Wearable Interfaces for a Video Diary: towards Memory Retrieval, Exchange, and Transportation

Tatsuyuki Kawamura, Yasuyuki Kono, Msatsugu Kidode
{*tatsu-k,kono,kidode*}*@is.aist-nara.ac.jp*
Graduate School of Information Science, Nara Institute of Science and Technology
8916-5 Takayama, Ikoma 630-0101, Japan

Abstract

In this paper, we discuss wearable interfaces for a computational memory-aid useful in everyday life. The aim of this study is to develop a Video Diary system with vision interfaces to aid in memory retrieval. The Video Diary system provides users with "Memory Retrieval, Exchange, and Transportation" through four types of indexes: 1) through the user's location, 2) through real world object(s), 3) through keyword(s) and 4) through the use of a summary or "the story of the day." The authors have developed the following two systems to achieve the above indexes: 1) a Residual Memory system, 2) a Ubiquitous Memories system. Residual Memory can index a user's location automatically by analyzing a video recorded from a wearable camera for "Memory Retrieval." Ubiquitous Memories provides users with the ability to associate augmented memories with real world objects for "Memory Exchange." We have integrated the above two systems for "Memory Transportation." We believe that the above interfaces can be integrated into the Video Diary system.

1 Introduction

We are in the process of developing man-machine interfaces for a next-generation wearable information play station [10]. To make these interfaces, we use vision-based systems, including various cameras, a head-mounted display (HMD), and other miniaturized visual devices. Images are observed by a wearable vision system with a wearable camera(s) and the system analyses the observed images. We focus particularly on a useful wearable vision system named the *Video Diary* (Fig.1), 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

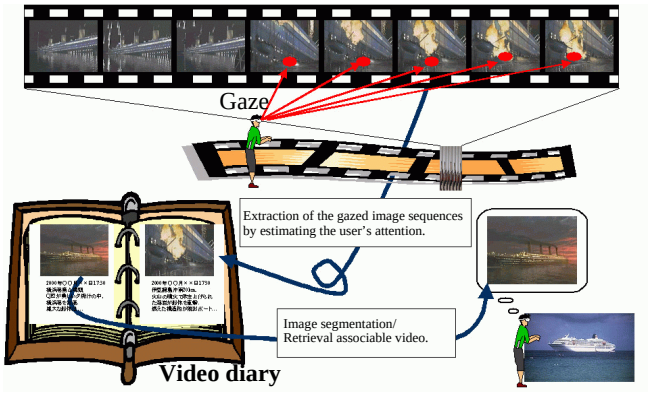

Figure 1: Concept of Video Diary

his/her experiences by reviewing a stored video which is retrieved by consulting the indexes, i.e., both the automatically stored and the manually annotated information. Note that we define "video," which was recorded from user's viewpoint, as an augmented memory. We consider the Video Diary to be a strong application for a wearable information play station. Fig.2 illustrates the overall architecture of the proposed Video Diary system.

Video Diary provides users with memory retrieval through the following four indexes. The first index can be automatically indexed, although the others are manually indexed by the user:

A1: User's location A person tends to recall an experience he/she had in association with the location where it happened. To make the Video Diary system, user's location information is automatically indexed both outdoors and indoors. Such indexes can be acquired by analyzing user's viewpoint images recorded by the wearable camera.

M1: Real world objects A person retains a relationship between real world objects and his/her memory associated with such objects [9]. If each

real world object is identified, then Video Diary will be able to 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 transport his/her memories of an object and exchange them with other users via the image linked to the object.

M2: Keywords In addition to the above two types of indexes, a person tends to remember an event he/she has experienced with conceptual/linguistic annotations, i.e., keywords. Video Diary provides the user with a method to make annotations in augmented memories.

M3: The story of the day A person summarizes the events he/she has experienced in a day explicitly/implicitly. For example, he/she may keep a diary to explicitly summarize events. In summarizing events, a person weaves a tale of his/her own context, and at diary contains the episodes in which he/she is interested. Providing the user with a means for editing and annotating his/her own diary is crucial to the Video Diary system.

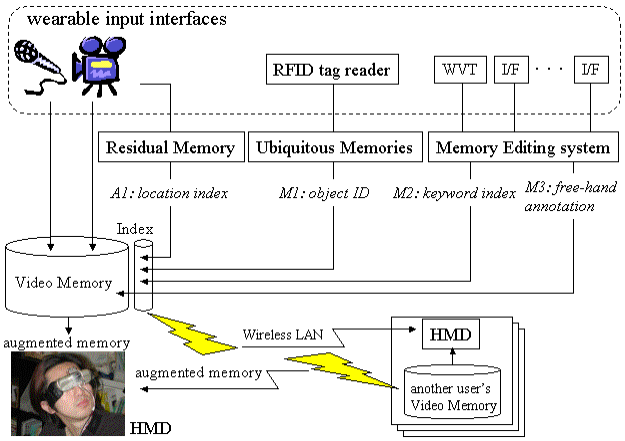

Figure 2: Proposed configration of Video Diary

We believe that the following interfaces can be integrated into the Video Diary system. We have developed the following two wearable vision interfaces using the first two types of indexes (A1 and M1) and one basic vision technique used to acquire the latter two types of indexes (M2 and M3):

Residual Memory: This system stores the continuous image sequences observed from the user's viewpoint while the user walks around an area or "scene." When the user arrives back at the location where he/she has been before, i.e., when the current input image is identified with the stored one, the system shows the stored one on the HMD. This system assists (as A1) the user in remembering the event that happened at that location.

Ubiquitous Memories: To establish associations between augmented memories and real world objects, a user is equipped with an RFID device on one of his/her wrists and each real world object is implanted/attached to a micro-miniature RFID tag. By simply "touching" an object, i.e., the user's wrist accesses the object, a tag ID of the object is identified. The M1 type of index is then realized by the system. The user is allowed to explicitly link his/her augmented memories with real world objects. When the user touches an object that has already been linked with his/her augmented memories, the ID of the object is identified and the system plays back one of the memories linked with the object. Ubiquitous Memories reduces a user's memory overload, because the user is able to recall his/her experience by the simple operation of "touching." Furthermore, Ubiquitous Memories helps its users exchange their experience by showing others' augmented memories.

Wearable Virtual Tablet (WVT) [15]: Wearable Virtual Tablet is a vision interface using only the user's fingertips which can input free-contents in the real world by utilizing the index of M2. The interface detects and tracks a user's fingertip and an arbitrary rectangular plane held by the user's hand. The plane is regarded as an input plane. The user's fingertip and the input plane are utilized as a pen device and an input plane, respectively. While the fingertip traces the plane, the traced locus on the plane is regarded as the input. Although we also try to realize the M2 and M3 types of indexes in this system, the interface does not have enough functions to use both indexes.

Both Residual Memory for "Memory Retrieval" and Ubiquitous Memories for "Memory Exchange" have been developed together as a prototype system. We have integrated the Residual Memory system with the Ubiquitous Memories system for "Memory Transportation" that is a sharing technique of augmented memories among remembrance functions. We are planning to integrate the Wearable Virtual Tablet interface with the above memory-aid system. In this paper, we present the above three systems, i.e., Residual Memory,

Ubiquitous Memories, and the integrated memory-aid system. Although the tablet is still at a basic stage, we believe that it is a crucial interface for coping with both M2 and M3 types of indexes in everyday life, in order to fit what we call editing activities of the diary itself "Memory Editing" to the Video Diary system.

2 Related Works

The "Forget-me-not" system [11] records a user's action history using sensors implanted in a laboratory, and active badges worn by users. The user can refer to his/her own history and can easily recall past events, e.g., a person he/she met or the time when he/she gave the person a document, by viewing a PDA. The remembrance agent system [13] supports the editing of documents related to a particular time/place by referring to the history of the editing operations by the user. The use of these two studies is limited to an indoor environment, because sensors are placed indoors.

In contrast to the above two studies, the following studies use a video and a stand-alone type wearable system. Clarkson and Pentland's system [3], however, cannot directly retrieve previous associable video data for a user who wants to know detailed location information. Aoki, Schiele and Pentland's system [2] also cannot quickly select a similar video of approximately the same viewpoint from continuously recorded video because an offline training sequence is used.

In the field of memory-aids on wearable computers used in everyday life, DyPERS [6], a video replay system, stores a user's visual and auditory scenes. This system can retrieve a video clip using a signal that was explicitly registered by the user by having the user push a button while looking at an interesting scene. StartleCam records a video data triggered by skin conductivity from a startled response from a user [7]. Aizawa et al. proposed a system that summarizes a video from a wearable camera using brain waves [1]. The above two systems automatically start recording when the user shows a noticeable interest. The VAM system detects a human face, which was recorded previously, and displays information about the retrieved person [4]. These types of research have developed human-centered computing systems that focus on recognizing the user's interests by representing augmented memory. Remembrance strategies, however, are not only operated with interests or information in the human brain, but are also involved with the relationship between events and real world objects.

3 Residual Memory

Figure 3: Similar images in terms of location

To achieve a location-based memory-aid, a system should retrieve video data, which includes similar images, recorded previously to the current input image. The two images in Fig.3, for instance, are slightly different because the hand of a user cuts into the scene in the right image. However, both images should be regarded as similar in terms of location. In this system, 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 data set 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.

When the system retrieves the stored video data that includes images similar to the input images, the system plays back the retrieved video data in the HMD.

3.1 Image Retrieval using Motion Information

A wearable computer must treat motion information. We divide motion information into two types. One type is the user's head motion information. Another type of motion information is moving object information. These two types of motion in a user's viewpoint image should be computationally, separated. The following three processes are used to recognize location-based similar images:

Tracking Head Movements: We have set two mono-gyro sensors at the center of the wearable camera. In order to remove the user's head motion information, an examination of the relationship between the amount of value transition with the gyro sensor and the amount of shift with images is necessary.

Tracking Moving Objects: We have adopted a block matching method that detects areas, each of which includes moving objects in a scene. We simplify calculated block motion vectors down to five states (up, down, left, right, and non-movement). If a motion vector is adequately small, this block is named a "non-movement."

Exclusion of Motion Information: In order to remove mutual motion blocks in each image from target searching blocks, a motion block mask should be made. First, the image matching process compares the same address block in two images with blocks called "non-movement" states. The block matching method, then, uses two non-movement states. The summed value calculated by the previous process is divided by the number of non-movement blocks.

Table 1: Relevance and recall rate

	relevance		recall	
	proposed	normal	proposed	normal
outdoor	0.90	0.54	0.98	0.96
hall	0.97	0.56	0.92	0.90
room	0.88	0.57	0.97	0.96
average	0.92	0.56	0.96	0.94

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. The "normal" method, which does not consider motion information, was performed to compare with our proposed method. Table 1 illustrates the relevance and recall rates of both methods. 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.

3.2 Video Scene Segmentation

The system detects scene changes by combining color differences appearing on the entire screen with a moving average method with two mono-axis gyro sensors. If the difference between the appearance of two images and the amount of the transition value of the gyro sensors are large between segmental images, then the system regards the images as the point of a scene change in obtaining the meta-trend of captured data. The method equation is as follows:

$$MA_T(t) = \frac{\sum_{i=t-T}^{t} f(i)}{T}. \quad (1)$$

Four values are calculated by the moving average method: Two values are calculated with yaw-axis gyro values, and the other two values are calculated with pitch-axis gyro values. The moving average value of a short interval (e.g. T=30) and that of a long interval (e.g. T=300) are calculated by each axis gyro value.

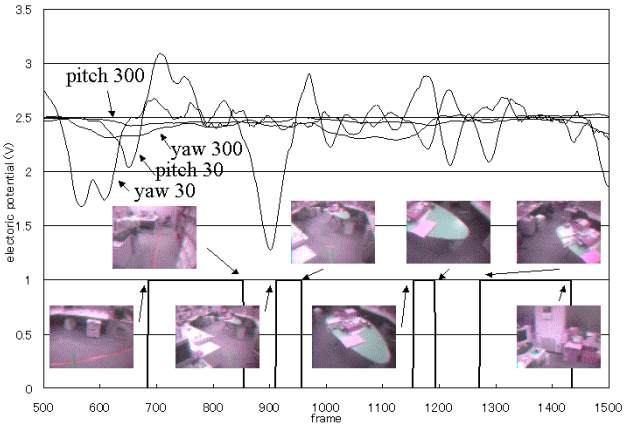

Figure 4: Making scene changes using two mono-gyro sensors

Both image and gyro data, which consist of 5584 frames (191.36 seconds, 29.19 fps), were recorded for evaluation. A test subject walked two times around a specified route in our laboratory. The remarkable result of the experiment is shown in Fig.4. The upper lines in the figure are the moving average values. The lower line shows the scene changes. 41 scenes were segmented from a video data set. The average interval of segmented scenes was 119.51 frames (4.09 seconds). The minimum and maximum interval of segmented scenes were 31 frames (1.06 seconds) and 756 frames (25.90 seconds). A minimum scene was made when the subject walked quickly through a narrow path. The maximum scene change was segmented on a long, straight, uniform hallway.

3.3 Real-time Video Retrieval

The proposed video retrieval method is based on similarity predictions, which divide video data into small segments and retrieve the associable video data, because the cost of the retrieval process increases as the video data set becomes large. In the retrieval process, all images in a segment l are compared with a current query image from the wearable camera. In the next process, the next segment $l+1$ is selected when the maximum image similarity, $HM(l)$, is under a threshold th. We consider the following hypothesis: Similar images form clusters in a sequential video data set.

$$l = \begin{cases} l, & \text{for HM(l)} \geq \text{th}, \\ l+1, & \text{for HM(l)} < \text{th}. \end{cases} \quad (2)$$

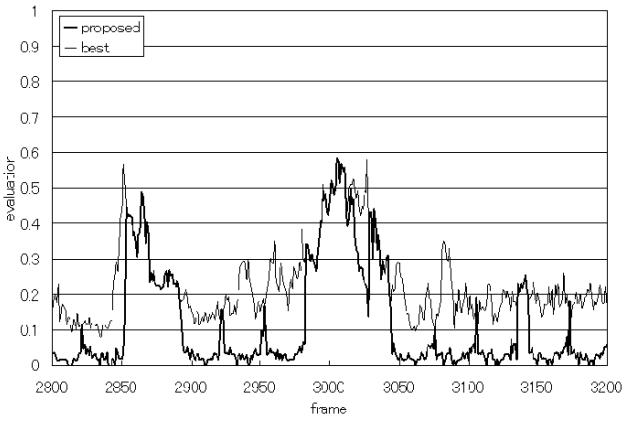

Figure 5: Similar images in terms of location

We compared the proposed method and the normal method. The data set is the same as that for the scene segmentation. We used the first half of the recorded data set as an associable data set, and the latter half as a query of one. We divided the associable data set into 30 small segments. We set a process condition that retrieves location-based similar images from the same segment when the evaluation value is over or equal to 0.2. The experimental result is shown in Fig.5. The calculation time of the process per frame reduced searching for a full data sequence to 1/30. The best evaluation value of all data set is tracked by comparing the same segment of divided video data to the query image when the evaluation value is maintained over or equal to 0.2.

4 Ubiquitous Memories

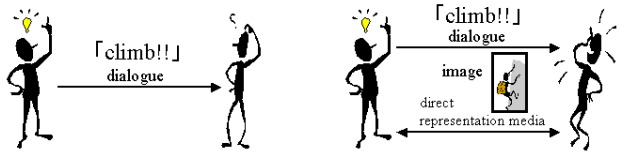

Figure 6: Direct representation media for memory exchange

When one person tells another person a story related to a common experience, a direct representation of the augmented memory of the experience aids in a collective recollection of the memory (see Fig.6). To realize such types of memory exchange, the system should provide users with the following abilities:

- The ability to link memory transportation media, i.e., augmented memories, with real world objects.
- The ability to share direct representation media, which can be used with verbal information.

4.1 System Hardware

A system with direct media representation must use "natural" operations such as the operation of "touching." A user picks up a real world object when the user is interested in that object or needs that object in everyday life. We have developed the Ubiquitous Memories system because the "touching" operation in this system is a natural operation used to establish a link between the user's external activities and an operational object.

The user wears a Head-mounted Display (HMD) and a camera, which is attached at the center of the HMD, on his/her head. The user is also 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 within 3cm. The entire system connects to the World Wide Web via a wireless LAN.

4.2 Operational Procedures

We have developed a prototype system of Ubiquitous Memories to create a real-world-oriented augmented memory concept. Fig.7 illustrates an overview of the Ubiquitous Memories system and the basic procedures

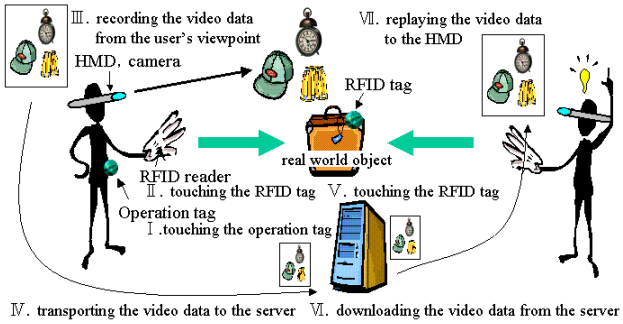

Figure 7: An operational overview of the Ubiquitous Memories system

(a) Selecting an object (b) Replaying the video

Figure 8: HMD view of the operation REMEMBER.

of the system's operation. The system 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 a 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 the gloved hand near a physical 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 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 he/she chooses to remember (shown in Fig.8 (a)) (V). The system retrieves the video data from the web server (VI). The retrieved video data is then replayed in the top-left area of the screen of the HMD (shown in Fig.8 (b)) (VII).

Two experiments have been conducted to evaluate the effectiveness of the Ubiquitous Memories system, and the result have shown that this system effectively supports memory recollection of past events for an individual user [9].

5 Integrated Memory-aid System

In this section we discuss our newly developed memory-aid system into which the Residual Memory system and the Ubiquitous Memories system have been integrated. A user employs a remembrance strategy among various strategies, to recall required information. The system, however, which provides a mono-remembrance strategy to a user, cannot always retrieve augmented memories that are required by the user. The system should have the ability of selecting a strategy automatically or providing an interface to memorize, retrieve, or share augmented memories manually. To provide a multi-selectable remembrance strategy, the system should have the following function:

- The function of memory transportation among remembrance functions each of which provides the user with a remembrance strategy.

Using the integrated system, the user can manually transport augmented memories by selecting a remembrance strategy from the location-based memory-aid or the object-based one.

5.1 Memory Transportation Procedures

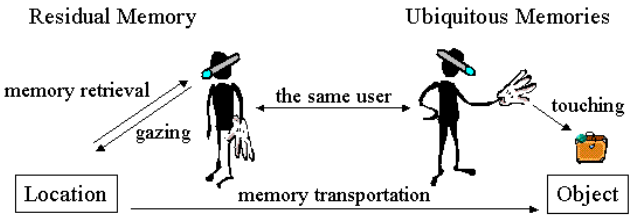

Figure 9: Memory Transportation

Fig.9 illustrates the basic procedures in the operation of our integrated system. In this study, we have developed a function of memory transportation from a user's location to a real world object. Using this integrated system, the user can link a real world object with an event that happened before at the location where he/she now stands. In other words, memory transportation allows the user to link an augmented memory with a real world object from an associated

past viewpoint retrieved by the location-based remembrance function. The system provides the user with the following procedures for memory transportation:

1) The user goes to a location where he/she has been before.

2) The system retrieves the augmented memory recording the event that happened at the location, so that the user can recall the event by simply "gazing" at the location.

3) The user performs the "touching" operation on a real world object, so that a link between the augmented memory, which is displayed on the HMD, and the object is established.

The integrated system use simple operations, such as "gazing" and "touching" to realize memory transportation, in everyday life.

5.2 System Configuration

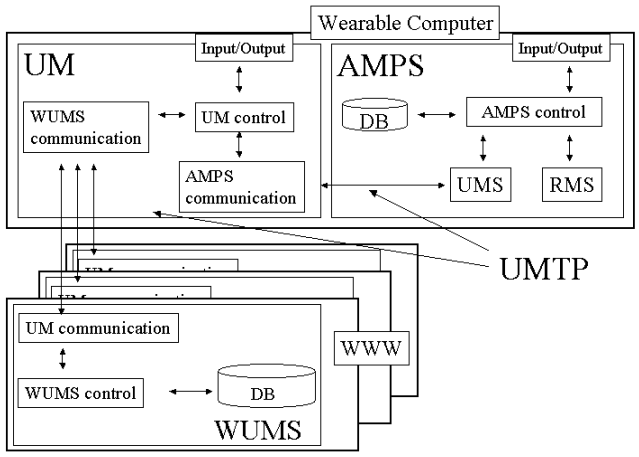

Figure 10: The configuration diagram of the integrated system

In Fig.10, the system components are UM (Ubiquitous Memories), WUMS (a Web-based Ubiquitous Memories Server), and AMPS (an Augmented Memory Processing System). These components communicate both control messages and augmented memories by using UMTP (a Ubiquitous Memories Transfer Protocol) with each other. WUMSs are set in the WWW, and UM and AMPS are installed into the PC that the user wears. These modules communicate with each other via the wireless internet.

AMPS: AMPS is the core component of the system, which provides services to the user. This component is constructed of the AMPS Control module, the UMS (Ubiquitous Memories Side) module, the RMS (Residual Memory Side) module, the Input/Output module, and the AMPS Database. The AMPS Control module manages the system by sending/receiving all the control messages to/from other modules. The UMS is a communication module used to connect to the UM component. The RMS corresponds to the Residual Memory system described in Sec. 3. The Input/Output captures a user's viewpoint images and signals from a wearable device, such as from, a wearable camera, gyro sensors, and accelerometers, and shows augmented memories on the HMD.

UM: The UM component manages links between augmented memories and objects implanted/attached to RFID tags. This component is constructed of the UM Control module, the AMPS Communication module, the WUMS Communication module, and the Input/Output. The UM Control module manages the system by sending/receiving all the control messages to/from other modules. The AMPS Communication module exchanges both augmented memories and control messages with UMS, and the WUMS Communication module exchanges with UM Communication module in WUMS. The Input/Output in this component employs a RFID tag reader/writer.

WUMS: WUMS components are set on the WWW. Each WUMS component stores augmented memories. This component is constructed of the WUMS Control module, the UM Communication module, and the WUMS Database. The WUMS Control module manages augmented memories that are sent from the user's PC. The UM Communication module exchanges augmented memories and messages from the UM component.

6 Concluding Remarks

In this paper we discussed vision-based wearable interfaces for the proposed computational memory-aid system. To realize the Video Diary system, various kinds of indexing methods are needed for "Memory Retrieval, Exchange, and Transportation." We introduced Residual Memory for "Memory Retrieval," Ubiquitous Memories for "Memory Exchange," and one integrated system for "Memory Transportation."

In the future, we plan to continue with the integration of remembrance functions for "Memory Retrieval, Exchange, Transportation, and Editing." "Memory Editing" will be particulary important because Video

Diary should provides the user with a method to make annotations in augmented memories. We believe that the future implimentation of Wearable Virtual Tablet 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.

In the integration of remembrance functions, several technological problems arise. Perhaps foremost, we need a common computational notation for the treatment of augmented memories among remembrance functions. In addition, other types of problems arise. Therefore, we divide the technological problems into functional problems, and also address problems of quantity, and problems of selection. In terms of functional problems, memory-aids need fundamental technologies, which assure reliability and robustness, regarding recognition and indexing in a really noisy world. In terms of the problems of quantity, we must pay attention, for example, to the funding and human resources needed to actualize serviceable memory-aid systems, because numerous remembrance strategies are necessary. In the problems of selection, the user needs an available interface to select a proper function from the various remembrance functions installed in the integrated system.

To realize the integration of the system, we are working on a common notation usable on a computer constructed from the memory-aid model. In addition, we also plan to use the integrated system as a place to test for the evaluation of the memory-aid model. We believe that the development of the Video Diary system and the construction of the memory-aid model will aid in "Memory Retrieval, Exchange, Transportation, and Editing" in everyday life.

Acknowledgements

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 Corporation (JST).

References

[1] K. Aizawa, K. Ishijima and M. Shiina, "Automatic summarization of wearable video-indexing sujective interest", *Proc. 2nd IEEE Pacific-Rim Conference on Multimendia* (PCM2001), pp.16–23, 2001.

[2] H. Aoki, B. Schiele, and A. Pentland, "Realtime Personal Positioning System for Wearable Computing", *Proc. ISWC'99*, pp.37–43, 1999.

[3] B. Clarkson and A. Pentland, "Unsupervised Clustering of Ambulatory Audio and Video", *Proc. ICASSP99*, 1999.

[4] J. Farringdon and Y. Oni, "Visual augmented memory (VAM)", *Proc. ISWC 2000*, pp.167–168, 2000.

[5] J. K. Hahn, J. L. Sibert and R. W. Lindeman, "Towards Usable VR: An Empirical Study of User Interfaces for Immersive Virtual Environments", *Proc. ACM CHI'99*, pp.64–71, 1999.

[6] T. Jebara, B. Schiele, N. Oliver and A. Pentland, "DyPERS: Dynamic Personal Enhanced Reality System", *Perceptual Computing Technical Report #463*, MIT Media Laboratory, 1998.

[7] H. Jennifer and W. Rosallind, "StartleCam: a cybernetic wearable camera", *Proc. ISWC'98*, pp.42–49, 1998.

[8] T. Kawamura, Y. Kono and M. Kidode, "A Novel Video Retrieval Method to Support a User's Recollection of Past Events for Wearable Information Playing", *Proc. 2nd IEEE Pacific-Rim Conference on Multimedia (PCM2001)*, pp.24-31, 2001.

[9] 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", *Information Science Technical Report #NAIST-IS-TR2002012*, Nara Institute of Science and Technology, 2002.

[10] 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.

[11] 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.

[12] D. G. Lowe, "An object recognition system using local image feature of intermediate complexity", *Proc. International Conference on Computer Vision (ICCV 99)*, pp.1150–1157, 1999.

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

[14] Y. Sumi, Y. Kawai, T. Yoshimi and F. Tomita, "Recognition of 3D Free-Form Objects Using Segment-Based Stereo Vision", *Proc. International Conference on Computer Vision (ICCV 98)*, pp.668–674, 1998.

[15] N. Ukita, Y. Kono, and M. Kidode, "Wearable Vision Interfaces: towords Wearable Information Playing in Daily Life", *Proc. 1st CREST Workshop on Advanced Computing and Communicating Techniques for Wearable Information Playing*, pp.47–56, 2002.