Proc. 2001 IFIP Conference on Human-Computer Interaction (Interact 2001), pp. 240-246, July 2001.

SnapLink: Interactive Object Registration and Recognition for Augmented Desk Interface

Takahiro Nishi*, Yoichi Sato+ & Hideki Koike*

*Department of Information Systems, University of Electro Communications

+Institute of Industrial Science, The University of Tokyo

{nishi, koike}@vogue.is.uec.ac.jp, ysato@iis.u-tokyo.ac.jp

Abstract: Identification of objects in a real world plays a key role for human-computer interaction in a computer-augmented environment using augmented reality techniques. To provide natural and intuitive interaction in such environments, it is necessary for an interface system to know which objects a user is using. In previously developed interface systems, real objects are identified by using specially designed tags attached to objects. In this work, we propose a new method for interactive object recognition and registration; the method permits more natural and intuitive interaction without using any tags. In particular, we introduce interactive object registration and recognition by combining direct manipulation by the user's hands with a color-based object recognition algorithm.

Keywords: human computer interaction, augmented reality, computer vision, object recognition, color histogram

1 Introduction

Graphical user interface (GUI) is commonly used as a standard interface on personal computers. GUI is well-matured, and it provides an efficient interface for a user to employ various kinds of applications on a computer. However, many users find that the capability of GUI is rather limited when they try to perform some tasks by combining physical objects such as paper documents on a desk and computer applications. This limitation comes from the lack of seamless integration between the two types of entities, i.e., physical objects and computer applications.

One of the earliest attempts to provide seamless integration of physical objects and associated computer applications was reported in Wellner's DigitalDesk (Wellner, 1993). In this work, the use of a desk equipped with a CCD camera and a video projector was introduced, and several tasks such as an electrical sketchpad and a calculator integrated with physical paper documents were described. However, interaction using the DigitalDesk was rather limited largely due to its insufficient sensing capability for tracking user's activities and recognizing real objects on a desk.

We have been studying an augmented desk interface system called the EnhancedDesk, which allows a user to perform various kinds of deskwork efficiently by using both computer applications and physical objects such as textbooks (Koike et al, 2000). While the design of our augmented desk interface system was inspired by the pioneering work by Wellner, the EnhancedDesk provides users more intuitive interaction by allowing them to use their own hands for direct and fine manipulation of both physical and projected objects. Unlike the DigitalDesk, the EnhancedDesk uses a fast and robust sensing technique for monitoring users' activities in real-time, particularly motions of hands and fingertips in uncontrolled environments such as ordinary office rooms (Sato et al, 2000).

On the other hand, the capability of recognizing real objects on the desk is rather limited in the EnhancedDesk, which is dependent on a specially designed 2D marker attached onto a real object to determine which object a user is using. As a result, a user is required to place a code onto a real object whenever a new object is introduced. Moreover, it proved to be impossible to put a code onto some objects due to their shape and size.

To overcome this limitation, we propose a new method for interactive object registration and recognition for augmented-desk interface systems (Figure1). The goal of the method is to provide a user a natural and intuitive way to incorporate real objects into an augmented desk interface system without being required to use tags attached to the objects.

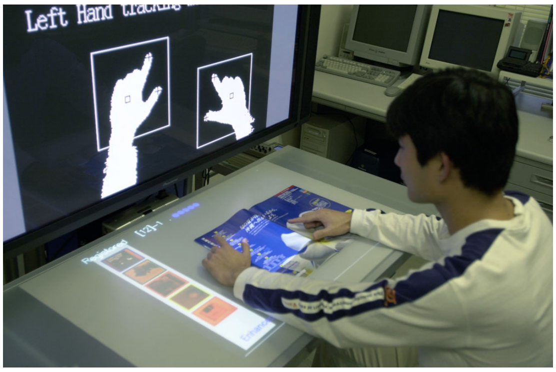

Figure 1: Interactive object registration and recognition in our augmented desk interface system

This paper is organized as follows. In Section 2, we review previous approaches for recognition of real objects for computer-augmented environments. In Section3, we propose interactive registration and recognition of real objects for an augmented desk interface system. Finally, in Section 4, we present our conclusions and discuss future research directions.

2 Registration and Recognition of Real Objects

For human computer interaction in a computer-augmented environment, it is essential for an interface system to have the ability to recognize real objects in the environment.

2.1 The use of ID tags

There are several techniques currently available for identifying real objects by using various kinds of ID tags attached to those objects. One of the most widely used examples of such techniques is based on radio frequency identification (RFID) tags. A RFID tag is a small badge equipped with electronics for radio frequency communication. RFID tags are widely used for industrial applications such as management control of parts distribution in factories.

RFID tags have been used by several researchers as a key component for human-computer interaction in a computer-augmented environment. For example, Want et al. proposed the use of RFID tags as a tool for bridging the physical world and a virtual world in their interface system (Want et al, 1999).

Some researchers have proposed other approaches for object recognition by using a marker and a video camera (Ishii et al., 1997; Rekimoto, 1997; Rekimoto et al, 1999). A marker with a specially designed pattern is attached to a real objects. Then, using computer vision techniques, the marker is identified in an input image obtained by using a video camera. This is considered as an extension of the barcode and barcode scanner that are widely used in many applications. Unlike the barcode/barcode scanner method, a user does not have to touch an object with a marker to scan the marker, but rather simply captures an image of the object by means of a video camera. Thus, the motion of a user is less restricted. This is considered to be suitable for providing a user-friendly interaction in a computer-augmented environment. These techniques for identifying objects with RFID tags or 2D markers are effective for certain types of applications. With such ID tags, objects can be identified even if multiple objects have identical appearances. Rekimoto proposed using a 2D matrix code for real-time object recognition for augmented reality (Rekimoto et al, 1998); this technique was also used for Navicam, which was proposed by Rekimoto and his colleagues (Rekimoto, 1997).

However, techniques based on ID tags suffer from a common limitation when they are applied to human-computer interaction. Markers need to be attached to all objects beforehand in order to identify those objects. This might not be a problem in some situations, but it becomes infeasible if users are required to place an ID tag to all objects that they use in a computer-augmented environment. In addition, it may be impossible or difficult to attach any ID tags to some objects due to their shape or size.

2.2 The use of image features

In the computer vision and image processing community, a number of object recognition algorithms have been proposed; these algorithms are based on various types of basic features extracted from input images. For instance, object recognition based on image features has been the main focus in the field of data retrieval for multi-media database systems. Typically, geometric features such as lines, points, and corners are extracted from an input image, and then the identity of an object seen in the input image is determined based on those extracted geometrical features (Ballard et al, 1982). Unfortunately, those techniques tend to be computationally expensive, and therefore less suitable for real-time applications.

Another interesting group of techniques for object recognition has been proposed recently in the computer vision community. The appearance of an object is recorded by observing the object under different viewing and illumination conditions; then the input image is directly compared with these recorded training images of the object. This group of techniques is called the appearance-based method because the appearance of an object is directly compared with input images. Unfortunately, this computation is generally too expensive to be used for real-time interaction. To overcome this problem, Murase and Nayar proposed an efficient method to perform this comparison; their method is based on principal component analysis of input images (Murase et al, 1995), and they have demonstrated that 100 objects can be recognized in real-time (Nayar et al, 1996). A similar technique based on vector quantization of binarized images was proposed by Krumm (Krumm, 1997), and his technique was later used for registering and recognizing real objects in a computer-augmented environment (Krumm et al, 2000).

Appearance-based methods for object recognition have been shown to be effective for recognizing a large number of objects in real-time. However, those techniques are not suitable for our application of interactive registration and recognition in an augmented desk interface where all training images of objects to be registered have to be collected and processed as a batch process beforehand.

Color information is also used for recognizing objects. A histogram of color image pixel values of an object is used as a model of object appearance. Then the object is identified by comparing a color histogram computed from an input image and color histograms of registered objects. Techniques based on color information have the advantage that color histograms are translation and rotation invariant. Also, those techniques are relatively insensitive to undesirable effects such as partial occlusion, distortion, and change in image backgrounds. Swain and Ballard demonstrated the use of color histograms for recognizing a number of objects (Swain et al, 1991); their technique was called *the color indexing algorithm*.

Unlike the appearance-based methods, object recognition methods based on color information use relatively simple representations as object models, e.g., a color histogram. This is particularly advantageous when real-time processing is necessary. For this reason, we decided to use an object recognition method based on color, in particular, the color indexing method proposed by Swain and Ballard in (Swain et al, 1991). Before we describe the interactive object registration and recognition framework used for our augmented desk interface, we briefly review the color indexing algorithm in the following subsection.

2.3 Color indexing method

An image of an object to be registered is captured as a reference image. Then each pixel is voted to a color space, e.g., a RGB color space, according to its color value. After all image pixels of the reference image are voted in the color space, the color histogram of the registered object is obtained as the resulting distribution in the color space. Let M_i' be a color histogram of a reference image where the index i represents each entry in the RGB color space, and the number of entries are n^3 if each axis of the color space is partitioned in n steps $(i = 1,..,n^3)$. The sum of values of all entries $\sum_i M_i'$ becomes equal to the number of image pixels in the reference image. Since we would like to treat reference images and input images with different sizes, color histograms need to be normalized by $M_i = Mi' / \sum_j M_j$ so that $\sum_{i=1}^{n^3} M_i = 1$.

The similarity between the color histogram of a reference image M and that of an input image I is defined as an intersection between those two color histograms. To be more precise, the similarity between those two color histograms is given as $\sum_i \min(I_i, M_i)$. If those two histograms are exactly the same, the value becomes 1, and if the reference image and the input image does not share any color, the value becomes 0.

When an input image is obtained, the color histogram of the input image is compared to all color histograms of reference images. If the largest intersection value is more than predetermined threshold, the object contained in the corresponding reference image is considered to be found in the input image.

3 Interactive Object Registration and Recognition Object for EnhancedDesk

In the EnhancedDesk reported in (Koike et al, 2000), real objects on a tabletop were identified by using a 2D matrix code attached to objects. This use of 2D matrix codes resulted in the same limitation described in Section 2.1. Users were required to place a code onto a real object whenever a new object was registered. Also, it was even impossible to put a code to certain objects, and therefore, those objects could not be integrated into the EnhancedDesk system.

This motivated us to seek another approach for recognizing real objects and for realizing intuitive interaction with our augmented desk interface system. In particular, a user should be able to register a real object with various kinds of information associated with that object without using any ID tags. Also, a registered object should be identified easily by the interface system when a user asks the system to obtain information associated with the object.

In this work, we propose a new method for interactive object registration and recognition for human computer interaction in computer-augmented environments. The key feature of the proposed method is that objects are registered and recognized with active participation of a user. Instead of trying to develop a computer vision technique for fully automatic registration and recognition of real objects, we designed our system in such that users can indicate interactively with their hands which object should be registered and where an object should be looked for. This is particularly favorable in an augmented desk interface system with the need for real-time processing and intuitive interaction. By having a user indicate the location of an object, the interface system does not have to solve the problem of object localization, i.e., the object's location, but it does have to solve the problem of object identification, i.e., which object it is. Thus, computational cost for object registration and recognition can be reduced significantly to achieve real-time processing performance with an ordinary PC.

In addition, the framework of interactive object registration and recognition can be effectively used with computer vision algorithms for object recognition. Even with significant advances in such algorithms, it is not realistic to expect any object recognition algorithm to work without any failures. However, it is a reasonable assumption that an object recognition algorithm can show a user a set of possible answers for object recognition. Then a user can select a correct answer from candidates shown by the system.

In the next section, we present an overview of our augmented desk interface system called the EnhancedDesk (Koike, 2000). Then we explain our proposed method for interactive object registration and recognition in the EnhancedDesk.

3.1 EnhancedDesk

The 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.

As shown in Figure 2, the EnhancedDesk is equipped with an infrared camera, a color video camera, two LCD projectors, and a plasma display. The color video camera of the EnhancedDesk is mounted so that it observes the desktop. Input color images are used for object registration and recognition based on the color indexing algorithm.

Real-time tracking of user's hands and fingertips is an essential part of the EnhancedDesk with which a user can simultaneously manipulate both physical objects and electrically projected objects on a desk by using natural hand gestures. In real-time, the system tracks positions of the centers and the fingertips of both right and left hands of a user (Sato et al, 2000). An infrared camera is used for reliable detection of the user's hands even when various kinds of information are displayed onto 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. Figure 3 shows examples of hand and fingertip tracking.

The EnhancedDesk displays information in three ways; it has a half-transparent tabletop panel, and two LCD projectors are used for displaying various objects onto the desk from its front and its back.

Since the EnhancedDesk uses object recognition algorithms based on color information, the rear-projection LCD projector is mainly used for information display. By using rear-projection display, observed color of real objects is not affected by information display. For instance, various kinds of information such as web pages are displayed using the rear-projection LCD projector.

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 the EnhancedDesk so that certain types of information are displayed by using the front-projection display. In addition, the front-projection display is used for illuminating real objects when those objects are registered and recognized, thereby contributing to increased robustness of the color indexing algorithm.

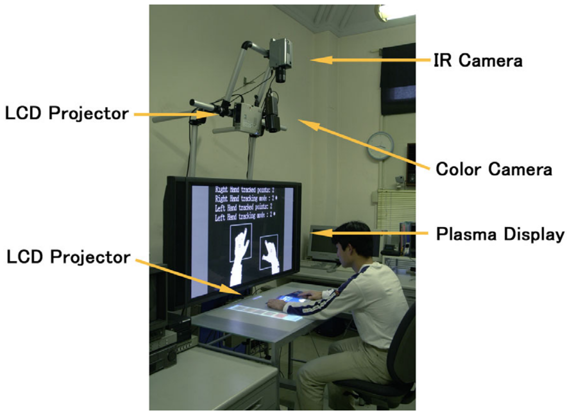

Figure 2: Overview of the EnhancedDesk

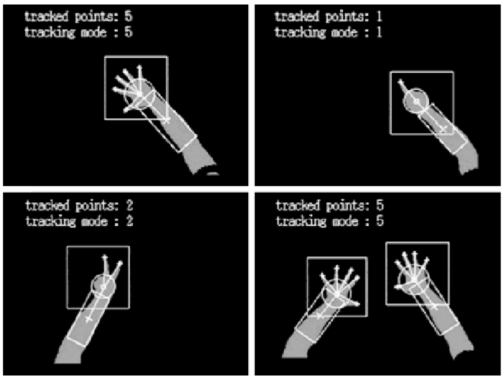

Figure 3: Example of tracking results

3.2 Object registration with snap-shot gesture

In our current design of the EnhancedDesk, an object is registered based on its appearance when a user indicates a region on the desk by making a snapshot gesture with four fingers as shown in Figure 4. When a user makes the gesture, a rectangle of a fixed size, e.g., 60x60 pixels, is overlaid onto the specified region to indicate to the user which portion on the desk is going to be registered. We have found this overlay display of a focused region helpful for users to adjust the portion for object registration. Figure 5 shows one example of overlay display for object registration when a user makes a snapshot gesture. The region to be registered is illuminated with a bright rectangle by using the front-projection LCD projector, and users can adjust the position of the region simply by moving their hands. After users hold the same hand gesture for a certain duration of time, they hear the sound of a mechanical shutter, indicating that the region has been registered by the system.

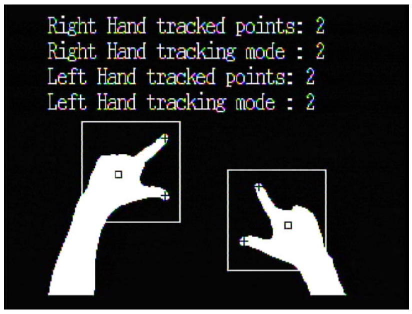

Figure 4: Recognition of snapshot gesture

Figure 5: Registration of object based on color histogram

When a user registers an object by using a hand gesture, the region specified by the user is recorded as a reference image. Then, the color histogram of the reference image is generated as a model for the object. To avoid the undesirable effect caused by the background of the image, the background subtraction is used with a background image that is automatically captured by the system occasionally. Figure 6 shows an example of a cropped image specified by a user and its resulting image after background subtraction is then used for creating a color histogram as an object model.

The color histogram can be recorded with any kind of information associated with the object. For instance, a user can register a real object and the EnhancedDesk records the address of a web page that the user is viewing. The important point is that a user can create links between real objects and associated information dynamically with simple and intuitive hand gestures. Our proposed method for interactive object registration and recognition provides a user such a tool for natural and intuitive interaction in a computer-augmented environment.

Figure 6: Reference image after background subtraction

3.3 Object recognition with pointing gesture

When a user wishes to retrieve information associated with a real object on a desk, a user simply points to the object with a single finger. If only one fingertip is detected for a certain duration of time, the system interprets the action as a trigger for object recognition and illuminates the pointed region with a bright rectangle displayed by using the front-projection LCD projector. If the user keeps the same hand gesture, the illuminated region is captured for object recognition based on the color indexing algorithm. A color histogram is generated from the illuminated region after background subtraction. Then the obtained color histogram is compared with each of the registered object models. Finally, the object model with the highest matching score is considered to be present in the region pointed to by the user.

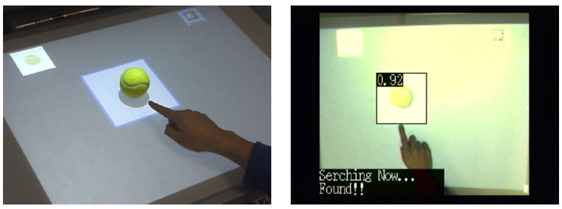

Figure 7: Object recognition by pointing gesture

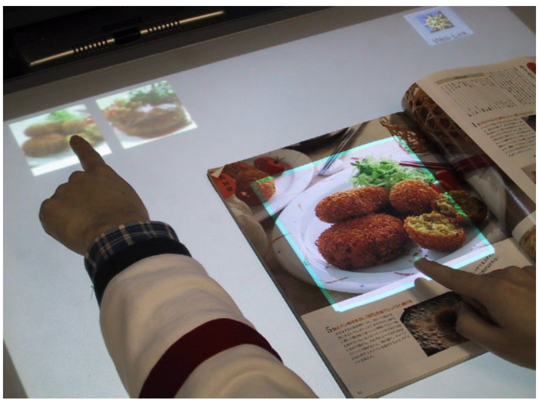

Figure 8: Selection of correct object by a user

However, as described at the beginning of Section 3, a wrong object may have the highest matching score accidentally due to various reasons such as imperfect background subtraction, partial occlusion, specular highlights on an object, and so on.

To overcome this problem, the system effectively uses feedback from the user instead of trying to return a single answer for a matched object. The EnhancedDesk shows a user several candidates for a correct object by displaying recorded images of registered objects that have a matching score higher than some threshold value (Figure 8). Then the user selects the correct object by simply pointing with a single finger of the other hand. Once the correct object is selected, the EnhancedDesk displays information associated with the object by using the rear-projection LCD projector.

4 Conclusions

In this paper, we have introduced a new method for interactive object registration and recognition for more natural and intuitive interaction in a computer-augmented environment. In particular, we described interactive object registration and recognition by direct manipulation with users' hands for our augmented desk interface called the EnhancedDesk.

Unlike other interface systems based on augmented reality technologies, our system does not require any ID tags to be attached onto real objects. Appearances of real objects are represented by using their color histograms. When a user inquires the identity of an object on a tabletop by pointing with a single finger, a correct object is obtained by using the color indexing histogram algorithm. To cope with imperfect performance of the object recognition algorithm, several candidates for a correct match are shown to a user by the system; from these, the user can select a desired object by pointing with his or her finger.

In our current implementation of the EnhancedDesk, over several dozen objects can be registered and recognized by using a user's hand gestures. We are currently extending our color-based object recognition algorithm to increase robustness and accuracy of recognition.

One of the interesting areas of or further research for human computer interaction in computer-augmented environments is the problem of establishing links between real objects in a physical world and associated information of various types.

In our current prototype applications with the EnhancedDesk, only limited types of information such as web page addresses are used. With the proposed framework of interactive object registration and recognition, we are planning to investigate further the problem of creation and management of links between real objects and associated information.

References

Ballard, D. H. & Brown, C. M. (1982), *Computer Vision*, Prentice-Hall, Inc., New Jersey

Ishii, H. & Ullmer, B. (1997), Tangible bits: Towards seam-less interface between people, bits and atoms, *In Proceeding of ACM Conference on Human Factors in Computing System (CHI'97)*, pp.234-241.

Koike, H., Sato, Y., Kobayashi, Y., Tobita, H. & Kobayashi, M. (2000), Interactive textbook and interactive Venn Diagram: natural and intuitive interfaces on augmented desk system, *In Proceedings of ACM Human Factors in Computing Systems (SIGCHI 2000)*, pp. 121-128.

Krumm, J. (1997), Object Detection with Vector Quantized Binary Features, *In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'97)*, pp. 179-185.

Campbell, R. & Krumm, J. (2000), Object Recognition for and Intelligent Room, *In Proceedings of IEEE Computer Vision and Pattern Recognition (CVPR2000),* pp.691-697.

Murase, H. & Nayar, S. K. (1995), Visual Learning and Recognition of 3-D Objects from Appearance, *In Proceedings of International Journal of Computer Vision*, Vol. 14, pp.5-24.

Nayar, S. K., Nene, S. A. & Murase, H. (1996), Real-Time 100 Object Recognition System, *In Proceedings of IEEE International Conference on Robotics and Automation (ICRA'96)*, pp.1223-1227.

Rekimoto, J. (1997), NaviCam: A Magnifying Glass Approach to Augmented Reality Systems, *Presence: Teleoperators and Virtual Environments*, Vol. 6, No. 4 pp.399-412

Rekimoto, J. (1998), Matrix: A Realtime Object Identification and Registration Method for Augmented Reality, *In Proceeding of 1998 Asia-Pacific Computer Human Interaction (APCHI'98)*.

Rekimoto, J. & Saito, M. (1999), Augmented surface: a spatially continuous work space for hybrid computing environments, *In Proceeding of ACM Conference on Human Factor in Computing System (CHI'99*), pp.378-385.

Sato, Y., Kobayashi, Y. & Koike, H. (2000), Fast tracking of hands and fingertips in infrared images for augmented desk interface, *In Proceedings of 2000 IEEE International Conference on Automatic Face and Gesture Recognition (FG2000)*, pp. 462-467.

Swain, M. J. & Ballard, D. H. (1991), Color Indexing, *International Journal of Computer Vision*, vol. 7, pp. 11-32.

Want, R., Fishkin, K. P., Gujar, A. & Harrison, B. L. (1999), Bridging Physical and Virtual Worlds with Electronic Tags, *In Proceedings of ACM Conference on Human Factors in Computing System (CHI '99)*, pp.370-377.

Wellner, P. (1993), Interacting with paper on the DIGITAL DESK, *Communication Of The ACM*, vol. 36, No. 7, spp.87-96.