

# INTERACTION FOR ENTERTAINMENT CONTENTS BASED ON DIRECT MANIPULATION WITH BARE HANDS

Kenji Oka\*  
Imari Sato\*  
Yasuto Nakanishi\*\*  
Yoichi Sato\*  
Hideki Koike\*\*

\* *Institute of Industrial Science, The University of Tokyo*

{oka, imarik, ysato}@is.u-tokyo.ac.jp

\*\* *Graduate School of Information Systems, University of Electro-Communications*

{naka, koike}@is.uec.ac.jp

**Abstract** We have developed an augmented desk interface system called the EnhancedDesk which allows us to interact with both physical objects and digital information projected onto a desk simultaneously; this interaction is accomplished by direct manipulation with our hands. During our technical demonstration of EnhancedDesk, a large number of visitors actually tried various applications developed for our system. In this paper, we describe those applications and the lessons that we learned during that demonstration.

**Keywords:** augmented desk interface system, direct manipulation, hand/finger tracking

## 1. Introduction

We have developed an augmented desk interface system called the EnhancedDesk [1]. While the design of EnhancedDesk was inspired by the pioneering work by Wellner, i.e., DigitalDesk [2], our augmented desk interface system has a distinct advantage in that it allows users, with their own hands and fingers, to perform various kinds of tasks by manipulating both physical objects and electronically displayed objects simultaneously for more natural and intuitive interaction [3].

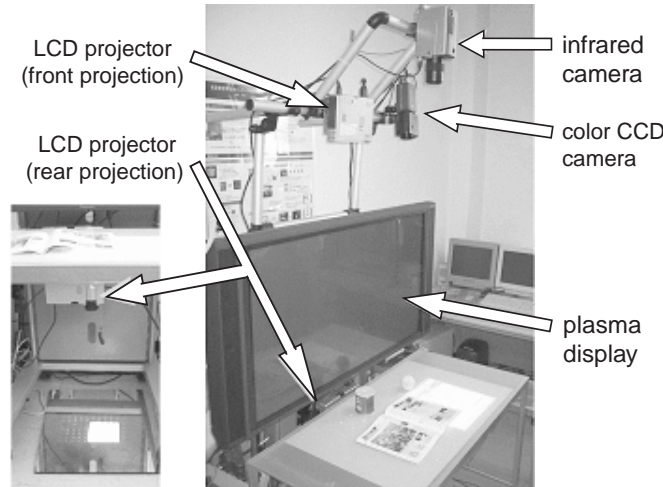


Figure 1. Overview of EnhancedDesk

We had conducted preliminary user studies of our augmented desk interface system in a laboratory setting. However, we had never had a chance to have a large number of users actually try our system outside our laboratory. During the past summer, EnhancedDesk was selected as one of the exhibits at Emerging Technologies of SIGGRAPH2001; more than 2000 visitors tried our system during a period of six days.

In this paper, we describe the details of EnhancedDesk, and four different applications that were presented at the demonstration. In addition, we report some lessons that we learned from this experience.

## 2. Description of EnhancedDesk

EnhancedDesk provides a user an intelligent environment that automatically retrieves and displays digital information corresponding to real objects, e.g., books, on a desk. The system also provides users the ability to directly manipulate digital information by using their own hands and fingers for more natural and intuitive interaction. The setup of EnhancedDesk is shown in Figure 1.

The most important technique of EnhancedDesk is real-time tracking of a user's hands and fingertips, which allows a user to simultaneously manipulate both physical objects and electrically projected objects on a desk by using natural hand gestures. In most of the previously proposed methods, image regions corresponding to human skin are typically extracted either by color segmentation or by background image subtrac-



Figure 2. Tracking of hands and fingertips

tion [5]. However, it is difficult to use those methods in our augmented desk interface system, because observed color of human skin and image backgrounds continuously changes due to projection by an LCD projector. To avoid this difficulty, we make use of an infrared camera. By adjusting the infrared camera to measure a range of temperatures that approximates human body temperature, e.g.,  $30^{\circ}\text{C} \sim 34^{\circ}\text{C}$ , values of image pixels corresponding to human skin become higher than other image pixels. Therefore, image regions which correspond to human skin can be easily identified by binarization of the input image with a proper threshold value even in complex backgrounds and under different lighting conditions.

Once a region of a user's arm is located in an input image, fingertips are searched for within that region. We make use of the observation that the projected shape of a fingertip in an input image appears to be a semi-circle. Based on this observation, our method uses normalized correlation with a template of a circle of the proper size for fast detection of multiple fingertips.

For the purpose of measuring motion of each fingertip, we, in addition, consider correspondences of detected fingertips between successive input image frames. First, locations of multiple fingertips are predicted based on a filtering technique; then, correspondences between the predicted locations and detected fingertips are examined. Our method enables us to determine trajectories of multiple fingertips reliably in real time as shown in Figure 2. Our method can also detect the center of a palm by applying morphological operations to each input image frame.

Based on the obtained tracking results, necessary information is displayed in three ways with EnhancedDesk. It has a half-transparent tabletop panel and a plasma display, and two LCD projectors, which

are used for displaying various objects by front projection and back projection onto the desktop. Since the rear-projection LCD projector does not cast shadows of the obstacles, e.g., user's hands and head, onto the desktop, various kinds of information such as web pages are displayed using it. However, objects on the desk may occlude the rear-projection display, particularly in a situation where many objects are placed on the desk. For this reason, we designed EnhancedDesk so that certain types of information are displayed by using the front-projection display.

### 3. Applications of EnhancedDesk

Four different applications, named F.Force, StickyFace, Narrative-Hand, and Rush&Crash, were newly prepared for the demonstration at SIGGRAPH2001. We developed these four different applications with several important aspects of our interface system in mind. In particular, we aimed to develop applications which emphasize the following three aspects.

- Intuitive interaction based on direct manipulation and symbolic gestures with users' own hands
- Integration between real objects and virtually projected objects
- Accuracy and reliability of our vision-based technique for tracking multiple hands and fingertips in an uncontrolled environment

#### 3.1. F.Force

The demonstration called F.Force, shown in Figure 3, illustrates the speed, accuracy, and reliability of our vision-based technique for tracking multiple fingertips of a user's hand.

When there is no hand over the desk, many small particles with random walking are projected onto the desk. As soon as a user's hand is placed on the desk, the system starts tracking the fingertips of the hand, and the projected particles are attracted to the tracked fingertips to form strings of bright spheres which show the trajectories of the fingertips.

#### 3.2. StickyFace

Intuitive interaction based on direct manipulation with both left and right hands is illustrated in StickyFace.

In this demonstration, users can grasp and stretch computer graphics with their own hands. Figure 4 shows such interaction with the face of the Mona Lisa, where a user can distort the face in a very interesting way.

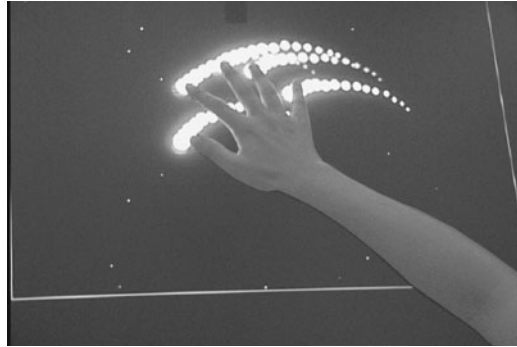


Figure 3. F.Force



Figure 4. StickyFace

This application is based on a demonstration program by Silicon Graphics, Inc. (SGI), in which a user can pick up and stretch graphics by clicking and dragging a mouse. In StickyFace, clicking is replaced by making a fist, i.e., closing a hand to create a situation where no fingertips can be detected, and dragging is replaced by movement of a hand which is in the fist position. Therefore, users can feel as if they are really manipulating the objects.

### 3.3. NarrativeHand

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 (Figure 5). For example, objects are squeezed gently when users close their own hands slowly. In contrast, when they close them quickly,



Figure 5. NarrativeHand

objects are smashed violently. Although the displayed hand is not a user's own hand, the user somehow feels a strange identification with the virtual hand.

Our vision-based tracking technique is used for determining the speed of the closing of a hand. At the same time that a user's hand is closed, a trigger signal and the speed of closing a hand is sent to a PC, where movies of crashing objects are projected.

### 3.4. Rush&Crash

In Rush&Crash, when users make drawing gestures such as a circle or a triangle, the meanings of the gestures are recognized and synthetic objects corresponding to the gestures are generated. Users can bounce those synthetic objects off their own hands in the same way as in a hockey game.

The other focus of that demonstration is interaction based on integration of both real objects and digitally projected objects. For example, a number of synthetic objects also bounce off real objects placed on the desk (Figure 6).

## 4. Results and discussion

While we did not conduct a formal user study during the exhibition, we obtained valuable comments and feedback from the many visitors who tried our applications with EnhancedDesk. In addition, we analyzed a video recording of our demonstration to examine users' reactions after the exhibition. In this section, we describe some lessons that we learned from these experiences.

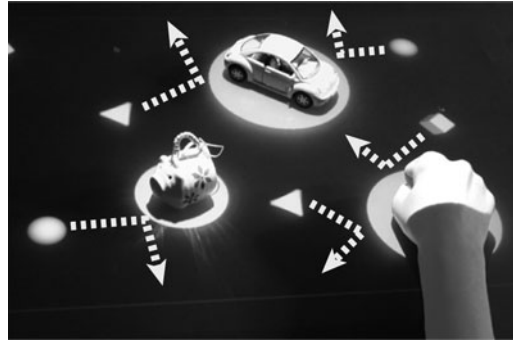


Figure 6. Ruch&Crash

Regarding direct manipulation based on users' own hands, we received mostly positive comments from participants. Most of them found direct manipulation by hands in EnhancedDesk to be natural and intuitive, and they preferred it to conventional input devices such as a mouse. This was most obvious in the case of StickyFace.

On the other hand, some participants found direct manipulation based on hands somewhat awkward for Rush&Crash. Based on our observation of users' reactions, we concluded that this awkwardness is due to the lack of haptic feedback onto users' hands. It is interesting that very few participants complained about the lack of haptic feedback for StickyFace, and that most of the participants were satisfied with only auditory and visual feedback. In contrast, the lack of haptic feedback was more frequently pointed out in the case of Rush&Crash. We consider that this tendency is closely related to the focus of users' attention. In StickyFace, participants enjoyed unexpected variations of the Mona Lisa's face, and therefore little attention was paid to the fact that haptic feedback was missing. In this sense, interaction in Rush&Crash was much simpler: patting virtual objects with a hand. Thus participants easily noticed the lack of haptic feedback which they normally expect from patting real objects with their hands.

Regarding the performance of our vision-based technique for tracking multiple fingertips in real time, the system never failed during the six-day exhibition period. It worked reliably for a large number of users with different sized hands and different ways of moving those hands. It should be noted that this reliability is important in order to use our vision-based technique for actual HCI applications. On the other hand, we found that it was necessary to improve the reliability of our tracking technique to handle users with hands too cold to enable reliable extraction of hand

regions in input images from the infrared camera. Interestingly, this seemed to happen more often to female participants.

Besides the lessons described so far, we acquired a number of lessons through demonstrations. One of them was the difference in hand gestures among different cultural backgrounds. For example, in Narrative-Hand, we defined the squeezing gesture as closing a hand with the thumb folded inside the palm. However, a number of participants made this gesture with their thumbs outside their palms. This difference sometimes caused errors in our gesture recognition system. We had never expected this problem until this demonstration, and it indicates that we should consider cultural variation of gestures in order to ensure the reliability of gesture recognition systems.

## References

- [1] H. Koike, Y. Sato, Y. Kobayashi, H. Tobita and M. Kobayashi, "Interactive textbook and interactive Venn diagram: natural and intuitive interface on augmented desk system," *Proc. ACM Conf. Human Factors in Computing Systems (SIGCHI 2000)*, pp. 121-128, 2000.
- [2] P. Wellner, "Interacting with paper on the DigitalDesk," *Communications of the ACM*, Vol. 36, No. 7, pp. 87-96, 1993.
- [3] T. Nishi, Y. Sato and H. Koike, "SnapLink: interactive object registration and recognition for augmented desk interface," *Proc. IFIP Int'l Conf. Human-Computer Interaction (Interact '01)*, M. Hirose (Ed.), IOS Press, pp. 240-246, 2001.
- [4] K. Oka, Y. Sato and K. Koike, "Real-time tracking of multiple fingertips and gesture recognition for augmented desk interface systems," to appear in *Proc. IEEE Int'l Conf. Automatic Face and Gesture Recognition (FG '02)*, 2002.
- [5] V. Pavlovic, R. Sharma and T. Huang, "Visual interpretation of hand gestures for human-computer interaction: a review," *IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI)*, Vol. 19, No. 7, pp. 677-695, 1997.