Functional Evaluation of a Vision-based Object Remembrance Support System

Takahiro Ueoka Tatsuyuki Kawamura Yasuyuki Kono Masatsugu Kidode

{taka-ue, tatsu-k, kono, kidode}@is.aist-nara.ac.jp

Graduate School of Information Science, Nara Institute of Science and Technology

8916-5 Takayama, Ikoma, Nara 630-0192, Japan

Abstract

People tend to forget where they placed an object, which is necessary to achieve a certain task, in their everyday circumstances. To support a user's object-finding tasks, we have proposed a wearable interface system named "I'm Here!" The system manages Augmented Memories, a video database of the user's viewpoint labeled with previously registered objects, to display the last video of the target object held by the user. In this paper we evaluate the function of the system with laboratory experiments, and discuss the positive and negative effect of the system with the experimental results.

1 Introduction

We have proposed a wearable interface system named "*I'm Here!*" to support a user's ability to find an object which is necessary for him/her to achieve a certain task [1]. People tend to forget where they place target objects in their everyday circumstances. The *I'm Here!* system records a video of a user's viewpoint with a head-mounted camera, and analyzes the video on a wearable PC, thus showing the user the last-recorded video of the object the user seeks.

The *I'm Here!* system is one of the key modules of the Augmented Memory albuming system [2], which supports a user's everyday memory activity. The Augmented Memory albuming system manages a video memory made from the video of the user's viewpoint. The video memory is an important information media in our everyday life because it records the user's everyday activities, for example, the environment he/she was in, the action he/she performed, and the target he/she accessed. By analyzing a user's own information, the Augmented Memory albuming system constructs an Augmented Memory, which is a video database with the detected context information of the user.

Computational augmentation of human memory has been studied recently. A Wearable Remembrance Agent [3] supports a user's need to remember relevant information in his/her real-time situations by proactively providing notes on a wearable interface. The Forget-Me-Not system [4] gives the user an alternative way of retrieving information that the user may have forgotten by collecting his/her everyday activities and displaying such information as a personal biography with the use of a small portable interface. Although these studies focus on building a human memory database that stores information of a user's everyday life, the evaluation of the computational memory augmentation system in terms of contribution as a user's memory aid has not been discussed.

In this paper we study the quantitative contributions of the *I'm Here!* system to support a user's object finding task with practical experiments. We discuss the positive and negative effects of the system from the result of the experiments.

2 Proposed system overview

Figure 1: Memory retrieval with the *I'm Here!* system

The interface of the *I'm Here!* system provides three phases, i.e., object registration, object observation, and object retrieval, to support a user's object finding task. In the object registration phase, the user registers an object to the system with a simple operation, namely, by holding and gazing at the object. The system records a video of the user's viewpoint and extracts the images of the object held by the user in several appearances as its visual features. Then the user inputs the name of the object. The name and the visual features of the object are registered into an object dictionary.

The object observation phase is continuously carried out in the user's everyday life. The system continuously records the video of the user's viewpoint and extracts the appearances of the object held by the user in the video scene. Based on the result of matching the extracted appearances of the object with the features of the objects registered in the object dictionary, the system continuously constructs a video of the user's viewpoint labeled with the name of the registered object as an Augmented Memory.

Figure 1 illustrates a scene of memory retrieval employing the object retrieval phase of the *I'm Here!* system. When the user wants to remember where one of the registered objects is, he/she selects the target object from the displayed index of the object dictionary. The system retrieves the name of the selected object from the Augmented Memory and obtains the last recorded video to show to the user. Viewing the video, the user can remember where and when he/she placed the object.

Other systems have also been developed to support object finding tasks in a user's everyday circumstances. The Hide and Seek system [5] navigates with the frequency of sound how far a target object is placed from a user. Each small device attached to a target object is assigned a unique ID, and the device emits the frequency of sound in response to a user's selection with a portable controller. In the *I'm Here!* system, however, the user simply has to wear certain devices, and there is no need to place such ubiquitous devices in the real world. Additionally the *I'm Here!* system has the advantage of operating in an environment such that the user can indicate the required object anywhere. In contrast, the Hide and Seek system only allows for the situation such the user and the target object are in the same room. Kawashima et al. [6] developed a wearable episode recording system to support a user's remembrance histories of his/her everyday life. The system records the video of the user's viewpoint, identifies his/her location and action with the video data, and constructs the episode database. The user can retrieve the database with a PDA interface. Although the wearable episode recording system is similar to the *I'm Here!* system in its concept and employment, the evaluation of contribution of the system as a memory aid for the user has not been addressed.

We have conducted experiments to evaluate the contribution of the *I'm Here!* system to support a user's object finding tasks. We have also evaluated both the positive and negative effects of the system on the behavior of the user by conditioning the accuracy of object recognition. To survey the system's effects and contributions, we have configured the system and environment for the experiments by virtually setting the accuracy of object recognition.

3 Experiments and results

3.1 Conditions

The wearable system for the experiments consists of a head-mounted display with CCD camera, a wearable PC, and a JogDial interface (Figure 2). Additionally, a pedometer is loaded to count the total number of steps in an experiment of the wearer. The experimental system simulates the retrieval phase. The CCD camera captures the video of a user's viewpoint only for displaying the real-time video to the user.

The Augmented Memories which were recorded in advance are loaded into the experimental system. The Augmented Memories are configured by applying a virtual object-recognition rate which is set as one of the following percentages: $\{0\%, 33\%, 66\%, 100\%\}$. The percentages represent the assortment of the positive and negative cases included in each trial of an object-finding task. The positive case represents the case where the retrieved video displays the view of the virtual subject, the experimenter, and placing the target object in its actual place in the target-object placing task. The negative case represents the case where the video displays a view of the virtual subject placing an object somewhere other than where the object was originally and actually placed. The case where an experimental setting doesn't allow a subject to use the interface of the system is named as the without-system case.

Figure 2: Hardware for the experiments

The experiments were performed in our laboratory environment. Twenty-one preset places where the objects could possibly be placed were defined in advance. Twenty-one objects were prepared and each object was placed at a predetermined specific place. Two patterns of correspondence between the objects and the places are defined. These patterns also determine the sequence of placing the objects.

The subjects are seven male students of our laboratory who are familiar with the environment. Each subject carried out two experimental trials with the interval of a day between trials. The first and second trials for each subject have different patterns of object-place correspondence, and are common to all subjects.

3.2 Methods

A trial consists of 1) an object-placing task with a pattern of object-place correspondence, 2) memory-stressing tasks, and 3) an object-finding task. At the end of the second trial, each subject performed 4) an auxiliary data-extracting task. After a month of doing task 4, each subject is 5) interviewed.

- **Task 1: The object-placing task**

In this task, a subject wearing the virtual online version of the *I'm Here!* system places all twenty-one objects in a designated target area one by one via his shuttling between the starting point and the target place. At the starting point, the subject gets the object with information displayed on a card and a map. The card indicates the name of the object with its image, and the map indicates the location of the target place to put the object. A view image of the place for detailed annotation of the target place is attached to the map. The pedometer counts the total number of steps walked by the subject. Both the steps and the time required by the task are recorded.

- **Task 2: The memory-stressing task**

When the subject finishes task 1, he then performs the memory-stressing tasks before the finding task. The subject memorizes the location of twenty-five numbers on a 5×5 array within ten seconds. Then he sequentially clicks grids of the array in ascending order of the numbers without displaying the numbers. Throughout this task each subject adds stress to his memory.

- **Task 3: The object-finding task**

In this task, the subject seeks three objects one by one while walking around the experimental environment. Before he starts walking, the card displaying the name and an image of the target object is shown to the subject. Next, the subject is allowed to operate the interface of the system only once, i.e., he selects an object with the Jog Dial interface, if the setting of the trial is not the without-system case. The virtual online system then displays the video previously associated with the object on the head-mounted display. Positive and negative cases in a task have been controlled as the setting of the task as below:

> *In the 1st experimental trial of subject P, the object-recognition rate is virtually defined as 33%. In P's object-finding task, the retrieved videos consist of one positive case and two negative cases, i.e., in a positive case the video shows the scene of placing the target object where it has actually been placed, and in each two negative cases, the video shows the scene of placing the target object where it had not actually been placed.*

The number of steps and the time required by the subject's walk are recorded in the same way that the object-placing task was done.

• **Task 4: The auxiliary data-extracting task**

In this task, the subject performs a short and long shuttle walk. Each walk consists of a starting, turning, and stopping motion, the same as the walk in the object-placing and finding tasks. The total steps and the time required by the walk are recorded as the normal walking data for the subject.

• **Task 5: The interview task**

In this task each subject is interviewed by an experimenter about his internal states of mind during the experiments. A subject answers some questions interactively, and the questions mainly consists of A) How he memorized the place of objects in task 1, B) What he was reminded of at the time the card indicating the name and an image of the target object is displayed in task 3, C) How he felt about the reliability of the displayed video in the case of using the system, and D) How he decided on his strategy to seek the target object.

3.3 Parameters

The parameters for calculating the quantitative contribution of the *I'm Here!* system are extracted from the recorded data. To extract the parameters, we assume the subject's walking model is denoted by the following variables: 1) W is the total number of steps, 2) L is the length of the optimum route to reach the given object, and 3) T is the time required to pass over the route. We also assume a relationship among the variables. Although the steps and the length are assumed to be in linear dependence, the length and the time are assumed to be divided into two cases, one case constitutes a constant velocity walk, and the other case, a zero velocity walk, i.e., temporarily stopping during a walking task.

The equations below denote the relationship between L and W, and the relationship between T and L in the constant velocity walk.

$$L = \alpha \cdot W + \beta. \tag{1}$$

$$T = \gamma \cdot L. \tag{2}$$

α, β, and γ are the coefficient parameters for a subject as decided by the normal walking data in task 4. When the data sampled in short normal walking is denoted as $\{W_0, L_0, and T_0\}$ and that in long normal walking as $\{W_1, L_1, and T_1\}$, the coefficient parameters for the subject are denoted as below:

$$\alpha = \frac{L_1 - L_0}{W_1 - W_0}. \tag{3}$$

$$\beta = \frac{L_0 W_1 - L_1 W_0}{W_1 - W_0}. \tag{4}$$

$$\gamma = \frac{T_1 - T_0}{L_1 - L_0}. \tag{5}$$

When the sampled data in task 1 are denoted as $\{W_p, T_p\}$ and the data in task 3 as $\{W_s, T_s\}$, the length of the route the subject walked in each task is assumed using equations 1, 3 and 4 as below:

$$L_p = \alpha \cdot W_p + \beta. \tag{6}$$

$$L_s = \alpha \cdot W_s + \beta. \tag{7}$$

L_p is the full length of the route in task 1, and L_s is that in task 3.

The parameters for evaluating the contribution of the *I'm Here!* system consists of L_f, T_f and T_z. L_f is the length of additional roaming in task 3, as calculated using equations 6 and 7 as below:

$$L_f = L_s - L_p. \tag{8}$$

T_f is the time required by the additional roaming denoted using equations 2 and 5 as below:

$$T_f = \gamma \cdot L_f. \tag{9}$$

T_z is the time required by staying denoted using the result of equation 9 as below:

$$T_z = T_s - T_p - T_f. \tag{10}$$

3.4 Results

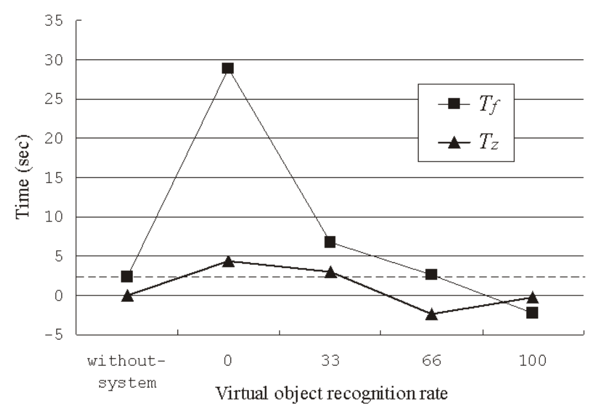

Figure 3: Contribution of the object-recognition rate to the object-finding task

Figure 3 denotes the experimental results evened off by the number of samples in each setting of the virtual object-recognition rate. Fourteen samples, which consist of the result of seven subjects each in two trials, are set as table 1. The horizontal dotted line denotes the score of T_f in the case that the subjects did not access the object retrieval function of the *I'm Here!* system. The line crosses to the result with the use of the system in about 67% of the recognition rate. The result implies that the system should embody a 67% accuracy at least.

Table 1: Configuration of the number of trials in the object-recognition rate settings

settings	without-system	0%	33%	66%	100%
number of trials	3	3	3	3	2

Table 2: Means and Standard Deviations of the T_f

cases	without-system	negative	positive
N	9	18	15
Mean	2.3	19.5	-1.3
S.D.	5.1	35.2	1.4

Table 2 denotes the mean and standard deviation (S.D.) of T_f in each group of the cases in the trials of the object-finding task. The without-system case represents the case without the support of the interface, and the positive/negative cases represent the cases of retrieved video that have true/false scenes. Via the result of ANOVA, we found the significant effect of providing the correct information of the indicated object for the time required by the subject's additional roaming in the object-finding task ($F_{(2,39)} = 3.41, p < .05$). The result of multiple comparisons based on the LSD method reveals that the mean of T_f in the negative case is significantly larger than in the positive case ($MS_e = 579.73, p < .05$). However, the result also reveals that the mean of T_f in the without-system case is not significantly different from that in the positive and negative cases.

Table 3: Means and Standard Deviations of the T_z

cases	without-system	negative	positive
N	9	18	15
Mean	-0.04	3.03	-0.29
S.D.	4.85	7.50	1.97

Table 3 denotes the mean and S.D. of T_z in each group of the trial case in the object-finding task. In contrast to T_f, no significant effect for the time required by temporarily stopping during a walking task from the result of ANOVA ($F_{(2,38)} = 1.56$) existed.

3.5 Discussion

On figure 3 we found that the *I'm Here!* system is effective with the user's object finding task when the system has more than 67% of its object-recognition rate. Additionally, there appears to be a negative effect in using the system when the accuracy of the system is below the rate. With the experimental results including interview logs, we analyzed noticeable cases where revealed worse results in the subjects' object-finding tasks with the system than the cases without the system. We found that the subjects in the worth cases had less reliability in their own memories. They blindly decided to rely on retrieved videos which actually represented negative cases. As the result of the decision, they were directed to where the target object was not placed, and roamed from there to find the target object.

On the interview logs, we found that subjects tended to evaluate the validity of a retrieved video by comparing episodes detected from the video with their own episodic memories. The episodes, for instance, if a user places a cup on a table, consist of the location of the cup on the table, other objects laid on there, appearance of the room from his/her viewpoint, a sequence of his/her activity, and so on. The episodic memories of the subjects more or less decreased with time.

We assume that a subject's reliance on his/her own episodic memories makes his/her evaluating validity of a retrieved video difficult. When a subject can not evaluate the validity, he/she has to make a choice between relying on the episodes detected from the video blindly and ignoring them. If he/she ignores them, the efficiency of his/her object-finding task will be the same as the task without the system. To make a user's evaluating validity of a retrieved video easier, extensional information, e.g. the time mark of the video, should be shown to him/her.

Additionally, we found that there were certain strategies applied to object-finding tasks by certain subjects who did not rely on the system. One of the strategies was to eliminate places where the target object would not be placed. Another strategy was to test the most possible candidate places where the target object would be placed. The enhancement of the *I'm Here!* system, e.g. displaying where the target object was often placed, will support these strategies and boost the efficiency of the user's object-finding task.

4 Concluding remarks

In this paper we presented the *I'm Here!* system and explained the detail of experiments to evaluate the contribution of the system quantitatively. The experimental results revealed the functional availability of the method for supporting a user's object-finding task if the system accurately recognizes an object held by him/her.

We discussed the negative effects shown in the results and assumed that subjects had trouble evaluating the validity of retrieved videos. If the system displays extensional information of the video, the user will be able to evaluate the validity easier, and the negative effects will be eliminated.

We plan to develop a method to display extensional information of the retrieved video to a user. We also plan to improve the *I'm Here!* system so it can be used on-line for additional experimental evaluations.

Acknowledgement

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

References

[1] T. Ueoka, T. kawamura, Y. Kono and M. Kidode, "I'm Here! : a Wearable Object Remembrance Support System," *Proc. International Symposium on Human-Computer Interaction with Mobile Devices and Services (MobileHCI2003)*, pp.422–427, 2003.

[2] T. Kawamura, Y. Kono and M. Kidode, "Wearable Interfaces for a Video Diary: towards Memory Retrieval, Exchange, and Transportation," *Proc. International Conference of Wearable Computing (ISWC2002)*, pp.31–38, 2002.

[3] R. J. Rhodes, "The Wearable Remembrance Agent: a System for Augmented Memory," *Proc. International Conference of Wearable Computing (ISWC'97)*, pp.123–128, 1997.

[4] M. Lamming and M. Flynn, "Forget-me-not: Intimate Computing in Support of Human Memory," *Proc. FRIENDS21: International Symposium on Next Generation Human Interface*, pp.125–128, 1994.

[5] M. Shinnishi, S. Iga, F. Higuchi and M. Yasumura, "Hide and Seek:Physical Real Artifacts which Respond to the User," *Proc. World Multiconference on Systemics, Cybernetics and Informatics (SCI'99/ISAS'99)*, Vol.4, pp.84–88, 1999.

[6] T. Kawashima, T. Nagasaki, M. Toda and S. Morita, "Information Summary Mechanism for Episode Recording to Support Human Activity," *Proc. International Workshop on Pattern Recognition and Understanding*, pp.49–56, Invited, 2002.