Augmenting and Sharing Memory with eyeBlog

Connor Dickie, Roel Vertegaal, David Fono, Changuk Sohn, Daniel Chen, Daniel Cheng, Jeffrey S. Shell and Omar Aoudeh
Human Media Lab, Queen's University
Kingston, ON. K7L 3N6, Canada
{ connor, roel, fono, sohn, chend, cheng, shell, aoudeh } @ cs.queensu.ca

ABSTRACT
eyeBlog is an automatic personal video recording and publishing system. It consists of ECSGlasses [1], which are a pair of glasses augmented with a wireless eye contact and glyph sensing camera, and a web application that visualizes the video from the ECSGlasses camera as chronologically delineated blog entries. The blog format allows for easy annotation, grading, cataloging and searching of video segments by the wearer or anyone else with internet access. eyeBlog reduces the editing effort of video bloggers by recording video only when something of interest is registered by the camera. Interest is determined by a combination of independent methods. For example, recording can automatically be triggered upon detection of eye contact towards the wearer of the glasses, allowing all face-to-face interactions to be recorded. Recording can also be triggered by the detection of image patterns such as glyphs in the frame of the camera. This allows the wearer to record their interactions with any object that has an associated unique marker. Finally, by pressing a button the user can manually initiate recording.

Categories & Subject Descriptors
H.5.m. Information Interfaces and Presentation, H.5.4. Hypertext/Hypermedia. H.3.1. Content Analysis and Indexing.

General Terms
Human Factors. Management.

Author Keywords
Attentive User Interface. Humanistic Intelligence. Ubiquitous Computing. Blogging. Eye Tracking. Wearable.

INTRODUCTION
We are entering an age of information overload. The pervasiveness of portable personal recording devices that produce large amounts of data or large numbers of files emphasizes this problem. We find ourselves on the brink of an era where the unlimited storage capacity first envisioned by Bush [2] will allow everyone to record and store their entire lives in personal libraries. The increasing amount of data accumulated by the average user is becoming more difficult to manage, and will continue to do so as storage space increases and recording devices approach ubiquity. Although Engelbart [3] had once envisioned computational devices that would augment human intellect, integrating these devices into our every day existence has proven to be challenging. The affordances of virtually unlimited storage are both empowering and constraining. As the notion of manually deleting older files to make room for new ones becomes obsolete, we will have to deal with the reality that finding a particular file, or relevant data within large files becomes more difficult. To combat this we must devise ways to either reduce the amount of recorded information to that which is relevant, or annotate files with information about their content. To this end we present eyeBlog; an automatic, attentive video blogging tool.

CARPE'04, October 15, 2004, New York, New York, USA.

Previous Work
The Humanistic Intelligence (HI) model proposed by Mann [4], describes a framework in which the human and computer interact in close synergy with one another. According to him, HI arises from the very existence of the computer, user and their close interaction with one another. The HI framework describes devices that are in close proximity to, and often worn by the user on a daily basis [5]. As such, meeting the requirements of the HI framework (see Fig. 1) enables the human and computer to achieve the augmentation of intellect envisioned by Engelbart. According to the HI model, the system must:

- Be unmonopolizing;
- Support unrestrictive interaction to the human and the outside world;
- Support observation of system properties;
- Be controllable: The device is responsive to the human's control at any one time;
- Communicate: The computer can communicate with others;
- Pay attention: The computer is attentive to the user and the environment.

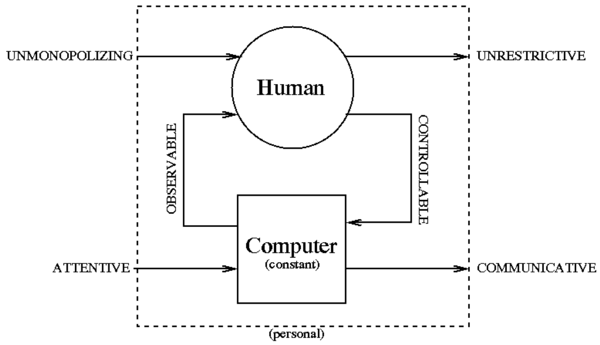

Figure 1. HI framework for augmentation [6]

EyeTap [6] is an illustration of a device that satisfies these properties. EyeTap are a pair of glasses that allow the user to continuously capture and record information. To record, the device employs a circular buffer system such that the last 10 minutes of video is retained. This enables the user to review the last 10 minutes of their life, allowing them to manually save the most memorable moments throughout the day. Although effective, manual initiation of recordings demands attention, and therefore goes against the attentive properties of HI. Manually editing and categorizing the recorded segments from an entire day is both tedious and time consuming. In this respect, the attentive property of the device is the most notable of the fundamental HI framework. That is, the device must be intelligently aware of the context in which it is operating. With this knowledge, we can automate the processes of determining which moments are relevant as well as the recording of metadata. The Attentive User Interface (AUI) paradigm by Vertegaal et al. [7] integrates well with the HI model. AUIs register the attention of the user through sensing proximity, body orientation and eye contact of users. The Attentive Cellphone prototype [8] is an early example of a system that exploits an AUI. Unlike Eye-aRe glasses [9] that used Infra Red emitters and receivers to detect when two pairs of glasses are mutually aligned, the Attentive Cellphone registered actual eye contact between the wearer and an onlooker. It used an eye-contact sensor mounted on the brim of a baseball cap to inform potential callers when the recipient was in a face-to-face conversation. However, due to the parallax between the eyes and camera, eye-contact detection with this design was not precise [8]. By mounting the camera on a glasses-like headset Mann's EyeTap avoids visible parallax. This allows video to be recorded from a first-person perspective while an associated wearable computer streams video content to the web. An example of an early wearable facilitating automated editing and annotation of video content is LAFCam [10]. In LAFCam, points of interest in the video are established by monitoring nonverbal utterances made by the videographer during filming. This metadata is recorded in sync with the video channel giving human editors a quick and easy way to find interesting segments in the footage.

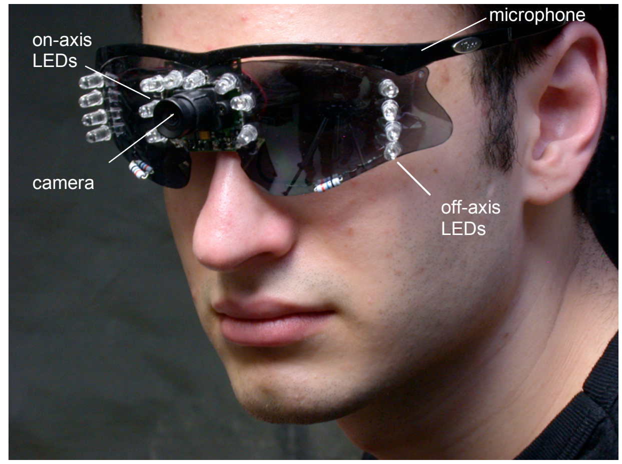

Figure 2. ECSGlasses: camera with LED illuminators

IMPLEMENTATION

Weblogs (or blogs) are regularly updated web pages with more recent posts appearing closer to the top of the page [11]. Blogging is a form of online communication that is rapidly gaining popularity with mainstream users due to its accessibility and ease of use. Typically blogs support single-click publishing of entries. According to Herring et al [12], the vast majority (over 70%) of blogs are authored by ordinary people for their own reference purposes and niche audiences. Blog entries are typically text, but can also include images, sounds, video and other rich media types. eyeBlog is essentially a video blogging system made up of three distinct components. ECSGlasses are a novel wearable camera system that allows a user to record video from a first person perspective upon detection of eye contact at the wearer. This allows the wearer to pay attention to the event at hand, rather than the device being used to record it. The associated recording software, running on a remote computer, uses the information from the ECSGlasses to start and stop recording. Recorded video segments are then appropriately compressed and filed for web publishing. Publishing is handled by an Apache web server and a modified installation of the Moveable Type blogging software and a proprietary Perl script called eyeBlogScript. This allows anyone with Internet access to annotate, grade, catalogue and search the video blog archive.

ECSGlasses

Eye-Contact Sensing (ECS) Glasses are a novel type of wearable input device for detecting the looking behavior of other people without any discernible parallax [1]. They are also capable of recognizing other image patterns such as glyphs. ECSGlasses consist of a standard pair of sunglasses, augmented with an eye-contact sensing camera [1]. We have developed both clear and dark glass versions, the latter is shown in figure 2. Unlike Eye-aRe [9], ECSGlasses do not require others to wear any machine recognizable tag. Instead, the glasses employ computer

Fig 3. Processed Image: circles indicate eye contact detection

Fig 4. eyeBlog: image preview

vision to search for human eye gaze directed toward the wearer's eyes as well as detecting glyphs that appear within the frame. The current design employs an analog Infra Red sensitive camera mounted between the wearer's eyes. On-axis with the camera is a collection of Infrared Light Emitting Diodes (LEDs) surrounding the lens (see Fig. 2). Two more sets of LEDs are situated off-axis near the arms of the glasses. The LEDs illuminate the scene in front of the camera. The Mammalian retina acts as a retro-reflector, bouncing light back in the direction of origin. When the on-axis LEDs are illuminated, a red-eye effect is observed in the pupils of an onlooker. When the off-axis LEDs are illuminated the camera records a similar image, but with dark pupils instead. By syncing the LEDs with the camera clock, a bright and dark pupil effect is produced in alternate fields of each video frame. A simple algorithm finds any eyes in front of the user by subtracting the even and odd fields of each video frame. The LEDs also produce reflections on the corneas. These glints appear near the center of the detected pupils when the onlooker is looking at the eyes of the user, allowing the sensor to detect eye contact without calibration (see Fig. 3). Eye-contact sensing range depends on focal length and the camera resolution. With a 12mm lens and a 320x240 effective resolution, range is 1m with a 50cm field of view. Glyph detection is handled by ARToolkit [14], which searches for specific geometrical shapes arranged in a sequence. It recognizes glyphs by analyzing each image from the ECSGlasses' camera and applying pattern recognition algorithms. The number of possible glyphs is constrained only by the resolution of the camera. These unique graphical identifiers are used to differentiate between objects. ECSGlasses uses an RJ-45 cable that connects to a compact box worn at waist-level. The box supplies power using a 12-volt DC battery, and contains the LED timer circuit board as well as the wireless video transmitter. The video feed is processed on a computer with an associated wireless video receiver. This allows the ECSGlasses to operate untethered. The computer remotely performs image analysis to recognize glyphs or determine if any eyes are oriented at the wearer. After processing image data, the computer streams information about the co-ordinates and orientation of pupils and detected glyph IDs to connected appliances over a wireless network.

Recording

eyeBlog is a video blogging application designed to automatically record and publish face-to-face conversations and interactions with tagged objects. It addresses one of the major problems in video blogging: that of immediate editing and publishing of video content. Digital cameras are typically quite demanding of a user's attention. More often that not, a videographer's attention is focused on the device being used to record an event, rather than on the event being recorded. This is particularly problematic when recording documentaries of one's life experiences. First-person perspective video and automatic recording eases the creation of online video blogs by allowing the user to participate in the event rather than simply monitor it. To limit the amount of footage, current video bloggers are required to engage in the video recording process or manually edit video footage before posting. eyeBlog facilitates this process by using the signal from the eye-contact and glyph sensing camera in the ECSGlasses to decide when to record video from the ECSGlasses' camera. eyeBlog, which runs on a remote computing unit, statistically processes eye contact and glyph data. Recording is triggered when the percentage of eye contact over time reaches a preset threshold value or an object of interest and its associated glyph appears within the frame. Recorded video segments have a resolution of 640x480 and are updated 30 times per second. Videos are compressed using the MPEG-1 compression scheme producing a final average data-rate of 100 kilobits per second. Each resulting .mpg file is stored in the eyeBlog video directory on the web server. A single frame from the beginning of each video segment is used to create an image preview. These image previews are then stored as a .jpg file in the eyeBlog image directory on the web server. By displaying an image preview for each video, users can quickly see the content of a video segment without having to download it. This is shown in Figure 3. This facilitates participation of those who are browsing eyeBlog through low-bandwidth devices such as mobile phones and analog modems.

Blog Publishing

Blog entries are dynamically created as new video segments are recorded. This is handled by eyeBlogScript, which is a Perl script that monitors the video and image preview directories on the web server. When files are added to these directories, eyeBlogScript creates a new blog entry and submits it to Moveable Type where it formats the content for display and adds support for annotation, grading, cataloguing and searching. When a blog entry is created, the image preview is displayed. When a user clicks on the image preview, the associated video segment is revealed in a new browser window. Users have the ability to download the videos to a local drive for offline viewing. There is a space below the image preview reserved for the eyeBlogger's personal annotations. Below this area is a link that allows browsers of eyeBlog to add their own annotations. In addition to this, Moveable Type automatically generates an XML representation of the annotation, grading and cataloguing information from recent entries and stores this as a Rich Site Summary (RSS) feed. RSS is a simple way for bloggers to syndicate their content in digest form. Users can subscribe to the RSS feed and automatically be notified when eyeBlog is updated. Many RSS readers support the ability to filter feeds so only blog entries that match the specified conditions are displayed.

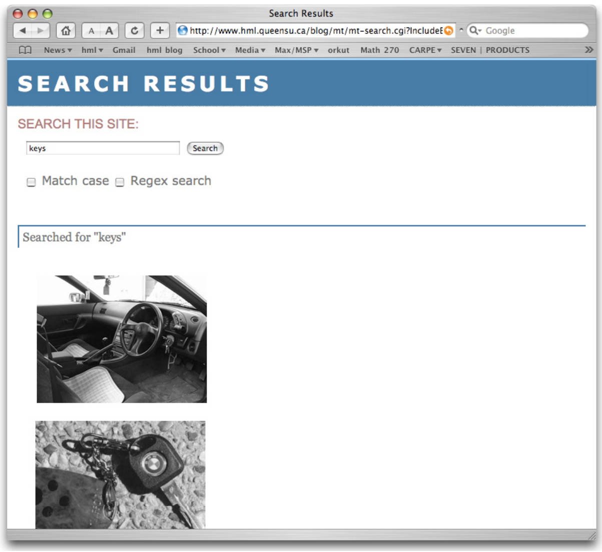

Fig. 5 eyeBlog: search results

SCENARIOS

We present three scenarios that depict how users can both augment memory and share experience with eyeBlog. eyeBlog bestows searchable videographic memory on the wearer. These memories can be accessed and effectively exploited by both the eyeBlog wearer and third parties. In scenario one we show how an individual with eyeBlog can mine their immediate personal history to inform their current decisions. The second scenario shows how future users of eyeBlog can recall, share and learn from the experiences of their ancestors. The final scenario shows how an eyeBlog wearer can enhance their knowledgebase.

Scenario 1: Augmenting Short-Term Memory

Maayke, is both an artist and film editor by trade. She is very busy and must manage her separate editing and portrait businesses. It is now 7:00 am and she is about to make her way to the gym for an early workout. As usual, she has misplaced her car keys. Because her keychain has an associated glyph attached to it, she knows that she can check eyeBlog to see where she left them the night before. She clicks the eyeBlog bookmark in her web browser and enters "car keys" into the search field of the eyeBlog interface. A collection of image previews appears in reverse-chronological order in her browser. No need to download the video segment, as she quickly sees that she left the keys in the ignition of her car (see Fig. 5). After her workout she stops for a cup of coffee before her big meeting with a major studio producer for a possible editing job. This is her third meeting with this producer and she really wants to make a good impression on him. She consults eyeBlog to see what transpired in their previous meetings. She constrains the search to "face-to-face," "business" and the "dates" that the meetings transpired. Another short collection of image previews appears. She clicks on the appropriate segments. From watching them she is reminded that that in their last meeting the producer was in a hurry, so he was unable to view her latest portfolio additions. She decides to quickly go back home and pick up the DVD with her most recent work.

Scenario 2: Augmenting Historical Memory

It is one hundred years in the future. Tarak and Julia are in 5th grade and have been paired together to work on a history presentation for class. They have chosen to learn about the pioneers who first colonized the Moon. Tarak is lucky because Julia's Grandfather was one of the initial colonists to migrate there. Because he was augmented with eyeBlog there is an ocean of first-hand data on the experience of being a Lunar colonist. Tarak and Julia surf through the contents of her Grandfather's eyeBlog and compile a collection of relevant links for the class presentation.

Scenario 3: Sharing Experience

Mark is retrofitting a jet engine on a commercial airliner. Although he has been an aviation engineer for over ten years, he still has a little difficulty remembering exactly how the fuel-line attaches. His colleague Owen has agreed to assist him when he reaches this critical step. Owen, who works across town has already subscribed to Mark's personal RSS feed and instructed his filter to alert him when Mark's eyeBlog registers the glyphs on the fuel-line. When the notification comes, Owen calls Mark to talk him through the procedure. After a few minutes Mark feels confident that he can finish the job on his own. He thanks Owen for his help and terminates the call. If he needs more information he can always consult Owen's eyeBlog entries that relate to the fuel-line.

FUTURE CONSIDERATIONS

Although applying glyphs is an easy and cheap way for eyeBlog to identify objects, it is not practical to augment all objects. Even if manufacturers agreed to augment all new objects with unique glyphs, there would still be the issue of untagged pre-existing objects. Object size is also of concern, as not all objects are large enough or of adequate shape to support the addition of glyphs. By adding better image recognition, eyeBlog would be able to recognize any arbitrary object that falls within the cameras field of view. This would allow a wearer who is interested in a particular object to indicate to the system that they are interested in this object, and to record any instance where the object appears in the frame of the camera. This same technology could also be used to recognize the face of interlocutors. Knowing the identity of people could be used to better categorize conversations. Adding speech recognition and a means for transcription would free the wearer of having to manually annotate eyeBlog entries. Incorporating other sensing mechanisms such as Global Positioning, electrocardiogram and brainwave telemetry in a fashion similar to Aizawa's video summary system [15] would further enrich meta-data and allow for more precise categorization of entries.

CONCERNS

eyeBlog is designed to record and publish a user's daily experiences. We have yet to incorporate a means to automatically maintain a user's privacy. As such, eyeBlog is bound to record instances that one may not want to make public. Currently users must manually delete sensitive material.

CONCLUSION

In this paper we presented eyeBlog, a personal recording and publishing system. eyeBlog automatically records events of interest. It determines interest by detecting mutual eye-gaze between interlocutors as well as registering the sightings of glyph-augmented objects. We gave example scenarios of the types of interactions and uses eyeBlog supports.

ACKNOWLEDGMENTS

We would like to thank all the students who comprise the Human Media Lab, particularly our Engineer, Changuk Sohn. David Fono and Daniel Cheng also deserve kudos for their invaluable software development skills.

REFERENCES

1. Dickie, C., Vertegaal, S., Shell, J., Sohn, C., Cheng, D., Aoudeh, D. Eye Contact Sensing Glasses for Attention-Sensitive Wearable Video Blogging. *Proc. CHI 2004*, Demo.
2. Bush, V. As We May Think. Atlantic Monthly. July, 1945.
3. Engelbart, D. Augmenting Human Intellect: A Conceptual Framework. *Research Report AFOSR-3223*, Stanford Research Institute.
4. Mann, S. Guest Editor's Introduction, Wearable Computing: Towards Humanistic Intelligence. IEEE Intelligent Systems Special Issue. Vol 16, No. 3. 2001.
5. Mann, S., Picard, R.W. A Historical Account of the Wearcomp and Wearcam Inventions Developed for Applications in Personal Imaging. *Proc. IEEE 1997*. First International Conference on Wearable Computing. pp. 66-73
6. Mann, S., Fung, J., Moncrieff, E. EyeTap Technology for Wireless Electronic News Gathering. *ACM SIG MOBILE*. Review Vol. 3, No. 4, pp 19-26. October 1999.
7. Vertegaal, R. Attentive User Interfaces. Editorial, Special Issue on Attentive User Interfaces. Communications of ACM 46(3), March 2003.
8. Vertegaal, R., Dickie, C., Sohn, C., Flickner, M. Designing Attentive Cell Phones Using Wearable Eye-Contact Sensors. *Proc. CHI 2002*, Extended Abstracts.
9. Selker, T, Lockerd, A. and J. Martinez. Eye-R, a Glasses-mounted Eye Motion Detection Interface. *Extended Abstracts CHI 2001*. Seattle: ACM Press, 2001, pp. 179-180.
10. Lockerd, A., Mueller, F. LAFCam: Leveraging Affective Feedback Camcorder. *Proc. CHI 2002*. Extended Abstracts.
11. Shiano, D., Nardi, B., Gumbrecht, M., Swartz, L. Blogging by the Rest of Us. *Proc. CHI 2004*, Late Breaking Results.
12. Herring, S., Scheidt, L., Bonus, S., Wright, E. Bridging the Gap: A Genre Analysis of Weblogs. *Proc HICSS 2004.*
13. Moveable Type website: http://www.moveabletype.org
14. ARToolkit website: http://www.hitl.washington.edu/
15. Aizawa, k., Ishijima, K., Shiina, M. Summarizing Wearable Video. *Proc. IEEE 2001*. Conference on Image Processing.