I'm Here! : a Wearable Object Remembrance Support System

Takahiro Ueoka, Tatsuyuki Kawamura, Yasuyuki Kono, and Masatsugu Kidode

Graduate School of Information Science, Nara Institute of Science and Technology
8916-5 Takayama, Ikoma, Nara 630-0192 Japan
{taka-ue, tatsu-k, kono, kidode}@is.aist-nara.ac.jp

Abstract. In this paper we propose a wearable vision interface system named "I'm Here!" to support a user's remembrance of object location in everyday life. The system enables users to retrieve certain information from a video database that has recorded a set of the latest scenes of target objects which were held by the user and were observed from the users' viewpoint. We propose the object recognition method to associate the video database with the name of objects observed in the video. The offline experiments demonstrate that the system is useful enough to recognize the objects.

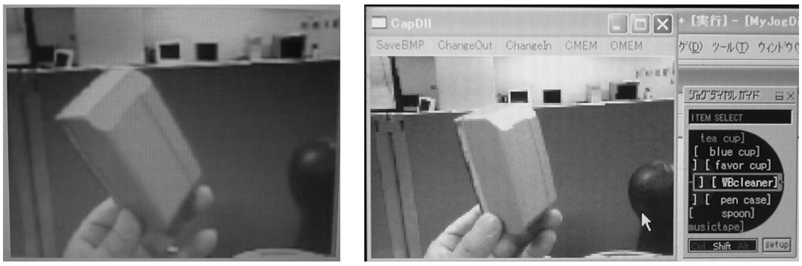
1 Introduction

We propose a wearable vision system to help the user remember where portable, rigid objects are placed in his/her everyday circumstances. This system retrieves the last recorded video of a user's viewpoint including the target object from a video database termed "video memory".

In the field of wearable computing for memory-aid, a Video Albuming system [1] is proposed to support a user's memory retrieval. The Video Albuming system always records a video of the user's viewpoint in his/her video memory, and analyses context information included in the video memory. The "I'm Here!" system follows the concept of the Video Albuming system, and focuses on the idea that users forget where they placed objects. DyPERS [2] assists users by reminding them of information about an object. This is accomplished by replaying a video explicitly associated with a snapshot of the object. In contrast, our system can retrieve the video without a user's explicit association, because the "I'm Here!" system automatically associates the video memory with the previously registered objects. The registration of an object is done by simply holding the object and gazing at it. Hide and Seek [3] is a system that navigates with the frequency of sound how far an object is placed from the user. The small devices attached to the target objects are each assigned a unique ID. In the "I'm Here!" system, a user simply has to wear some devices and there is no need to place such devices in the real world.

2 Design Concepts of the System

The “I’m Here!” system shows a user the latest video recorded when he/she lastly held a target object through a head-mounted display. Viewing the video that was observed by his/her head-mounted camera, he/she can remember where and when he/she placed the object. Ultimately we expect that the system will act as if the object itself sends a message, such as “I’m Here!” to the user.



(a) A video of user’s viewpoint at object registration

(b) The system interface and a displayed video at object retrieval

Fig. 1. Support for object remembrance using video memory

As depicted in figure 1(a), the user registers an object by holding and rotating it in view of his/her head-mounted camera. The system extracts visual information of the object from the video and records its features with the object’s name as assigned by the user.

In the user’s everyday life, the system continuously records a video of the user’s viewpoint in a video memory. Simultaneously, the system identifies an observed object, held by the user, as the registered one. Using the result of identification, the system then associates the name of the identified object with the video and automatically constructs an index of video memory, which is named “augmented memory.”

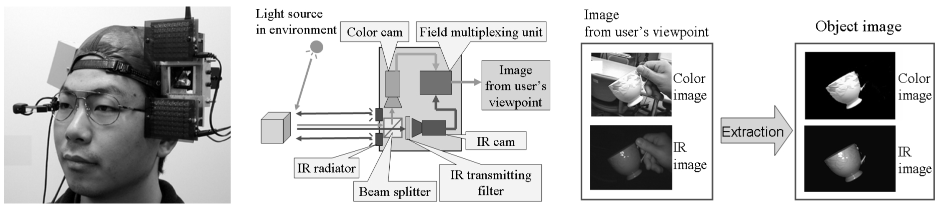
Figure 1 (b) shows a scene of object retrieval. The user, via the selective list of registered objects, assigns a name to the object so as to remember where it is placed. The retrieved video is displayed as shown in figure 1 (b).

3 System Implementation

We have developed a prototype of the “I’m Here!” system. The hardware mainly consists of a head-mounted camera called “ObjectCam”, a wearable PC, and a head-mounted display (HMD). The ObjectCam captures the user’s viewpoint image. The PC executes the processes of the system. The HMD shows the user system information and the retrieved video.

3.1 Camera and Display Devices

Figure 2(a) denotes head-mounted camera and display devices. We have newly developed an “ObjectCam,” which is a head-mounted combined camera device,



(a) Head-mounted devices (b) Block diagram of the ObjectCam (c) Extracting an object image

Fig. 2. Newly developed camera and display device

to extract an object image from a user's viewpoint image. A frame of the image consists of a color and an infra-red (IR) image for each field (figure 2(b)). An IR image displays the reflected IR luminance caused by the IR light source on the devices [4]. The system obtains the object image by eliminating background regions from the viewpoint image with the luminance of the IR image, and hand regions by using skin color (figure 2(c)).

3.2 System Functions

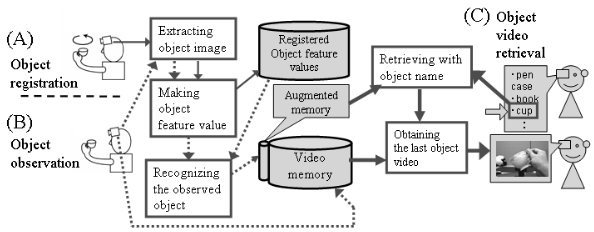


Fig. 3. System functional diagram

Figure 3 shows both the system function provided to the user and the process diagram of each function. In "object registration" (figure 3(A)) the system records a video of the object held and manipulated by the user. The object is observed in several appearances from the user's viewpoint in everyday life. The system thus extracts the images of the object in several appearances from the video memory. The image groups are made from the extracted object images, based on the appearances of the objects. The system constructs the feature values from representative images of each group.

In "object observation" (figure 3(B)) the system abstracts the object feature values from the user's viewpoint images just as in the case of figure 3(A). Comparing the target object with registered objects, the system makes a decision based on their feature values. The system records the user's viewpoint image, and labels the image with the registered object name if the target object is recognized as one of the registered objects.

When the user wants to remember where one of registered object is, he/she selects the target object from the list of registered objects (figure 3(C)). The system retrieves the name of the target object from the augmented memory, and obtains the last recorded video of the target object. Lastly, the system displays the retrieved video to users through the HMD.

3.3 Object Recognition

A $\{H\text{-}Z\text{-}C\}$ feature value of an object image consists of $\{H, Z, C\}$ elements. These elements are obtained from each pixel of the image. $\{H\}$ is a hue value. $\{Z\}$ is an IR luminance value. $\{C\}$ is a group of pixels divided by the distance from the median point of the silhouette of the object image. The $\{H\text{-}Z\text{-}C\}$ feature value denotes a 3D distribution. The equation 1 and 2 show the $\{H\text{-}Z\text{-}C\}$ feature value expressed by Integrated Probabilistic Histogram value (IPH) in the case of $j \in H, k \in Z,$ and $l \in C$.

$$IPH(j,k,l) = \sum_i N\big(j - H_i, \sigma_H(S_i)\big) \times N\big(k - Z_i, \sigma_Z(S_i)\big) \times C_{il} \tag{1}$$

$$C_{il} = \begin{cases} 1 : l = \left[\dfrac{L_i}{L}\right] \\ 0 : l \neq \left[\dfrac{L_i}{L}\right] \qquad L : const. \end{cases} \tag{2}$$

When $\{H_i, Z_i, S_i\}$ means a value of the ith pixel, $N(j - H_i, \sigma_H)$ and $N(k - Z_i, \sigma_Z)$ are normal distribution functions. $\sigma_H(S_i)$ and $\sigma_Z(S_i)$ denote dispersion values of distribution based on the exponential decreasing function including S_i in the exponent part. The element L_i denotes the distance between the median point of the silhouette of the object image and the ith pixel.

The similarity between two $\{H\text{-}Z\text{-}C\}$ feature values is calculated by the Sum of Absolute Difference (SAD) method. The system compares a feature value of an observed object with feature values of registered objects to identify the observed object. The system then selects the most similar object from the entire group of registered objects when the similarity shows higher than the preset threshold.

4 Experimental Results

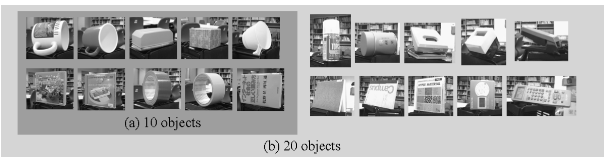


Fig. 4. Test objects

Table 1. Recognition rate

	Object feature value			
	H	H-C	H-Z	H-Z-C
(a) 10 objects (%)	99.2	99.2	91.7	96.7
(b) 20 objects (%)	81.7	88.8	87.5	94.1

We have estimated the recognition performance of the system through an offline experiment. The test images were recorded under fluorescent lights. As depicted in Figure 4, Test group (a) consists of the images of ten objects, and test group (b) consists of twenty objects including the objects in (a), and an additional ten objects. The images of an object consist of twenty different configurations of distance and perspective.

We compared the $\{H\text{-}Z\text{-}C\}$ method with the $\{H\}$, $\{H\text{-}C\}$ or $\{H\text{-}Z\}$ method in terms of recognition rate. The result of recognition allows a matching between patterns in the same object. Table 1 shows results of the experiment. The recognition rate using the $\{H\text{-}Z\text{-}C\}$ method is the highest among the methods. Furthermore, the decrease in the recognition rate caused by the increase of the number of test objects is shown, but the decrease using the $\{H\text{-}Z\text{-}C\}$ method is significantly smaller than other methods.

From the result, we found that the $\{H\text{-}Z\text{-}C\}$ method is more robust when the number of registered objects increases. We believe that the proposed $\{H\text{-}Z\text{-}C\}$ method is appropriate for object recognition in "I'm Here!"

5 Concluding Remarks

In this paper we proposed the "I'm Here!" system, which is a wearable interface system for remembering where an object used in everyday life was placed. The performance of object recognition directly affects the performance of the system in its adequacy of support.

The proposed "I'm Here!" system supports retrieving only the objects placed by the user himself/herself. We are planning to apply the system to support the case where the objects have been moved and placed by others. In everyday life, a human sequentially handles objects to perform a task. For instance, when he/she wants to have a cup of coffee, he/she prepares his/her cup, boils water in a kettle, and stirs the coffee with a spoon. We are also extending the "I'm Here!" system to recognize the sequence of accesses to objects in a task to suggest to the user what objects should be used and where they are placed.

Acknowledgements

This research is supported by Core Research for the Evolutional Science and Technology (CREST) Program "Advanced Media Technology for Everyday Living" of the Japan Science and Technology Corporation (JST).

References

1. Tatsuyuki Kawamura, Yasuyuki Kono, Masatsugu Kidode: Wearable Interfaces for a Video Diary: towards Memory Retrieval, Exchange, and Transportation. In Sixth International Symposium on Wearable Computers, Seattle, Washington (2002).
2. Tony jebara, Bernt Schiele, Nuria Oliver, Alex Pentland: DyPERS: Dynamic Personal Enhanced Reality System. In Proceedings of the 1998 Image Understanding Workshop, Monterrey CA, November 1998. Also appears as Vision and Modeling Technical Report #463.
3. M. Shinnishi, S. Iga, F. Higuchi, M. Yasumura: Hide and Seek: Physical Real Artifacts which Responds to the User. In Proc. of World Multiconference on Systemics, Cybernetics and Informatics (SCI'99/ISAS'99) Vol.4, pp.84-88 (1999).
4. Cameron M. Lee, Konrad E. Schroder, Eric J. Seibel: Efficient image segmentation of walking hazards using IR illumination in wearable low vision aids. In Sixth International Symposium on Wearable Computers, Seattle, Washington (2002).