Towards Wearable Gaze Supported Augmented Cognition

Andrew Toshiaki Kurauchi
University of São Paulo
Rua do Matão 1010
São Paulo, SP
kurauchi@ime.usp.br

Carlos Hitoshi Morimoto
University of São Paulo
Rua do Matão 1010
São Paulo, SP
hitoshi@ime.usp.br

Diako Mardanbegi
IT University, Copenhagen
Rued Langgaardsvej 7
2300 Copenhagen
dima@itu.dk

Dan Witzner Hansen
IT University, Copenhagen
Rued Langgaardsvej 7
2300 Copenhagen
witzner@itu.dk

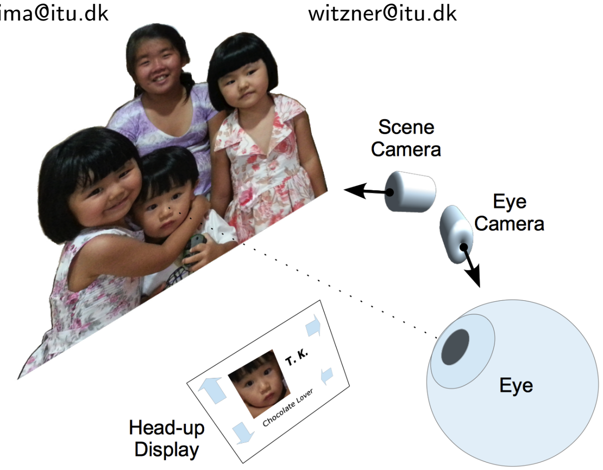

Figure 1: The user has instant and up-to-date information about a person and can interact using gaze alone, gaze and a button, and gaze and head gestures.

Abstract
Augmented cognition applications must deal with the problem of how to exhibit information in an orderly, understandable, and timely fashion. Though context have been suggested to control the kind, amount, and timing of the information delivered, we argue that gaze can be a fundamental tool to reduce the amount of information and provide an appropriate mechanism for low and divided attention interaction. We claim that most current gaze interaction paradigms are not appropriate for wearable computing because they are not designed for divided attention. We have used principles suggested by the wearable computing community to develop a gaze supported augmented cognition application with three interaction modes. The application provides information of the person being looked at. The continuous mode updates information every time the user looks at a different face. The key activated discrete mode and the head gesture activated mode only update the information when the key is pressed or the gesture is performed. A prototype of the system is currently under development and it will be used to further investigate these claims.

Author Keywords
gaze interaction; wearable computing; augmented cognition

ACM Classification Keywords

H.5.2 [Information interfaces and presentation: user interfaces]: .

Introduction

In this paper we explore how gaze interaction might enhance the usability of wearable computers by creating simpler interaction mechanisms and show how such mechanisms can be applied in applications for cognitive augmentation. But first we will discuss some design issues to better understand the benefits of gaze interaction for wearable computing applications,

Time and space are important to define the physical context but gaze may yield aspects of attention.

Wearable computing devices such as the EyeTap [7] combine a scene camera and a head-up display (HUD) to enable mediated reality, the ability to computationally augment, diminish, or alter our visual perception. The EyeTap configuration allows the camera to capture the same image as it would be captured by the eye, providing very realistic visual effects and life logging data that can be shared and used as the user's extended memory [5].

Similar but simpler configurations such as the "Memory Glasses" by DeVaul [4], may place a wearable HUD to (or instead of) the lens of the eye glasses. In such configuration, the useful display area covers just part of the visual field of view of one of the user's eyes, reducing the quality of the mediated reality experienced. Based on the announced Google Project Glass and the Vuzix M100, the next generation of smart phones is moving from mobile to wearable by using an HUD for "hands free" constant information and communication access.

Constancy is an important characteristic of wearable computers. Because the applications can always be on and available, having information popping up at any time may distract the user and become a hazard in particular situations, such as competing for (or even obstructing) the user's attention when crossing a street.

Therefore, the design of wearable applications must consider different design issues than desktop applications. In particular, as pointed out by Rhodes [10], typical WIMP interfaces require fine motor control and eye-hand coordination on a large screen, while many typical wearable computing applications are secondary tasks (e.g. reminders) or support a complex primary task. Even when the wearable application is the primary task (such as text editing), the environment might intrude and, therefore, there is a need to design for low and divided attention.

Bulling and Gellersen [1] provide a recent discussion on the current state of mobile gaze trackers and describe ways of using them in mobile applications. Due to the developments in wearable eye trackers that have just recently become more portable and easy to use, it is not surprising that there are only a few wearable computer systems that use gaze information.

For example, [2] suggests the eye movements data from a EOG eye tracker to determine context information to wearable applications, but does not use gaze information. Data input is carried out using a chord keyboard. One concrete example of a wearable augmented reality system using gaze interaction was described by Park et al. [9]. Their system relies on scene markers to position virtual objects. Gaze information is used to point and objects can be selected by dwell-time.

Because most of the work on gaze interaction has been done assuming a desktop or mobile device scenario, we discuss next different principles that can be used to design gaze supported wearable computing applications.

Current gaze interaction applications are not designed for low or divided attention.

Augmented cognition should be effortless.

Interaction with wearable computers

Because wearable computers provide support while the user is performing other activities, freeing the hands (or at least one hand) from computer interaction is an important feature. Typically, chord keyboards are used as input devices with the HUDs. Though chord keyboards can be very efficient for data entry, becoming an efficient typist might required a great effort [4]. To overcome this difficulty, speech and hand gestures have also been used.

Due to its ability to augment and mediate reality, wearable computing applications can provide support for complex real-word activities, with applications areas such as military, education, medical, business, and many others. But as identified by many wearable computing researchers, augmented cognition applications will be a major factor in the development of wearable computers. Augmented cognition applications can help the user to perform mental tasks. Because wearable computers are always on and available, they can be incorporated by the user to act like a prosthetic and become an extension of the user's mind and body.

Examples of augmented cognition applications are described in [6, 4]. Mann [6] gives examples of how diminished reality, i.e., removing clutter from the scene such as advertising and billboards, can help the user by avoiding information overload. The use of an EyeTap facilitates the substitution of planar patches of the scene by virtual cues. Another possible application is to place virtual name tags on each person within the field of view.

DeVaul [4] proposes the use of software agents to provide just-in-time information based on the user's local context. Using an HUD with a chord keyboard, his system (called Memory Glasses) was able to present short text messages on the HUD related to personal annotations typed using the chord keyboard, helping the user to remember related issues stored in the system. Today, with the current state of mobile computing, related information could be searched in the Internet.

During the development of the Memory Glasses, DeVaul [4] defined the following principles of low-attention interaction for wearable computing:

1. Avoid encumbering the user, both physically and perceptually, referring to the hardware, peripherals and interface.
2. Avoid unnecessary distractions, by minimizing the frequency and duration of the interactions, and using appropriate context information.
3. Design interfaces that are quick to evaluate, so the user, even when interrupted, is always in control.
4. Simplify the execution as much as possible, but no further. Easy things should be easy, hard things should be possible.
5. Avoid pointers, hidden interface states, non-salient interface changes, and never assume the wearable interface has the user's undivided attention.

These principles will be used in the design of a cognitive augmented application for memory aids, described next.

Gaze Supported Augmented Cognition

The "extended mind" conjecture of Clark and Chalmers [3] states that not all cognitive processes are in the head. The claim is based on the idea of epistemic actions, i.e., actions that alter the world to help cognitive processes. Because gaze, attention, and cognitive processes are so interrelated, it seems natural to use gaze information to automatically filter, control, and mediate the contents of wearable computing application, but the description of actual systems combining both gaze and wearable technologies are still rare in the literature. As an initial effort to combine previous experiences from both areas, we have followed the principles proposed by DeVaul [4] for low-attention interaction, to design three interaction modes for a gaze supported augmented cognition applications.

The objective of the application is to provide the user information about the person currently being observed.

The objective of the application is to provide the user information about the person currently being observed, similar to the automatic name tag application proposed by Mann [6], but using a simpler setup. The basic components of the system are shown in Figure 1. Two cameras are required for the wearable gaze tracker, one pointing to the scene and a second looking at the eye. An HUD is used to display relevant information to the user. Observe that it is also possible to use gaze information for interaction with the HUD.

Due to the low resolution screen of the HUD, when multiple people are seen by the scene camera, presenting information about every person at once might be confusing, since it might be difficult to associate a name to a given face. Following DeVault's first principle, to avoid encumbering the interface, our system is designed to provide information about a single person at a time, corresponding to the face being looked at.

To minimize the frequency and duration of the interactions, the information about the person can be updated every time the user's gaze lies on a new face. We will call this interaction mode continuous (C). Because the information is always presented in the same location on the HUD, this information can be easily ignored by the user.

We are also developing a discrete (D) mode, that updates the information on the HUD after a key press to determine if continuous updates are distracting. A third discrete mode controlled by head gestures (G) is also being developed. The head gesture mode allows for completely hands-free operation, while not overloading the eye with a control task. Because the head can perform simple gestures independently of the eye natural behavior, head gestures are more appropriate than eye gestures for wearable computing.

These three modes follow the simplicity of execution principle for the task of associating names to faces. For more complex tasks, e.g., to show more information about the person, the D and G modes could facilitate the interaction because they can be easily extended, using a double click or a different yet simple head gesture. Because the HUD can also be used for gaze interaction, a point and click (or point and gesture) interface will also be developed.

For the continuous mode, dwell-time and eye gestures could also be used for interaction with the HUD, but because these interaction modes would require longer interaction times and require full attention of the user, they would not be appropriate. Also for the C mode, to avoid the information to change when the user is looking at the HUD in case a person is positioned in that direction, its region is masked out, so no face is detected within the HUD region. As pointed by DeVaul's, context information could be used to improve the quality and timing of the information, and it should clearly be considered in a real application. The use of context information is not though the focus of this paper.

DeVaul's third principle states that the interface should be quick to evaluate. Designing for divided attention also

Figure 2: Low cost wearable head mounted eye tracker.

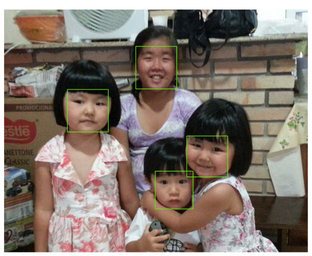

Figure 3: Faces detected using the Viola-Jones algorithm.

requires the user to be reminded of the last face seen, in case of distraction. Therefore the information is presented with a cropped region computed automatically by the face detector algorithm, showing the detected face. This feature also allows the user to avoid detection errors by the system. The last principle is a list of things to be avoided and that has been followed by our design.

System implementation
Figure 2 shows a low cost wearable head mounted eye tracker being used in our experiments. It uses two USB webcams, one pointing towards the scene and the second looking at the eye. The eye camera has two IR leds to provide robustness to illumination conditions. Both cameras are mounted on a baseball cap. The gaze tracking software is based on the open source Haytham gaze tracker (available at eye.itu.dk), that has been ported to run on a Linux platform. A 4 point calibration is used to compute a homographic transformation.

The wearable gaze tracker has not been integrated with an HUD yet, so the proposed memory aids methods will be demonstrated on videos projected on a large screen. The projected videos will be scaled to show the faces close to their actual size. Though this is not an ideal situation, we expect the video to cover the field of view of the user, so the HUD display can be simulated as part of the projected screen, and placed somewhere on the lower left of the video.

Faces are automatically detected using live video from the scene camera of the wearable gaze tracker using the Viola-Jones algorithm [11]. A result from this algorithm is shown in Figure 3. Once the user's gaze is detected within a face region, an estimator based on our gaze-to-face mapping algorithm is used to recognize the face and information about the person is displayed according to the current interaction method (C, D, or G). For the D mode, the left button of a wireless mouse is being used.

For the recognition of head gestures in the G mode, we are using the method introduced by Mardanbegi et al. [8]. Their method uses a combination of head gestures and a fixed gaze location for interaction with applications running in large displays and small mobile phone screens. Because the head gestures are estimated directly from the eye movements without the need of extra sensors such as accelerometers, the whole gaze interaction system can be made very light and comfortable to wear, as seen in Figure 2. Figure 4 show two images of the eye when fixating at a target and performing a vertical head movement (initially down and moving upwards, while looking forward). When a user keeps the gaze on a specific target, the vestibular-ocular reflex makes it possible to measure head movements because the eye moves in the opposite direction of the head. Therefore, head movements are measured indirectly from the eye movements detected from the eye camera.

Conclusion
A typical wearable computing application is always on and available, so it must be designed for divided attention. Gaze based applications, on the other hand, have been mainly developed for desktop computing. Therefore, the direct port of gaze based applications to wearable computing is not recommended since gaze and attention are so interrelated. More importantly, the use of most gaze interaction paradigms, such as dwell-time and gaze gestures are not appropriate for wearable computing, since they not only require full attention by the user to interact, but they misappropriate the natural behavior of the user's gaze.

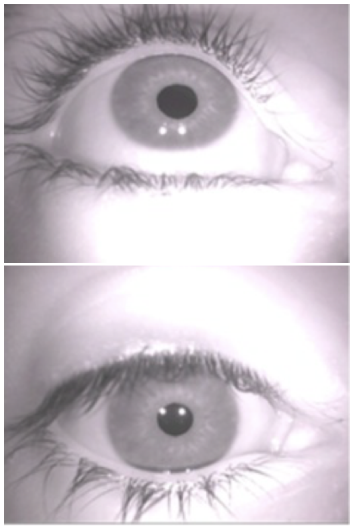

Figure 4: Images of the eye when looking at a target and performing a head gesture.

Nonetheless, we do believe gaze can revolutionize the way we interact with wearable computers. For that purpose, we have described our on going research on wearable gaze supported augmented cognition. By applying design principles learned from the wearable computing community, we proposed three gaze-based interaction modes that are appropriate for low and divided attention. The continuous mode updates information at every new event (such as looking at a different face), a key activated discrete mode, and a head gesture activated mode.

Though speech and gestures have also been used to interact with wearable computers, gaze interaction offers more privacy and discreteness and we expect it to offer faster interaction (though not faster than a chord keyboard, but definitely easier to learn). Maybe the most important characteristic is that gaze can potentially be used to interact with scene objects (with the help of computer vision algorithms), besides the head mounted display. A prototype of the system is currently under development and it will be used to further investigate these ideas.

References

[1] Bulling, A., and Gellersen, H. Toward mobile eye-based human-computer interaction. *IEEE Pervasive Computing 9*, 4 (2010), 8–12.

[2] Bulling, A., Roggen, D., and Tröster, G. Wearable eog goggles: eye-based interaction in everyday environments. In *CHI Extended Abstracts* (2009), 3259–3264.

[3] Clark, A., and Chalmers, D. The extended mind. *Analysis 58*, 1 (1998), 7–19.

[4] DeVaul, R. *The memory glasses: wearable computing for just-in-time memory support*. PhD thesis, Massachusetts Institute of Technology, April 2004.

[5] Ishiguro, Y., Mujibiya, A., Miyaki, T., and Rekimoto, J. Aided eyes: eye activity sensing for daily life. In *Proceedings of the 1st Augmented Human International Conference*, AH '10, ACM (2010), 25:1–25:7.

[6] Mann, S., and Fung, J. Videoorbits on eyetap devices for deliberately diminished reality or altering the visual perception of rigid planar patches of a real world scene. In *Proceedings of the Second IEEE International Symposium on Mixed Reality* (2001), 48–55.

[7] Mann, S., Fung, J., Aimone, C., Sehgal, A., and Chen, D. Designing eyetap digital eyeglasses for continuous lifelong capture and sharing of personal experiences. In *Proc. CHI 2005 Conference on Computer Human Interaction* (2005).

[8] Mardanbegi, D., Hansen, D. W., and Pederson, T. Eye-based head gestures. In *Proceedings of the Symposium on Eye Tracking Research and Applications - ETRA'12*, ACM Press (2012), 139–146.

[9] Park, H. M., Lee, S. H., and Choi, J. S. Wearable augmented reality system using gaze interaction. In *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, ISMAR '08, IEEE Computer Society (2008), 175–176.

[10] Rhodes, B. The wearable remembrance agent: a system for augmented memory. *Personal Technologies Journal Special Issue on Wearable Computing , Personal Technologies*, 1 (1997), 218–224.

[11] Viola, P. A., and Jones, M. J. Robust real-time face detection. In *ICCV* (2001), 747.