized the following aspects of our system: intuitive interaction based on direct manipulation, symbolic gestures with users' own hands, accuracy and reliability of our vision-based technique for tracking multiple hands and fingertips in an uncontrolled environment. In particular, we produced Narrative Hand with keeping in mind that our finger-tracking system was very fast.

2. Narrative Hand

In NarrativeHand, we see various kinds of objects such as a tomato or a piece of tofu. These objects are squeezed or smashed by a hand projected on a desk depending on how fast users close their own hands. For example, objects are squeezed gently when users close their hands slowly. In contrast, when they close them quickly, objects are smashed violently (Figure 1). Although the displayed hand is not a user's own hand, the user somehow feels a strange identification with the virtual hand. In our work, the shown objects are a tomato, a kiwi, an egg, an orange, a paprika, a piece of tofu, spaghetti, tagliatelle, blueberry jam, a cup of yogurt and a tube of mayonnaise. Our concept text is as follows.

> *Quietly watching the object in your virtual hand, you crushed it. You might have some such experiences in your life. You can damage the existence of the object by tearing off, twisting, beating or smashing. If you select a different action (the verb), your "intention" will be different for the same object (the noun). When an "adverb" that describes the action is different, what will be different in your mind? "She gently crushed the tomato in her hand." "She violently crushed the tomato in her hand". What will exist in the time when her fingers touched the tomato peel and in the space where the pulp of the tomato splashed?*

Figure 1. Examples of shown movies

Figure 2 shows the system architecture of our work. It consists of a finger-tracking system, a server program and a client program that shows movies. The finger-tracking system sends the recognized result to the server program as an event, and the server program forwards the event to the client program. The client program developed with Macromedia Director selects the movie to show according to the recognized result.

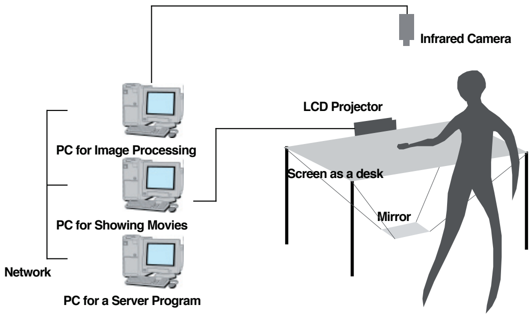

Figure 2. System architecture

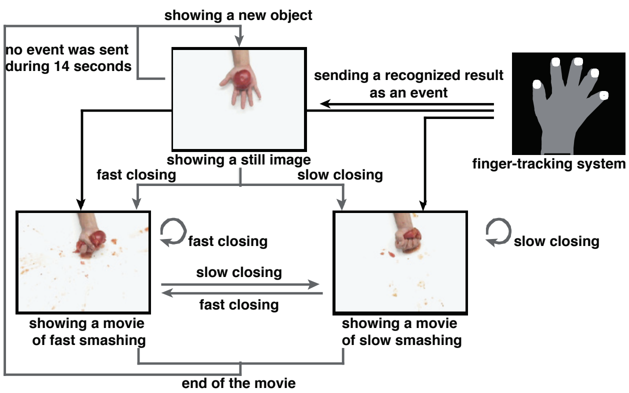

Figure 3. System flow

Figure 3 shows the flow of our work. Our vision-based finger-tracking technique is used for determining the speed of the closing of a hand. At the same time that a user's hand closes, a trigger signal and the speed of closing of the hand is sent to a PC, where movies of crashing objects are projected. When the replaying of a movie is finished, the system selects a new object to show at random, and the still image of that object is shown. Because the client program changes the frame number of a movie to replay when a finger recognition system sends a recognized result, it begins to replay a movie according to the speed of the user's hand closing in the midst of the replaying of that movie. Therefore, users are also able to smash the same object again and again at different speeds.

3. Finger Tracking

With using a infrared camera, image regions which correspond to human skin can be easily identified by binarization of the input image with a proper threshold value. Then, for the purpose of fast search of multiple fingertips, a search window of a fixed size is set so that it includes a hand part of the arm region based on the orientation of the arm. Once a search window has been determined for a hand region, fingertips are searched for within that window. The overall shape of a human finger can be approximated by a cylinder with a hemispherical cap. Thus, we use normalized correlation with a template of a circle with the proper size for detecting fingertips, and the locations of fingertips are detected. Then, our finger tracking system measures trajectories of multiple fingertips by taking correspondences of fingertips detected in each image frame between successive image frames with utilizing Karman filter [3] (Figure 4). Our finger-tracking system is fast (30frames/sec), and we can recognize not only gestures of hands but also speeds that gestures were done. In this work, we utilized the transition time that the number of recognized fingertips changed from five to zero as the speed of closing a hand. When that time is longer than the defined threshold value, the system recognized that a user closed his/her hand slowly. Otherwise, it recognized that he/she closed his/her hand fast.

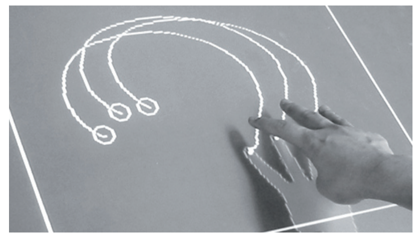

Figure4. Tracking locations and trajectories of fingertips

4. Conclusion

Many visitors experienced our work in SIGGRAPH2001. Some people enjoyed smashing various objects, and others enjoyed smashing the same objects at different speeds again and again. Many previous works have utilized image-processing techniques; however, their speed was not very fast. The visitors who tried our system of our work would have taken in the possibility of the human-computer interaction realized by fast image processing. Our work was selected as one of nominated works at 5th media arts festival held by the Agency for Cultural Affairs of Japan, and the movie of our work was exhibited at the Tokyo Metropolitan Museum of Photography in February 2002.

Reference

[1] Koike, H,, Sato, Y., and Kobayashi, Y.: Integrating paper and digital information on EnhacedDesk: a method for real-time finger tracking on augmented desk system, ACM Trans. on Computer-Human Interaction, Vol.8, Issue.4, pp.307-322 (2001).

[2] Chen, X., Koike, H., Nakanishi, Y., Oka K., and Sato, Y: Two-handed drawing on augmented desk system, Proc. 2002 Int'l Working Conf. Advanced Visual Interfaces (AVI 2002) (2002).

[3] Oka, K., Sato, Y., and Koike, H.: Real-time tracking of multiple fingertips and gesture recognition for augmented desk interface systems, Proc. of 2002 IEEE Int'l Conf. Automatic Face and Gesture Recognition (FG 2002), pp.429-434 (2002).