Using Eye-Gaze and Visualization to Augment Memory
A Framework for Improving Context Recognition and Recall

Jason Orlosky[1], Takumi Toyama[2], Daniel Sonntag[2], and Kiyoshi Kiyokawa[1]

[1] Osaka University, Osaka, Japan
{orlosky@lab.ime,kiyo@ime}cmc.osaka-u.ac.jp
[2] German Research Center for Artificial Intelligence, Kaiserslautern, Germany
{takumi.toyama,sonntag}@dfki.de

Abstract. In our everyday lives, bits of important information are lost due to the fact that our brain fails to convert a large portion of short term memory into long term memory. In this paper, we propose a framework that uses an eye-tracking interface to store pieces of forgotten information and present them back to the user later with an integrated head mounted display (HMD). This process occurs in three main steps, including context recognition, data storage, and augmented reality (AR) display. We demonstrate the system's ability to recall information with the example of a lost book page by detecting when the user reads the book again and intelligently presenting the last read position back to the user. Two short user evaluations show that the system can recall book pages within 40 milliseconds, and that the position where a user left off can be calculated with approximately 0.5 centimeter accuracy.

1 Introduction

It has been long known that humans often fail to convert short term memory into long term memory, and are inherently forgetful. We often mistakenly judge certain events as being unimportant, but which turn out to be important at a later time or in a different context. To help cope with this memory deficiency, technology has been used as a form of cognitive offloading to assist and sometimes even function as a substitute for memory intensive tasks. Good examples include digital calendars, reminder systems, life logging applications, and the use of search engines for information not committed to long term memory [1], [4], [7], [8]. Our research builds on this idea by augmenting memory through the use of eye-tracking and an AR display as shown in Figure 1. When a user returns to the situation in which a memory occurred, eye gaze can be used to detect context and more accurately present the user with previously stored information. Eye tracking is first used to identify a user's point of attention and to outline an area for recognition, such as text or an environmental object. That text or object is then inserted into a database along with relevant tags such as date, time and location. "Memories," represented by an array of contextual and temporal tags in the database, can be recalled later with keyword searches or object detection triggers.

N. Streitz and P. Markopoulos (Eds.): DAPI 2014, LNCS 8530, pp. 282–291, 2014.

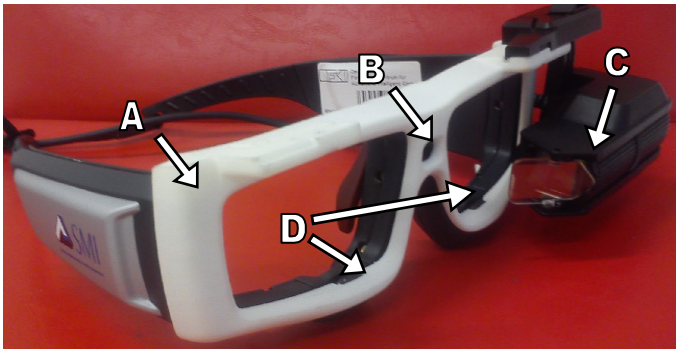

Fig. 1. Hardware setup showing A) 3D printed connector, B) outward facing camera, C) head mounted display, and D) inward facing eye-tracking cameras

Examples of applications of this technology include recalling items such as forgotten page numbers in documents, the location of misplaced keys, or patient information prior to surgery. Interfaces such as this one also have the potential not only for consumer use, but for use with clinical patients suffering from memory related illnesses such as vascular dementia or Alzheimer's disease. In addition to describing a general framework that facilitates the encoding of temporal events into digital form, we describe the hardware setup shown in Figure 1 and specific software implementations within the framework. These implementations include a system that can help a user recall a lost book page and a system that can encode an event, such as placing one's keys on a desk, into the database for later recall. Though a variety of implementations within this framework are possible, we chose page recollection and simple event storage since they are prime examples of how this framework can translate to practical application. To our knowledge, this is the first attempt at introducing a combination of eye-gaze interaction and AR into this kind of memory assistive model.

2 Prior Work

2.1 Research on Memory and Context

One widely explored field of research related to memory is that of physical systems that serve as memory aids. One such example is the SenseCam, which takes intermittent photos throughout the day and serves as a retrospective memory aid [4]. Detailed studies using the SenseCam show that memory can be improved by reviewing images taken by the system, especially long term memory [13]. A similar device called the EyeTap has been used as a form of capturing life experiences and sharing these experiences with others [7]. Although a large number of other software memory aids such as calendars and reminder systems are available, a majority of them only exist as mobile or smartphone based applications. Several systems are also available that utilize sensor data in order to extract context. One such system by Belimpasakis proposes a client-server platform that enables not only life logging, but richer social

experiences by extracting more meaningful contextual information from data [1]. The above systems all have the potential to be combined with or improved by various models for memory and decision making, such as those proposed by Hutter et al. [5]. They also fall into the broader goal of creating a complete database of all life events [2].

2.2 Eye Gaze and Augmented Reality

Another set of closely related studies are those which use computer vision and eye tracking for context recognition. One major branch is the study of object recognition, which can be conducted using hysteresis, feature tracking, and other algorithms [6], [16]. This type of method can help with context recognition since it has the potential to extract semantic information from objects in one's environment. In addition to recognizing objects, location can also be extracted using gaze and other sensors [14]. In conjunction with HMD systems, activity can also be recognized using other types of mobile sensors [12]. Once context, location, or other relevant content has been determined, information visualization methods can be used to place the information in a relevant location in the environment [10]. This can prevent information from becoming a distraction, and can make recalled information easier to view. Our framework uses a combination of elements from life logging, context recognition, memory models, and AR in order to assist users with event recall.

3 Hardware Design and Setup

We first construct a 3D gaze tracking system combined with an HMD that does not require the use of external tracking or projection hardware. The device is composed of a pair of eye tracking goggles, custom 3D printed attachment, and HMD. The devices are all connected, and can be calibrated as a single system.

3.1 Hybrid Eye Tracker and HMD

To start, we needed an apparatus for eye and vergence tracking that could be used simultaneously with a head mounted display placed near the user's eye. We decided to use a pair of SMI Eye Tracking Goggles, which can be worn like glasses and leave enough room to attach an HMD. The HMD part of our system consists of an 800 by 600 pixel AirScouter HMD, which includes digital input via USB and depth control. The focal depth can be set from 30 centimeters (cm) to 10 meters (m).

In order for gaze to be measured appropriately in the HMD, a user's eye convergence must be consistent and eye tracking hardware must provide enough accuracy to ensure consistent gaze on a target object of interest. In addition, we needed a way to make sure that the distance between the tracker and HMD would remain at the same during use. To ensure these conditions, we created a 3D printed fastener that fixed the distance between the prototype HMD and the eye tracker as shown in Figure 1 A. This setup allows for both left and right eye configurations.

3.2 Aligning Virtual and Real Environments

In order to provide information back to the user in an intelligent fashion, in many cases we have to align digital text with objects in the scene. In the case of recalling a book page or sentence in a document, text and pointers must be displayed in line with the targeted object and text. First, when a scene image is taken from the camera, the image is blurred by a Gaussian kernel and thresholded into a binary image in order to detect the centroid of each word region. The retrieval process is done by matching extracted features to the features of books, documents, and other media previously stored in the database [17]. Since we apply an image based method, we can deal with a variety of different paper mediums, fonts, and sizes. By matching the features between the scene image and the retrieved database image, we also calculate the homography between them. Based on this homography, the pose, rotation, and transformation of text in the scene image can be is estimated. This data can also be used both for HMD calibration and correct projection of overlaid data.

3.3 Gaze Calculation, Calibration and Focus Detection

Though the calculation for aligning virtual and real environments is independent of eye tracking, calibration for the eye tracker, HMD, and document pose estimation can be done all at once. From the eye tracker, we first extract a 3D vector of the direction of each eye, represented by GL=(glx,gly,glz) and GR=(grx,gry,grz) in Figure 2. Using this data, the approximate intersection of the two gaze vectors in 3D space is calculated. Though the specific details of the eye tracking process are proprietary, there are several basic steps that occur. Images are first taken from two infrared eye cameras and one scene camera. Each eye is then illuminated by six infrared light sources and the system tracks the changes of the six reflections off of the eye. Adjustments can be made for the height and width of a user's eyes, but the distance between eye tracker and HMD will still remain constant because the two pieces are connected.

Next, we calibrate the whole system, which is done in two steps. In the first step to calibrate gaze, the wearer to looks at one or more arbitrary points in the real world, allowing the system to adjust for that specific user's eyes. After this step, the system can be used for object recognition and non-environmentally aligned display of recalled information.

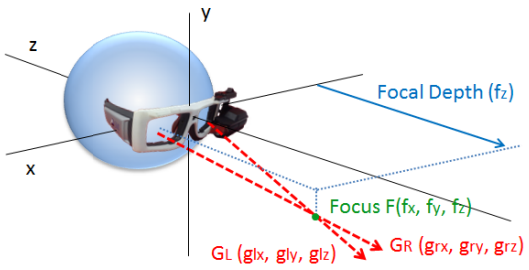

Fig. 2. Visual representation of the gaze vectors G_L and G_R used to compute gaze direction

In order to allow the HMD to align text with books or documents in the real world, a second calibration is used to determine the size and position of documents in the real world relative to the HMD camera and display screen. This calibration is done by asking a user to gaze at calibration point on a specific document, allowing the system to determine the size of the document relative to the user's field of view, thereby finalizing the entire calibration process.

When trying to extract a user's focus on an object or line of text, we utilize temporal fixation detection [16]. Once an object or document in the environment is detected, a threshold which can be set based on user preference functions as a trigger for encoding an event. In the case of recalling an item, such as a book page, detection of a book's cover or document title will trigger the display to output the last page or location viewed by the user. Next, we describe the framework for storage and recall.

4 Framework

Data processing within this framework primarily occurs in one of three steps, as shown in Figure 3. The first phase is interaction, where the primary sources of input are the position extracted from the eye-tracking interface, the environmental image from outward facing camera, and sensors such as GPS, accelerometer (for determining activity through methods such as those by Ravi et al.), and system time [12]. The second step is the encoding of this information into the database. Input data is stored in the database as an array of searchable keywords, and elements like time are stored as a chronological array. Finally, recollection of events is triggered by user initiated keyword search or by recognition of current context, and relevant database entries are displayed back to the user through the mixed-reality display.

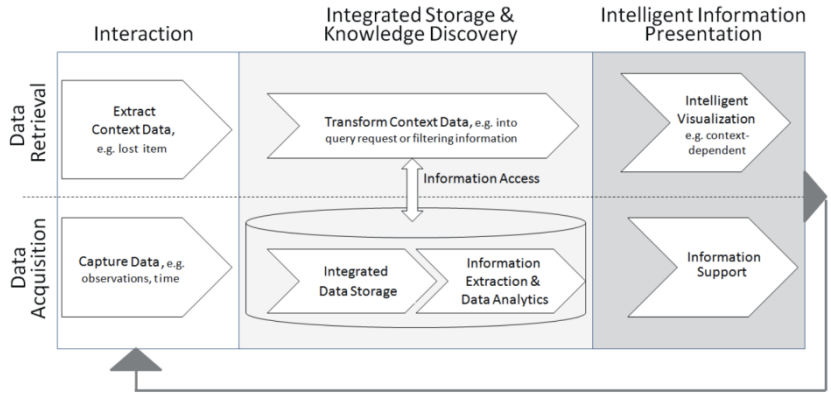

Fig. 3. Modified design of information framework showing flow of information [15]. Original image courtesy of Siemens AG.

4.1 Gaze Interaction and Context Recognition

The first and likely most important part of this framework is the method by which we recognize a user's context. Although there are a variety of methods used to detect objects in the environment such as those by Lowe et al. and Belongie et al., we use image retrieval for document recognition, and can also extract text in the environment through an optical character recognition (OCR) algorithm [2], [6]. Our system can also accept predefined markers as input, which provides a diverse test bed to show the interaction between recognizing objects in various contexts and storing/recalling information from the database. Using gaze as a form of interaction and search, we can recognize books, store names and locations, room numbers, and other text-based information that tends to be easily forgotten. We use marker tracking in place of object recognition which allows us to focus on methods for storage into the database and recall into the user's current context. As an example, recognition of text on a door plate and a marker on a set of keys can be stored in the database as an event. Alternatively, restaurant names and preferred menu items can be recognized and stored.

4.2 Database Design and Storing Events

Once an object or set of objects has been recognized, it must be stored in the database in a particular context. One important dimension for context is time, since events that occur closer to one another are likely more closely related. Time is also important in human memory, like our ability to remember procedural tasks or sequential events better than randomly distributed ones. This is the reason why many people must sing a song from the beginning in order to recall a particular phrase in the song. Other dimensions include semantic relevance, physical location, and custom input for more specific applications. These dimensions can also be cross-referenced to improve recall. Though the number of dimensions could be expanded with additional implementations, the current database elements include 1) an event, which represents the essence of the memory, 2) time, which is the moment in time when the event occurred, 3) location where the memory occurred, 4) semantic context, which includes any available information extracted from OCR or other contextual data extracted from sensors, 5) keyword, which represents an optional additional relevant contextual cue, and 6) an arbitrary field for use with specific implementations, such as the book page recollection algorithm. Though the mechanisms behind human memory are not fully replicated in our framework, these methods can serve as rough metaphor for basic storage and recall of past information.

Database queries can be manually engaged by a user, or automated based on triggers from a certain event or idea. If a user were searching for his or her keys for example, he or she would input "keys" as a keyword search and would be presented with a list of terms from the database contextually related to the word "keys," such as room numbers or objects detected in the immediate vicinity or time frame of the nearest occurrence of keys. This method is comparable to personal information models such as those proposed by Maus et al., but takes advantage of augmented reality for reduced interaction and faster presentation [8].

4.3 Information Presentation and View Management

Once information has been recalled from the database, it must be presented to the user in an intelligent way so that it is in context and does not induce confusion. In the example of finding the last sentence a user was reading in a certain book, simply displaying the first word of the sentence in the HMD would not be enough for a user to find his or her place quickly. Instead, our system uses a document image retrieval and projective calculation to determine the position of the book, and appropriately displays a notification or pointer to where the user left off in the real world. Finally, view management can move resulting notifications to ensure that recalled information does not interfere with reading, walking, searching, or other visual tasks [10].

4.4 Software Implementation

Here we present a software implementation of our framework which accounts for a certain type of cognitive task. Like many easily forgotten events, leaving a reading task without marking the page is a frequent occurrence. By implementing one type of recollection method within our framework, we can solve this problem. Using the same steps outlined in the framework section, this particular method detects when a user is reading a specific document or book, searches the database for any memories related to reading that particular book, and displays navigation cues to the reader to show the page and location where he or she last left off, as shown in Figure 4. In addition to displaying the correct page, pointers show the user the direction of their last reading position, and a line is displayed under the last word read.

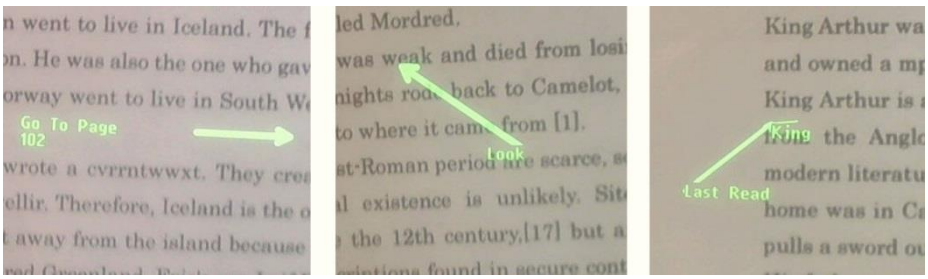

Fig. 4. Images showing recognition of an incorrect page (left), correct page with bookmarked text located above the HMD viewing field (center), and correct position with pointer (right)

Memory Logging and Recall Mechanisms. To provide a visual representation of how entries are recalled from the database, Table 1 shows how entries from a single day would appear in the database, and Figure 5 shows a view through the HMD showing a corresponding list of results using the keyword “book” as a search string queried from that database. The keyword search would be narrowed down further upon adding additional search strings such as time range or location. In the case of visual book recognition, instead of presenting search results, the document recall algorithm takes over, and displays the page and navigation instructions from Figure 4.

Table 1. Sample database entries for a single day

Event	Keyword	Time	Location	Context	Custom
book	Moby Dick	2014-01-22:21:15	home	read	pg88&x36&y773
book	Moby Dick	2014-01-22:18:22	home	read	pg5&x120&y150
magazine	Modern Art	2014-01-22:01:34	library	view	null
book	Relativity	2014-01-22:10:34	school	read	pg200&x52&y318
memo	groceries	2014-01-22:07:21	home	view	null

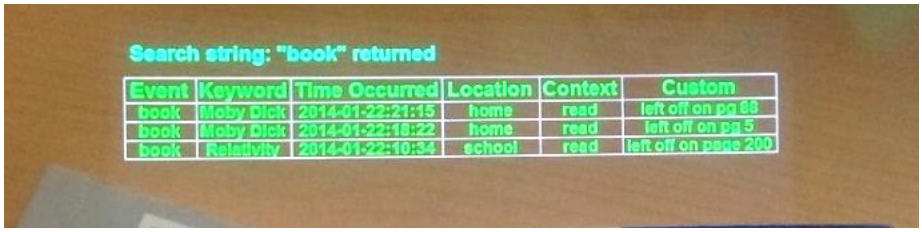

Fig. 5. Segment of an image taken through the HMD viewing screen of returned search output

5 Evaluations of Time-to-Recall and Accuracy

To provide a simple evaluation of book page recall, we conducted two short experiments. The first was designed to measure the time it takes to recognize a page when a user first looks at a document, and the second was designed to determine how accurately the exact reading position could be measured for re-display.

Time-to-Recall. The first experiment was conducted by asking 10 users to wear the display system. Each user was then presented with both a document presented on a computer monitor and a printed sheet, both of which had the same size and text. They were then asked to read each document as if they typically would any other type of text, and the recall algorithm was applied to each frame throughout both reading tasks. Reading angle was also measured for each participant to test whether we could still recall the text despite different viewpoints. Results show that for both the digital and physical documents, the recognition accuracy for each frame was 100% for reading angles between 70° and 90°. There was only a 0.54% decrease in accuracy for viewing angles between 50° and 70°. Other informal experiments showed that for over 50° of deviation from vertical, accuracy of recall decreases rapidly. However, considering that most participants chose to read the texts at between 50° and 90°, we can safely assume that the method is effective for the general viewing of text.

Reading Position Test. In the second experiment, we asked another set of 13 users to read through a document and pause at four different words over the course of two minutes. For each word, we measured the distance between the center of the requested word and the point provided by the eye tracker. On average, the deviation from each word was approximately 0.5 cm across all participants for the two minute

period, and showed a minor decrease in accuracy over the first minute. This distance is equivalent to either one line of text in the vertical direction or one to two words in the horizontal direction, meaning that a user would never have to read more than one or two lines of text away from his or her last reading position. With this level of accuracy, we can conclude that recall of page and position is effective for general use.

6 Discussion

In addition to the general recall of information, we have also explored the possibilities of the eye tracker and HMD setup for recalling patient faces and virtual display of patient records [15]. A generalization of this approach is the exploitation of eye movements in the context of more complex activities for which the role of vision has yet to be explored. New application domains should take daily activities into account and provide for cognitive assistance in those activities. The aim of our current studies is to determine the potential impact of such cognitive assistance for specific user groups in both medical and consumer applications. Our augmented reality setup can also potentially be used to interpret the center of gaze and fixations of dementia patients, which can be used to recall assistive information from the database.

7 Conclusion

In this paper, we propose the use of a combined eye-tracking HMD interface for detecting context, storing events into a database, and virtually presenting those events back to the user at a later time. Within this framework, we implement both the database for storing and recalling events, and a more specific method for recognizing documents, which virtually projects a pointer to the last location in the real world where the user left off. We then conduct two short evaluations testing the accuracy of document recall and reading position, finding that both are effective for practical use. This system can function as a cornerstone for the development of other context sensing AR interfaces, and we hope it will encourage further research on memory assistive technology.

Acknowledgements. Many thanks for the support of DFKI's MedicalCPS, Kognit, and ERmed projects for their contributions to this work. Another thanks to friends, family, and mentors for continued support and encouragement.

References

1. Belimpasakis, P., Roimela, K., You, Y.: Experience explorer: A life-logging platform based on mobile context collection. In: Third International Conference on Next Generation Mobile Applications, Services and Technologies, NGMAST 2009, pp. 77–82. IEEE (2009)

2. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(4), 509–522 (2002)
3. Gemmell, J., Bell, G., Lueder, R., Drucker, S., Wong, C.: MyLifeBits: Fulfilling the Memex vision. In: Proceedings of the Tenth ACM International Conference on Multimedia, pp. 235–238. ACM (2002)
4. Hodges, S., et al.: SenseCam: A retrospective memory aid. In: Dourish, P., Friday, A. (eds.) UbiComp 2006. LNCS, vol. 4206, pp. 177–193. Springer, Heidelberg (2006)
5. Hutter, M.: Universal artificial intelligence: Sequential decisions based on algorithmic probability. Springer (2005)
6. Lowe, D.G.: Object recognition from local scale-invariant features. In: The Proceedings of the Seventh IEEE International Conference on Computer Vision, vol. 2, pp. 1150–1157. IEEE (1999)
7. Mann, S., Fung, J., Aimone, C., Sehgal, A., Chen, D.: Designing EyeTap digital eyeglasses for continuous lifelong capture and sharing of personal experiences. In: Alt. Chi, Proc. CHI (2005)
8. Maus, H., Schwarz, S., Dengel, A.: Weaving Personal Knowledge Spaces into Office Applications. In: Fathi, M. (ed.) Integrated Systems, Design and Technology 2012, pp. 71–82. Springer, Heidelberg (2013)
9. Montemerlo, M., Pineau, J., Roy, N., Thrun, S., Verma, V.: Experiences with a mobile robotic guide for the elderly. In: AAAI/IAAI, pp. 587–592 (2002)
10. Orlosky, J., Kiyokawa, K., Takemura, H.: Dynamic text management for see-through wearable and heads-up display systems. In: Proceedings of the 2013 International Conference on Intelligent User Interfaces, pp. 363–370. ACM (2013)
11. Pollack, M.E., Brown, L., Colbry, D., McCarthy, C.E., Orosz, C., Peintner, B., Ramakrishnan, S., Tsamardinos, I.: Autominder: An intelligent cognitive orthotic system for people with memory impairment. Robotics and Autonomous Systems 44(3), 273–282 (2003)
12. Ravi, N., Dandekar, N., Mysore, P., Littman, M.L.: Activity recognition from accelerometer data. In: AAAI, pp. 1541–1546 (July 2005)
13. Sellen, A.J., Fogg, A., Aitken, M., Hodges, S., Rother, C., Wood, K.: Do life-logging technologies support memory for the past?: An experimental study using sensecam. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 81–90. ACM (2007)
14. Sonntag, D., Toyama, T.: Vision-Based Location-Awareness in Augmented Reality Applications. LAM Da 2013 5 (2013)
15. Sonntag, D., Zillner, S., Schulz, C., Weber, M., Toyama, T.: Towards Medical Cyber-Physical Systems: Multimodal Augmented Reality for Doctors and Knowledge Discovery about Patients. In: Marcus, A. (ed.) DUXU 2013, Part III. LNCS, vol. 8014, pp. 401–410. Springer, Heidelberg (2013)
16. Toyama, T., Kieninger, T., Shafait, F., Dengel, A.: Gaze guided object recognition using a head-mounted eye tracker. In: Proceedings of the Symposium on Eye Tracking Research and Applications, pp. 91–98. ACM (2012)
17. Toyama, T., Dengel, A., Suzuki, W., Kise, K.: Wearable Reading Assist System: Augmented Reality Document Combining Document Retrieval and Eye Tracking. In: 2013 12th International Conference on Document Analysis and Recognition (ICDAR), pp. 30–34. IEEE (2013)