SARA: A Framework for Augmented Memory Albuming Systems

Yasuyuki KONO, Tatsuyuki KAWAMURA, Takahiro UEOKA, Satoshi MURATA, and Masatsugu KIDODE
Graduate School of Information Science, Nara Institute of Science and Technology
8916-5 Takayama, Keihanna Research Park, 630-0192, JAPAN
{kono, tatsu-k, taka-ue, satosh-m, kidode}@is.aist-nara.ac.jp

Abstract

In this paper, we discuss wearable interfaces for a computational memory-aid useful in everyday life. The aim of this research is to develop an albuming system for augmenting human memory. SARA, a framework for augmented memory systems, provides users with four types of functions: Memory Retrieval, Memory Exchange, Memory Transportation, and Memory Editing. To be equipped with various strategies of human remembrance, modules for ``Memory Retrieval'' should be easily integrated into a SARA-based system. The authors have developed the following two ``Memory Retrieval'' modules: 1) a Residual Memory system, and 2) a ``I'm Here!'' system. The Residual Memory can index a user's location automatically by analyzing a video memory recorded from his/her head-mounted camera. The ``I'm Here!'' retrieves the last recorded video including a target object from the video memory. The Ubiquitous Memories provides users with the function for associating augmented memories with real world objects for both ``Memory Transportation'' and ``Memory Exchange.'' We believe that the above interfaces can be integrated into the memory albuming system.

1. Introduction

Our research goal is to investigate methods and their computational components for augmenting human activity, especially for augmenting human memory in everyday life. Augmented memory technology has been investigated extensively in recent years[1,2,3]. Such technology makes it easy for a user to refer to video data to recall a person's information, for instance, when the person stands in front of the user[4,5]. The VAM system[5] detects a human face, which was previously recorded, and displays information of the retrieved person. In these augmented memory systems, a user wears both a head-mounted camera for continuously recording his/her viewpoint images and a head-mounted display (HMD) for viewing information given by the system, e.g., a pre-recorded video of his/her experience.

This paper proposes a framework for augmented memory albuming systems, named SARA (Sceneful Augmented Remembrance Album), where 1) a user's viewpoint images are always observed, 2) the observed images along with the data observed by other wearable sensors are analyzed to detect context, 3) the observed images are stored with the detected context as the user's augmented memories, 4) the stored data can be additionally annotated/indexed by the user for later retrieval, and 5) the user can recall his/her experiences by reviewing a stored video which is retrieved by consulting the indexes, i.e., both automatically stored and manually annotated information. We consider the augmented memory albuming to be a strong application for a wearable information playing station (WIPS)[6].

2. Architecture of SARA

2.1. Functions for Albuming Memories

Figure 1 illustrates the overall architecture of the proposed augmented memory albuming system. We believe that the following four functions are essential to realize SARA:

Memory Retrieval: A memory retrieval function provides a user with the ability for retrieving proper multimedia data from a huge multimedia database termed ``video memory.'' A human chooses his/her best remembrance strategy from various remembrance strategies and applies it to the context. We have developed the following two modules for Memory Retrieval: ``Residual Memory'' for supporting a location-based user's recollection and ``I'm Here!'' for an object-based one. ``Ubiquitous Memories'' mentioned below also works as an object-based memory retrieval module.

Memory Transportation: A memory transportation function provides a user with the ability for associating either the event he/she is facing or the augmented memory retrieved by one of memory retrieval modules with other features. Using such function the user is able to rearrange his/her memories for later retrieval. Using the Ubiquitous Memories the user can associate augmented memories with real world objects. He/she can also hold and convey the memories with the associated objects.

Memory Exchange: A user would augment his/her problem-solving ability by referring to others'

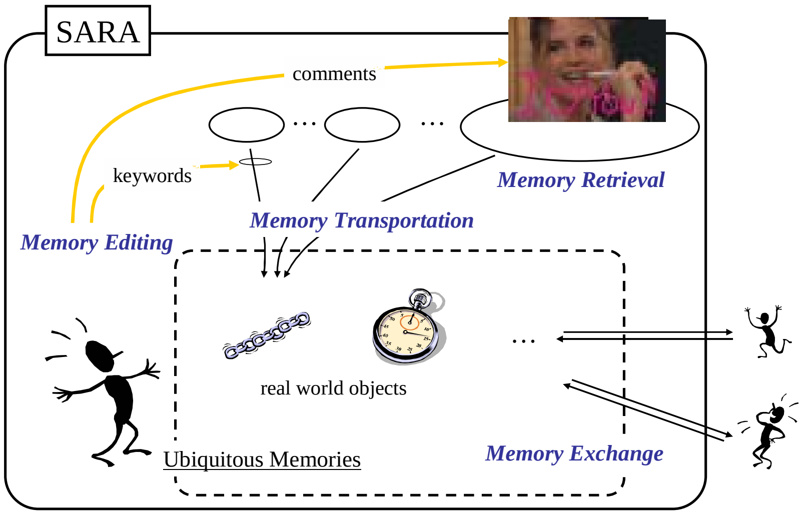


Figure 1: Basic functions for operating augmented memories on SARA

experience if it is properly associated with the given problem. The Ubiquitous Memories helps its users exchange their experience. By accessing a real world object, a user can view all the augmented memories associated with the object if the owner of each memory has approved of other users viewing it.

Memory Editing: A person is to remember the event he/she has experienced with conceptual/linguistic indexes. A memory editing function provides a user with making annotations in his/her memories by adding them keywords or free-hand comments, or by reordering them based on his/her own intention.

Utilizing these functions, the user can reuse human experience by remembering his/her own augmented memories or by viewing other users' augmented memories rearranged in a real world object.

2.2. Ubiquitous Memories

The Ubiquitous Memories is the core module of SARA framework. It is concerned with three of the four functions mentioned previously, i.e., *Memory Transportation*, *Memory Exchange*, and *Memory Retrieval*.

A person retains a relationship between real world objects and his/her memory associated with such objects[7]. If each real world object is identified, SARA can provide the user with the ability to link augmented memories to real world objects, as well as the ability to recollect his/her experience by accessing the object linked with the experience. Furthermore, the user is able to convey his/her memories with an object and exchange them with other users via the image linked to the object.

It is crucial for such a system to provide ``natural'' operations. In everyday life, a user picks up a real world object when the user is interested in that object or needs that object. ``Touching'' is a natural operation used to establish a link between the user's external activities and an operational object. So, we decided that the operational object is basically selected by the ``touching'' operation in the system. The user is equipped with a Radio Frequency Identification (RFID) device that is incorporated into a glove. Each object is implanted/attached to a RFID tag. A WWW server is employed to maintain links between objects and augmented memories. The RFID device can immediately read the RFID tag data when the device approaches the tag. The entire system connects to the World Wide Web via a wireless LAN.

Figure 2 illustrates the basic operations of the Ubiquitous Memories. It has two base states, i.e., MEMORIZE and REMEMBER. The state of the system is normally at REMEMBER. The user is allowed to select and execute one of the defined operations of the system by touching the RFID operation tag with a special ID associated with the operation, called an ``operation tag.'' Operation tags are worn on the opposite wrist of the user's gloved hand. Ubiquitous Memories allows the user to use the following two operations:

MEMORIZE: The state of the system is changed from REMEMBER to MEMORIZE when the user touches the operation tag MEMORIZE (I). The user then puts

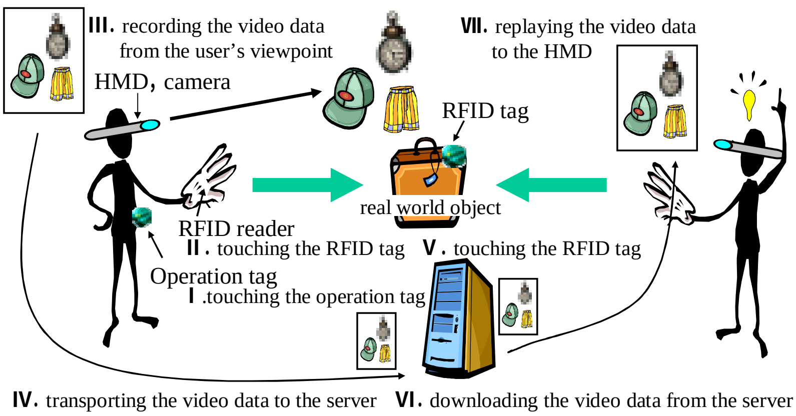


Figure 2: Operational overview of the Ubiquitous Memories

(a) Selecting an object (b) Replaying the augmented memory

Figure 3: HMD view of the operation REMEMBER

the gloved hand near an object (II). Next, the system records a video data from the head-mounted camera (III). Finally, the system connects with the web server to store the video data (augmented memory) linked with the RFID tag ID attached to the object (IV).

REMEMBER: The previous user or another user simply touches a RFID tag attached to an object when he/she wants to remember a memory (Figure 3(a)) (V). The system retrieves the augmented memory associated with the object from the web server (VI). The retrieved memory is then replayed in the top-left area of the screen of the HMD (Figure 3(b)) (VII).

The RFID devices are not essential to implement the concept of the Ubiquitous Memories, e.g., an object recognition technology as described in the next section can be applied to identify the object the user is touching. RFID devices are currently suitable for discriminating real world objects and implementing the Ubiquitous Memories, because a RFID tag does not require batteries to operate.

3. Memory Retrieval Modules

We have designed two memory retrieval functions/modules: one is a location-based function/module and another is an object-based one. Both locations and objects can be triggers for memorizing/remembering human experiences. The Ubiquitous Memories described in the previous section is also regarded as a memory retrieval module where a user retrieves augmented memories associated with an object by simply touching it.

3.1. Residual Memory

When the user arrives back at the location where he/she had been before, i.e., when the current input image from his/her head-mounted camera is identified with the stored one, the system shows the stored one on the HMD. The Residual Memory assists the user in remembering the event that happened at that location[9].

To achieve location-based memory retrieval, a system should retrieve video data, which includes similar images, recorded previously to the current input image. For instance, two images are slightly different when the hand of a user cuts into his/her view even if the posture of the head-mounted camera is stable. However, both images should be regarded as similar in terms of location. In the Residual Memory, to cope with the recognition of location-based image similarity between the input and stored images, the detection of video scenes from continuously recorded video data, and the retrieval of a required video data from large video memory with the input query images, the following methods are employed:

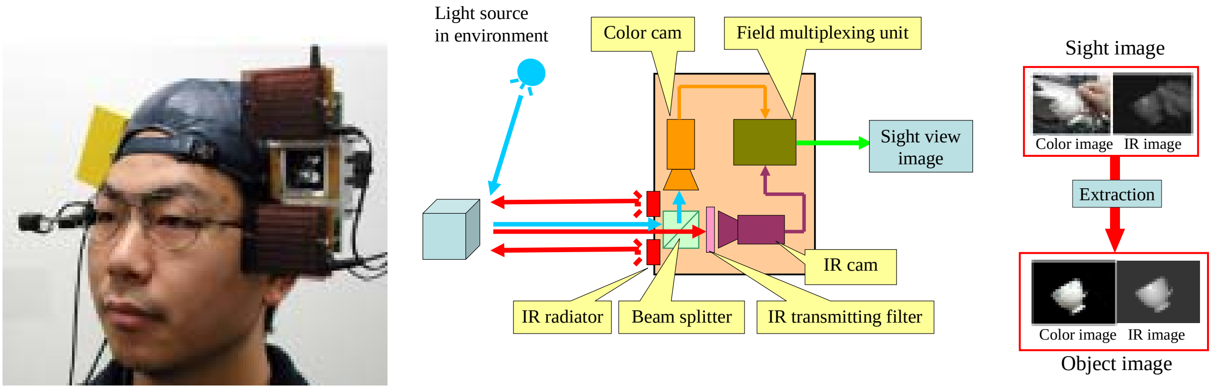
- The image differences caused by head motions and moving objects that cut into the images are corrected using gyro sensors on the head.
- The video scene differences caused by a user's activities are observed and analyzed using gyro sensors on the head.
- The associable video data are retrieved using a characteristic of user's viewpoint images continuously recorded before.

We have done an experiment to demonstrate the effectiveness of the above all methods. The experiment took place during the daytime, outdoors, in a hall, and in a room. Our method is well suited for retrieving similar location-based images because the relevance rate of the proposed method performed 1.3 times better than the ``normal'' method, which does not consider motion information.

Figure 4: HMD view of the ``I'm Here!''

3.2. I'm Here!

The ``I'm Here!'' retrieves the latest video recorded when he/she lastly held a target object. 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 tells the user ``I'm Here!''



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

Figure 5: Overview and configuration of the ObjectCam

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. The ``I'm Here!'' continuously identifies the observed object held by the user as one of the registered objects. Using the result of identification, the system then associates the name of the identified object with the observed video. Figure 4 shows a scene of object retrieval. The user assigns the name of the object he/she wants to remember where it is placed among the list of registered objects. The retrieved video recorded when he/she lastly held it is displayed.

We have developed an ``ObjectCam,'' which is a head-mounted combined camera device, to extract an object image from a user's viewpoint image (Figure 5(a)). A video frame consists of a color and an infra-red (IR) image for each field (Figure 5(b)). An IR image displays the reflected IR luminance caused by the IR light source on the device. 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 5(c)).

We employ an Integrated Probabilistic Histogram value (IPH) to represent the feature of an image of the object held and manipulated by the user. In object registration, the system records a video of the object held and manipulated by the user. The system divides the images of the object into several image groups which are made from the extracted object images, based on the appearances of the objects. The system constructs the feature value from the representative image of each group.

A {*H-Z-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, and {*C*} is a group of pixels divided by the distance from the median point of the silhouette of the object image. The {*H-Z-C*} feature value denotes a 3D distribution. The equation (1) and (2) show the {*H-Z-C*} feature value expressed by an IPH value in the case of $j \in H$, $k \in Z$, and $l \in C$.

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

$$C_{il} = \begin{cases} 1 : l = \left[\frac{L_i}{L}\right] \\ 0 : l \neq \left[\frac{L_i}{L}\right] \end{cases} \qquad L\text{: const.} \quad (2)$$

When $\{H_i, Z_i, S_i\}$ means a value of the *i*-th 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 *i*-th pixel.

The similarity between two {*H-Z-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. Integrating the Modules

A SARA-based system provides its user with functions for freely retrieving, transporting, exchanging, and editing augmented memories. By integrating the functions/modules the user can augment his/her memory capability and remembering means.

Suppose that the user participated in his/her graduation exercise in the auditorium of his/her university wearing his/her WIPS, and that a few years after graduation he/she has a chance to visit his/her alma mater. When he/she go into the auditorium, the Residual Memory retrieves the augmented memory of the graduation exercise and replays it in his/her HMD. He/she temporally associate the memory with his/her ID card he/she is porting and he/she transfers to his/her graduation certificate after he/she comes back home. After he/she stably arrange the memory by these two *Memory Transportation* operations, he/she or anyone approved by him/her, e.g., his/her university schoolmate, can review the scene at any time by touching the certificate.

5. Concluding Remarks

In this paper, we discussed the essential functions, the framework, and their technologies for enabling SARA, the Augmented Memory Albuming System. To realize the system, more and more kinds of indexing/retrieval methods are needed for *Memory Retrieval*. We introduced the Residual Memory and ``I'm Here'' for *Memory Retrieval*. We also introduced the Ubiquitous Memories for both *Memory Transportation* and *Memory Exchange*.

We are planning to continue implementing and integrating modules for *Memory Retrieval*, *Exchange*, *Transportation*, and *Editing*. *Memory Editing* will be particularly important because the Augmented Memory Albuming system should provides the user with a method to make annotations in augmented memories. We believe that the future implementation of Wearable Virtual Tablet[8] must play the role. On the other hand, we are also constructing a memory-aid model for computational augmentation of human memory in everyday life. We consider the two progresses as important points that must be resolved in the research field of memory-aids in everyday life.

Bridging between real world and symbolized world is also essential to improve the Augmented Memory Albuming functions. In everyday life, a human sequentially handles plural 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 system to recognize the task the user performs from the sequence of symbols, i.e., accesses to identified objects. Such function enables the system to suggest to the user what objects should be used and where they are placed.

References

[1] T. Kawamura, Y. Kono and M. Kidode, ``Wearable Interfaces for a Video Diary: towards Memory Retrieval, Exchange, and Transportation,'' *Proc. ISWC2002*, pp.31--38, 2002.

[2] 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.

[3] R. J. Rhodes, ``The Wearable Remembrance Agent: a System for Augmented Memory,'' *Proc. ISWC'97*, pp.123--128, 1997.

[4] J. Farringdon and Y. Oni, ``Visual augmented memory (VAM),'' *Proc. ISWC2000*, pp.167--168, 2000.

[5] T. Kato, T. Kurata and K. Skaue, ``VizWear-Active: Towards a Functionally-Distributed Architecture for Real-Time Visual Tracking and Context-Aware UI,'' *Proc. ISWC2002*, pp.162--163, 2002.

[6] M. Kidode, ``Design and Implementation of Wearable Information Playing Station,'' *Proc. 1st CREST Workshop on Advanced Computing and Communicating Techniques for Wearable Information Playing*, Nara, Japan, pp.1-5, 2002.

[7] T. Kawamura, T. Fukuhara, H. Takeda, Y. Kono and M. Kidode, ``Ubiquitous Memories: Wearable Interface for Computational Augmentation of Human Memory based on Real World Objects,'' *Proc. 4th International Conference on Cognitive Science*, 2003 (To Appear).

[8] N. Ukita, A. Terabe, Y. Kono, and M. Kidode, ``Wearable Virtual Tablet: Fingertip Drawing Interface using an Active-Infrared Camera,'' *Proc. IAPR Workshop on Machine Vision Applications (MVA2002)*, pp.98-101, Dec. 2002.

[9] T. Kawamura, Y. Kono and M. Kidode, ``A Novel Video Retrieval Method to Support a User's Recollection of Past Events Aiming for Wearable Information Playing,'' *Proc. Second IEEE Pacific-Rim Conference on Multimedia (PCM2001)*, pp.24-31, Oct 2001.