Table-Top Interface Using Fingernail Images and Real Object Recognition

Kenta Hara[1], Noriko Takemura[1], Yoshio Iwai[2], and Kosuke Sato[1]

[1] Graduate School of Engineering Science, Osaka University, Machikaneyama 1-3, Toyonaka, Osaka 560-8531, Japan
[2] Graduate School of Engineering, Tottori University, 4-101, Koyamachominami, Tottori-shi, Tottori, Japan 680-8552
{hara,takemura,sato}@sens.sys.es.osaka-u.ac.jp, iwai@ike.tottori-u.ac.jp

Abstract. Many researchers have proposed the development of table-top interfaces in the last decade. In a table-top system, for user satisfaction, they must be able to operate digital and analog media seamlessly. In this paper, we propose a table-top interface system that allows users to intuitively operate digital media by gesture recognition. The proposed system can capture the image of an object placed on the table by recognizing pressing gestures from fingernail images, and can transfer digital content by recognizing user's shaking gestures. The evaluation experiments show that users can intuitively operate the proposed system without being aware of the data transmission.

Keywords: table-top interface, object recognition, gesture recognition, fingernail image.

1 Introduction

Recently, many researchers have proposed the development of table-top interfaces and electronic book (e-book) devices, especially to support the wide distribution of e-book among the generic public. However, when a user simultaneously interacts with digital and physical media through an e-book device, complicated operations of data transmission to and from the e-book device adversely affect user convenience and satisfaction. This is because of the difference in operations between the physical space and the digital space. For example, when a user handles an object such as a document on the table-top display that shows digital content, a burdensome operation of transmission, such as scanning, prevents the user from intuitively operating the digital media as if it were physical media.

Ways of data exchange between users differ in digital and physical media. They can place physical documents on a table and pick them up easily; however, in order to exchange digital data, users have to transmit them by sending an e-mail or using some type of storage media such as USB flash drives and SD cards. In order to bridge the gap in data transmission between digital and physical media, a system that provides intuitive operations for data transmission is required.

D. Keyson et al. (Eds.): AmI 2011, LNCS 7040, pp. 21–30, 2011.

In this paper, we propose a system that allows users to interact with digital content by using physical gestures that will help in modeling the operations of document handling on a table. The proposed system recognizes a pressing gesture captured from a user's fingernail image by a high-mounted camera, of which the user is unaware.The system also recognizes shaking and tapping gestures from sensors incorporated in an e-book device that trigger data transmission between digital and physical media.

2 Related Studies

2.1 Gesture Recognition

A touch panel, such as a table-top system, has been conventionally used for integrating input and output spaces, and several other systems that use touch panel have been proposed. Sugita et al. proposed a method for detection of contact on a surface and estimation of the direction of fingertip pressure by detecting a change of fingernail color when the fingertip touches a hard object[1]. When the user's finger pressed a surface, the fingernail color turned white and the brightness of the fingernail image varied significantly. Therefore, whether the finger stayed – in contact or not – was determined by the variance threshold. The pressing direction was estimated by detecting the origin of pressure change on the fingertips. Although this method enables touch sensing everywhere, there is a hardware restriction because this method requires a high-resolution fingertip image. Enhanced Desk, proposed by Oka et al., recognized user gestures by processing images using a camera[2]. However, this system did not attempt to recognize whether the user touched the surface. Han et al. proposed an input method based on frustrated total internal reaction (FTIR) phenomenon[3]. This system not only supported multi touch gestures but could also be applied to larger equipments. However, it could not detect gestures when a hand was not in contact with the surface of this system. Currently, touch panels are certainly inexpensive and already available in the general consumer market. However, their sensing area is limited to the surface of the display.

In our system, fingernail image analysis is adopted to detect not only positions but also states of fingers. By limiting the system to pressing gesture recognition and by introducing a novel method for estimating fingertip states, the proposed system can recognize states from low-resolution images. Thus, high-resolution images are not required; a camera can be mounted at a high position so that the user is not aware of it. Moreover, this method enlarges the sensing area of the surface of not only the display but also the object placed on a table.

2.2 Table-Top Systems

When users work in an environment where physical and digital media are combined, they prefer to use an intuitive interface that maps a movement in the real world to digital operations. Various systems that attempt to achieve this capability have been proposed(e.g., [4], [5]). These systems projected digital documents and images on real-world objects or used a table as a display which allowed users to operate them with their fingers.

Hartmann et al. proposed a system that supported collaborative work by capturing images of real objects placed on a desk and projecting it[6]. When an object was recognized, digital handles were projected around the corners of the object; the user could capture the object's image onto the table by dragging the handle with his or her finger. However, this system required a complicated hardware setting, and it did not support a more intuitive operation of pointing to the object itself rather than that to the digital handle. Snap Table, proposed by Koshimizu et al., enabled users to transfer digital documents or images displayed on a table to an e-paper device by using physical gestures[7]. A photo-addressable e-paper was used so that digital media could be immediately transferred to the e-paper device. However, this system required prior access to digital documents, because it did not support capturing data from real media.

As mentioned above, our proposed system can recognize pressed states of fingertips, such that the user can directly press down a real object placed on a table into a display that is digital space. Moreover, this system provides a novel feedback to users, which is not provided by the conventional method. Moreover, as an extension of the operations performed with documents placed on the table, this system allows users to transfer digital media between the table and an e-book device by using gestures. Through these processes, the proposed system enables users to interact more seamlessly with physical and digital space.

3 System Overview

3.1 Process Flow

Fig. 1. shows the process flowchart of the proposed system. Images of a table-top display are captured by a camera mounted high over the table. The system estimates the positions of an object placed on a table, a user's fingertip, and an e-book device. From the captured images, the system also estimates the state of the fingertip by recognizing pressing gestures and the relationship of the e-book device and the object on the table. Using this information, the system interacts with the user by updating the table-top display and communicating with the e-book device.

3.2 Pressing Gesture Recognition

First, the system extracts the skin region H from the captured image. In this paper, we assume that region H is regarded as the region of the hand. Fingertip region T is determined from region H. Finger positions are defined as the centers of each fingertip region T, and sub images of each fingertip region T are normalized into an image H_f with a size of 10×10 [pixels]. The fingertip image H_f is defined as a 300-dimensional feature vector (10×10 [pixels]$\times$RGB). The state of a fingertip is determined by a linear discriminant function that is generated by the support vector machine (SVM) (Fig. 2). In the SVM-based method,

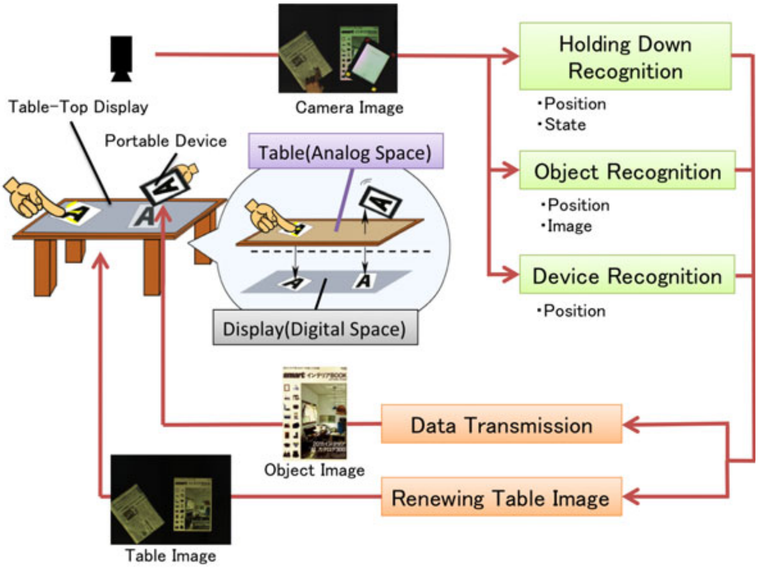


Fig. 1. Process flow of proposed system

the maximum-margin hyperplane can be obtained and suitable generalization performance can be achieved.

T is determined by scanning region H in raster-scan order using an $n \times n$ window R as follows:

$$T = \left\{ (x, y) \middle| [(x, y) \in H] \wedge \left[t_{min} < \frac{1}{n^2} \sum_{p=x-\frac{n}{2}}^{x+\frac{n}{2}-1} \sum_{q=y-\frac{n}{2}}^{y+\frac{n}{2}-1} g(p, q) < t_{max} \right] \right\} \quad (1)$$

where t_{min} and t_{max} are determined empirically and $g(p, q)$ is defined as

$$g(p, q) = \begin{cases} 1 & \text{if } (p, q) \in H, \\ 0 & \text{if } (p, q) \notin H. \end{cases} \quad (2)$$

The number of skin-color pixels in window R is counted in real time because the system uses an integral image technique[8].

(a) Non-pressed

(b) Pressed

Fig. 2. Change in fingernail color with pressing gesture

3.3 Table Image Update

When an object is placed on a table, the region of the object is extracted from an input color image by background subtraction using V-value in the HSV color space. When the difference between the centroid points of the region in the current frame and those in the previous frame is smaller than a certain threshold, the captured region is determined to be the object region. The system stores the captured image in the storage for future use.

When the system is in this state, the user can select an object from the ones placed on the table by pressing the object on the table surface. The pressing gesture is recognized from H_f described in the previous section. In order to reduce errors in recognition, the system recognizes the user's pressing gesture only when the discrimination function recognizes pressing gestures during a certain time t_{step}. When the system recognizes the user's pressing gesture, the system updates the table-top image by using the captured object image described in the previous section. In this way, the proposed system uses an intuitive operation that allows an object placed on the table to be captured from the table-top by using a pressing gesture.

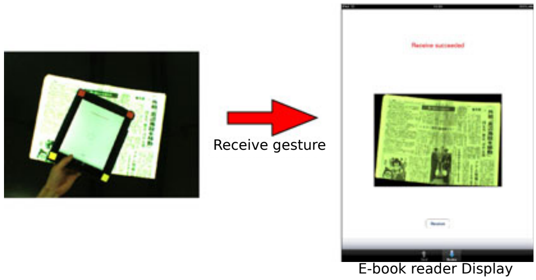
3.4 Bidirectional Transmission between Table and E-Book Device

Four color markers are attached at each corner of an e-book device in order to estimate the position of the e-book device from a captured image. The e-book device is detected when the distance and angle between same color markers are smaller than certain threshold values. This process is required to distinguish the device from objects on the table. When the e-book device is detected, the position of the e-book device, P_{pad}, is estimated from the geometry of the markers.

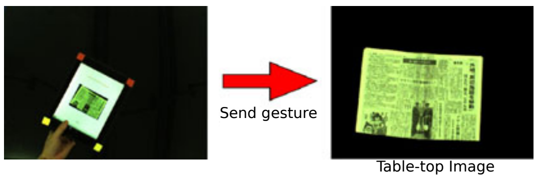
After the position estimation, the system starts a session to communicate between the e-book device and the table. The communication application in the e-book device has two modes – receive and send – and these modes get activated by shaking or tapping gestures. A shaking gesture is recognized by using an accelerometer in the e-book device, while a tapping gesture is recognized by using a touch sensor in the e-book device.

During the receive mode, the user selects an object from the table-top image and holds the e-book device over the object while performing a shaking gesture. When the system detects the e-book device and estimates its position P_{pad}, it retrieves an object image corresponding to P_{pad} from the storage. When the e-book device recognizes a shaking gesture or a touching gesture, it receives the retrieved object image from the table-top system through a wireless network and displays the image (Fig. 3 (a)).

During the send mode, the user selects an image on the e-book device that has to be sent and then performs a shaking gesture over any position on the table. When the e-book device recognizes the shaking gesture, it sends the image to the table-top system through a wireless network. When the table-top system receives the image, it displays the image at position P_{pad} (Fig. 3 (b)).



(a)



(b)

Fig. 3. Bidirectional transmission

4 Evaluation

4.1 Experimental Setup

Fig. 4. shows the configuration of the proposed system. Since the camera is installed at a height of two meters, images of the entire display can be captured, which enables users to interact easily with any part of the display. The table-top interface is developed by attaching a clear acrylic plate onto the display. The resolution of the camera is 800×600 [pixels] and the frame rate is 15 [fps]. The Apple iPad is used as the e-book device.

Fig. 4. Overview of experimental system

4.2 Experiments of Pressing Gesture Recognition

In order to confirm the effectiveness of our discrimination method, we conducted experiments with 14 subjects and compared our method to the conventional method[1]. Nine thousand fingernail images were obtained as training data from eight men and women in their 20s. Tables 1. and 2. show the recognition rate of subjects whose data was gathered during training and that of the other subjects, respectively. These tables indicate that our method performs better than the conventional method. The recognition rate of subjects whose data was obtained during training is 90.2%, which confirms the effectiveness of our method.

We also examined the relationship between the gesture recognition rate and the time needed for recognition t_{step}(Fig. 5.). A pressing gesture and a non-pressing gesture were performed for 30 [s] by five of the subjects whose data was used for training, and these gestures were recognized with each t_{step} varied from 0 [s] to 5 [s] at intervals of 0.25 [s]. The recognition rate varies with changes in t_{step}, and when t_{step} is 1 [s], up to 98% recognition rate was obtained.

Table 1. Recognition rate(trained subjects)

	Press	Non-press	Average
Proposed	**92.8%**	**87.7%**	**90.2%**
Conventional	73.2%	87.1%	80.2%

Table 2. Recognition rate(untrained subjects)

	Press	Non-press	Average
Proposed	**88.8%**	79.3%	**84.1%**
Conventional	72.9%	**87.4%**	80.2%

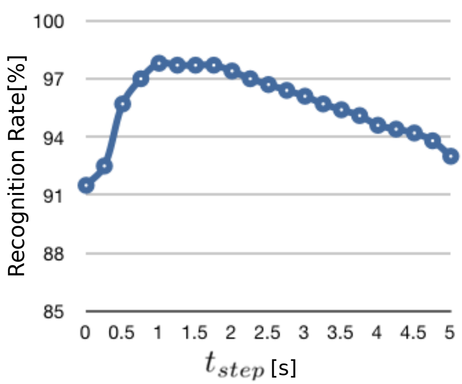


Fig. 5. Recognition rate of each t_{step}

4.3 Evaluation of System Usability

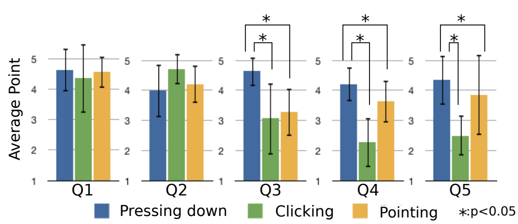
In these experiments, 14 subjects performed two tasks: capturing objects and transferring contents between the table and the iPad. The system was evaluated subjectively by a questionnaire.

In the operation of capturing objects, the subjects captured several objects sequentially. We compared three methods used to capture objects : (1) capturing by pressing the table (the proposed method), (2) pointing to out of reach objects, and (3) using a mouse to click on an object image displayed on an additional window. The subjects performed the task using these three methods. In the operation of transferring content between the table and the iPad, the subjects transferred the captured image on the display to the iPad and then transferred the transferred image in the iPad to the table. We employed a tapping gesture on the display of the iPad and a shaking gesture of the iPad as gestures for triggering the transfer. The subjects performed both gestures in this experiment.

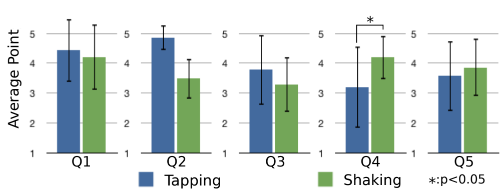
After the tasks were completed, the system was evaluated subjectively with the following questionnaire for each of the two tasks. The subjects selected a number along a scale, like "1" for very dissatisfied and "5" for very satisfied.

Q1. Was it easy to remember how to perform the operations?
Q2. Could you perform the operations as wanted to?
Q3. Do you feel that the operating action was natural?
Q4. Could you perform the operations without detecting digital processing?
Q5. How attractive was the operation you performed?

According to the results, the responses to Q4 are significantly different between the methods in terms of both capture operation and communication operation (Fig. 6.). When the subjects answered the questions about the capture operation, the responses to Q1 and Q2 were high in all the methods and there was not much difference between them. The evaluations of Q3 and Q5 in the proposed method were higher compared to the others, and there was a 5% level of significant difference. When the subjects answered the questions about communication, the shaking gesture was evaluated higher than the button operation in Q4 responses. And, responses to Q5 are higher than their median in both.



(a) Evaluation for the capture operation



(b) Evaluation for the communication operation

Fig. 6. Questionnaire responses for system usability

5 Discussion

From the experimental results in Section 4.2, we observed that the proposed system is advantageous over the previous system because the recognition rate of the proposed system is higher than that of the previous system regardless of the fact that the data were collected during training. In the previous system, the recognition of a pressing gesture was performed using a fingertip image that had a size of 20×20 [pixels]. Although the image size was 10×10 [pixels] in the proposed system, the proposed system shows better recognition performance than the previous system. Moreover, the results show that the proposed system can even handle small fingertip images.

As shown in Table 2., the recognition rate for untrained subjects is 84.1%, which is lower than that for trained subjects, but it is higher than that of the previous method. For some subjects, however, the recognition rate is very low because there was no fingernail color variation in the fingertip images of these subjects, which is considered as a different situation in the training data.

Because the accuracy of recognition of pressing gestures depends on the training data, the accuracy is degraded when the color variation is very different from the training data. In real-life situations, the number of users of a desk is limited. When the users are identified in advance, training data are collected from them and the recognition rate reaches 90%, as shown in Table 1. Fig. 5. shows an improvement in the recognition rate, which occurred by changing the duration time t_{step}; when t_{step} is set to around 1 [s], the recognition rate reaches 98%.

Pointing and pressing gestures have the same performance results as mouse operations for typical computer applications because the evaluation of an operation in the proposed system is equal to that in the conventional systems, as demonstrated from the results in Q1 and Q2 in Section 4.3. The recognition rate of pressing gestures is large enough to satisfy the user and does not prevent operability. As shown in Fig. 6. (a), the capturing of an object using a pressing gesture is similar to a user's mental model that an object is pressed onto the table by pushing the object. The users can perform an operation without discomfort and the system is evaluated highly in Q5 for the above reason.

The responses for the communication operation are significantly different with 5% error level, as shown in Fig. 6. (b), and the shaking gesture is evaluated highly because the user can seamlessly perform the operation without being aware of any digital process. The use of gestures is effective in concealing digital processes. The evaluations of both shaking and pressing gestures are attractive, as shown by the results of Q5, and the method used to select an object by holding the iPad over the object is naturally accepted. Three of the subjects failed to hold the iPad over a real object before the system captured an image of the object. From this observation, we realized that some subjects did not differentiate between real objects and object images during the experimental tasks. This phenomenon fulfilled our final goal of this study, which is filling the usability gap between real objects and digital data.

6 Conclusion

In this paper, we have proposed an interface in which mental models in the real world can be applied to an operation against digital data in an environment consisting of a table-top display and an e-book device.

In the proposed system, digital media and a real object can be dealt with simultaneously without a feeling of unfamiliarity. The system allows the user to interact with digital media using specific predefined gestures such as pressing a finger on a table-top display and shaking gestures of a device. In this system, we proposed a method for recognizing a pressing gestures on the basis of a fingernail image with a size of 10×10 [pixels]. We built a prototype system and confirmed that our recognition method is effective and that our interaction method does not cause users to feel unfamiliar or uncomfortable.

Improving the recognition accuracy and speed are our future goals. We also aim to improve the method for achieving a more comfortable interaction.

References

1. Sugita, N., Iwai, D., Sato, K.: Touch Sensing by Image Analysis of Fingernail. In: SICE Annual Conference, pp. 1520–1525 (2008)
2. Koike, H., Sato, Y., Kobayashi, Y.: Integrating Paper and Digital Information on EnhancedDesk: A Method for Realtime Finger Tracking on a Augmented Desk System. ACM Transactions on Computer-Human Interaction 8(4) (December 2001)
3. Han, J.Y.: Low-cost Multi-touch Sensing through Frustrated Total Internal Reflection. In: Symposium on User Interface Software and Technology, pp. 115–118 (2005)
4. Kane, S.K., Avrahami, D., Wobbrock, J.O., Harrison, B., Rea, A., Philipose, M., LaMarca, A.: Bonfire: A Nomadic System for Hybrid Laptop-Tabletop Interaction. In: Symposium on User Interface Software and Technology, pp. 129–138 (2009)
5. Iwai, D., Sato, K.: Limpid Desk: See-Through Access to Disorderly Desktop in Projection-Based Mixed Reality. In: Symposium on Virtual Reality Software and Technology, pp. 112–115 (2006)
6. Hartmann, B., Moriss, M.R., Benko, H., Wilson, A.D.: Pictionaire: Supporting Collaborative Design Work by Integrating Physical and Digital Artifacts. In: Conference on Computer Supported Cooperative Work, pp. 421–424 (2010)
7. Koshimizu, M., Hayashi, N., Hirose, Y.: SnapTable: Physical Handling for Digital Documents for Electronic Paper. In: The Third Nordic Conference on Human Computer, pp. 401–404 (2004)
8. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 511–518 (2001)