

Using Eye- and Gaze-tracking to Interact with a Visual Display

Slavko Milekic
University of the Arts
320 S. Broad St. Philadelphia, PA 19102
smilekic@uarts.edu

The analysis presented covers three broad areas of implementation of eye- and gaze-tracking (with the description of current research and functional prototypes):

- 1) **Gaze-tracking for mobile devices:** Recent advances in eye- and gaze-tracking make it possible to use Web-cams or smart phone cameras for eye-tracking.
- 2) **Gaze-tracking in a museum/gallery context:** Ability to track visitors' gaze direction offers a wide range of new possibilities for enhancing visitor experience in a museum. Just focusing (looking) at a detail of a painting a visitor can gain more detailed information. Collecting visitors gaze direction data also opens up a whole new area of research.
- 3) **Gaze-tracking for medical information display:** Most of the information in medical imaging field already exists in digital format (PET, CAT, MRI scans). This information is crucial when performing surgery and so far has been delivered by hand. The use of a gaze-sensitive display would allow a surgeon to access relevant information just by looking at a display.

The author of the paper holds US patent for "interacting with visual display using eye and gaze-gestures" (US 7,561,143 'B1).

Eye-tracking. Gaze-tracking. Visual display. Mobile devices. Museum. Medical applications.

1. INTRODUCTION

Evolutionarily speaking, humans, unlike most of other mammals, are biologically hard-wired to track each others gaze direction. This is evident even in the biological "design" of the human eye. Our pupils and corneas are surrounded by a visible white area (sclera) which makes gaze direction detection possible even in the absence of other directional signs, like head turning (see Figure 1).

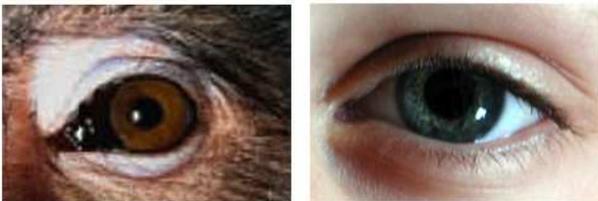


Figure 1: Difference between animal and human eye

As a method for collecting data, eye-tracking has been around for a surprisingly long time – more than seven decades (Yarbus 1967). However, it was only in the past two decades, and fuelled by technological advancements, that its use has expanded into a wide variety of areas ranging from

Web page design, marketing research, psycholinguistic studies to art projects.

This paper, rather than being a comprehensive overview of the technologies used, will focus on three emerging areas that have the potential to dramatically change the way we interact with visually presented information using our gaze. These are:

- the use of eye- and gaze-tracking for interaction with mobile devices (smart phones, tablets, MP3 players);
- the use of eye- and gaze-tracking in a museum/gallery context; and
- the use of gaze-tracking for interactions with visual displays in the field of medicine.

Although these areas of application may seem very different, they are brought together by the same underlying technology and interface solutions.

2. EYE-TRACKING FOR MOBILE DEVICES

With the advent of the iPhone and iPad by Apple Inc., we have witnessed dramatic advancements in the support for diverse user interactions with visual

displays. These range from simple awareness of the display regarding its orientation to detection of user movements and gestures, sensitivity to geographic location, and most recently, AI-guided speech recognition. Based on recent collaboration with (and perhaps partial acquisition of) Tobii, the largest European eye- and gaze-tracking equipment manufacturer by Apple Inc., as well as recently filed Apple patent (United States Patent Application 20120036433 2012) this paper examines the next logical step in the evolution of modes of interaction with digitally displayed information. The author anticipates and provides an analysis of the next revolution in interaction with digital displays in which, the display will not only “know” its orientation or location, but will also “know” exactly what a user is looking at.

2.1 Problems with mobile eye-tracking

Eye- and gaze-tracking with hand-held mobile devices introduces a number of problems specific to this area. A number of these are identified in a recent paper (Miluzzo, Wang & Campbell 2010). Broadly speaking, one can divide these problems into ones that arise from the specific context of use of mobile devices, and the ones that are related to technological constraints imposed by the small size of devices.

2.1.1. Variable distance & angle

Stationary eye-tracking equipment (like the models produced by Tobii, EyeLinc, Mirametrix) is usually already built into the display bezel, or can be affixed to it. In this case the viewing angle and the distance from the monitor are relatively stable and allow precise calibration and tracking. However, in normal use of mobile devices, the viewing angle and the distance from the display can dramatically change even within the small period of time, for example, while walking, gesticulating or changing position.

2.1.2. Variable ambient light

Changes in the ambient light (for example, stepping outside from a room) significantly affect accuracy of eye position, especially when eye-tracking is based on normal video signal processing.

2.1.3. Technological constraints

Hand held mobile devices pose a significant challenge in terms of miniaturizing the exceedingly large number of components and sensors. So far the industry has responded quickly and modern smart phones have built-in miniature cameras, accelerometers, gyroscopes, etc. Yet, introduction of another layer of interaction (like eye-tracking) would require either miniaturization of the existing components, an approach that seems to be taken by Tobii (Tobii IS-2), or entirely new way of detection, for example by using the whole display

as a camera (United States Patent 20060007222 2006).

2.1.4. Interface problems

While it is (now) relatively easy to track one's eyes using gaze as a “pointing device” is still problematic. The main problem is in distinguishing between two modes: a) when a viewer is just looking at something (observation mode), and b) when a viewer uses gaze to initiate actions (control mode). Historically, the solution to the problem was to use either a time-based approach (focusing on something for a predetermined period of time would trigger an action) or location-based (focusing on a certain location would define the control function of gaze fixation). Nearly ten years ago I proposed yet another way of solving the problem by using “eye-gestures” (Milekic 2003).

3. EYE-TRACKING IN A MUSEUM CONTEXT

With the dramatic drop in the price of technology, increased processing speeds of modern computers and more sophisticated (often open-source) software the use of eye- and gaze-tracking in a museum and gallery context has suddenly become a real possibility. In the following section I will outline several possible development directions as well as a number of art projects using this technology. Although eye-tracking can be achieved in a variety of ways, for example by wearing special goggles (Bulling, A., Roggen, D. & Tröster, G. 2008) in this section I will focus mostly on non-intrusive head-free technologies.

3.1 I know you are looking at me

The most basic form of eye-tracking involves just detecting one's gaze – that is, how many eyeballs are turned towards an object or a surface. A relatively cheap technological solution called EyeBox already exists. The advantages of this technology are that a number of simultaneous inputs can be detected, and the detection is effective at distances (currently) of 10 feet or so. At present this approach is mostly used for collecting data of how many eyeballs looked at certain advertisement (or a portion thereof) or to make an advertisement “aware” of gaze fixation. Amnesty International has made the first “gaze aware” advertisement to raise awareness of domestic abuse. This technology has also made possible a number of interactive artistic projects which tie a viewer's visual attention to an event (Knep 2003).

3.2 I am recording your gaze

Most eye-tracking applications are currently used to record the pattern of gaze-fixations and extract information about viewer's interest. While this would definitely be valuable information for visual

perception studies there are ways of making these recordings more meaningful and interactive for museum visitors.

Velichovsky developed a concept of “attentional landscape” (Challis & Velichovsky 1999) some time ago. Attentional landscape corresponds roughly to the area of foveal vision including an adjacent area of peripheral vision, while the rest of the visual field is blurred. Using this concept I developed an application (Basset & Milekic 2001) with a goal to recreate a personal experience that one had in a physical museum and somehow make it available on a museum Web site. The application allowed a viewer to explore a painting using the “attentional landscape” while the program recorded in real time the visual exploration path (Figure 2).



Figure 2: Partial “attentional landscape” path while examining the painting

When a viewer was satisfied with the exploration and had seen enough to form a personal opinion s/he was asked to write a personal essay about the painting and post it on the museum Web site. However, uploading the essay on the Web site also uploaded the individual “attentional landscape” exploration path. As a part of an on-line museum course the idea was to present a museum Web site visitor with the original painting that could be explored *only* using a recorded exploration path from the *previous* viewer, by providing exactly the same visual information from the exploration sequence of the original viewer. Knowing that the second viewer was exposed to exactly the same visual information, s/he was also asked to write a personal essay on her/his experience. After submitting the essay the second viewer could see the essay (personal experience) of the first viewer. This created an interesting dynamic, comparing one’s own experience with the experience of another person who had exactly the same (visual) experience. Although in the original project movement of attentional landscape was tied to mouse/cursor movement, later on I connected the original image exploration to gaze direction, and thus created a truly personal recorded “view” that

could be made available to another person. In essence, one person could “see” through another person’s eyes.

One can imagine further elaborations on this topic by allowing museum visitors to share their own visual exploration experience (via a Web link) with their family and friends.

The concept described above was pushed further into the direction of physical experience in a museum by two of my students (MacDuffie Woodburn & Miller 2010). They created a physical device (Figure 3) that literally allowed one person to see what the other person was “looking at”. The original idea (although it was not implemented in the physical prototype) was that the device used by the person who was “looking” would also track gaze direction and create a visual “attentional landscape” by (visually) emphasizing the focus of the gaze while blurring the rest of the visual field for the person who was “watching”. Even though in the original prototype the two units were physically connected, it would be easy to connect them wirelessly or, for that matter, have a “looking” person in a physical museum and make the visual feed available on the Web, or via a mobile device.



Figure 3: “Pixel” prototype showing the use of two units, one “seeing” and one “looking” through another person’s eyes. Very soon the gallery visitors discovered that they can make another person “look at themselves” by pointing their unit in the right direction

3.3 I am the voice in your head

Remote and non-intrusive eye-tracking can also be used in a museum context to deliver additional information about particular detail that attracted a viewer’s attention. A prototype of such an installation is depicted in Figure 4. The artefact is a large mural depicting scenes from Gandhi’s life (A). Under the mural is the location of camera (B) with

telescopic lens which is focused on the oval window area of the semi-transparent stand (C).

was achieved with the use of a ceiling mounted computer projector (E). The reason for this is that the entire setup is controlled by a (hidden) computer which is both analysing viewer's gaze

fixations and displaying them in real time on a visual display in the form of a spot light. In this case the visual display was a projection matched in scale to the observed artefact.

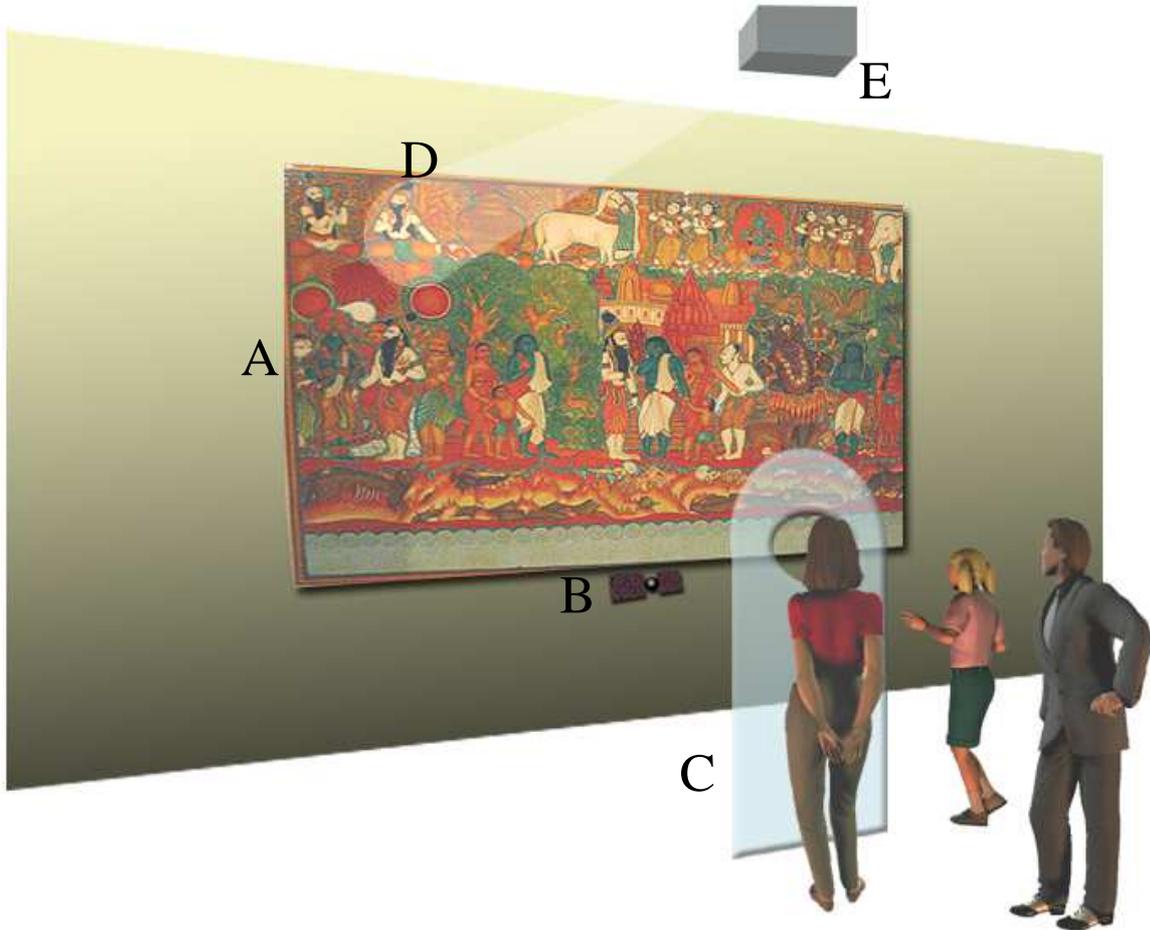


Figure 4: Audio delivery based on eye-tracking data (see explanation in text)

On each side of the camera are infrared LED light emitters (invisible to human eye) which make eye-tracking more efficient even in changing light conditions. The stand fulfils several functions: first, it indicates the right position for eye-tracking enhanced viewing. Second, it “frames” and positions a visitor's head in particular location (that the camera is already focused on) making it possible to track eyes of different visitors without the need for individual calibration. Third, the frame has built-in speakers for delivering the audio relevant to a viewer's focus of attention.

The whole process is made transparent to a visitor by tying the viewer's gaze direction to a spot light (D). In this particular prototype, the spot light effect While the technical description may sound complicated a visitor's experience is almost magical. Just by stepping to the “window” a spot of light is tied to observer's gaze, and whatever one

focuses on, a “voice in the head” provides additional information.

A functional prototype of such an application with a laptop equipped with (albeit primitive) eye-tracking capability was recently demonstrated at a Museums & Web '10 conference in Denver (Milekic, Roberts & Miller 2010). For the purposes of demonstration, a laptop screen displayed an actual painting from a museum, while providing visual feedback to the viewer about their gaze focus with a magnifying glass (thus, magnifying any part of the image that the viewer was looking at).

Figure 5 depicts a visual exploration path of an actual viewer. The red dots indicate points of visual fixations (with their size indicating the duration of fixation), and the balloon captions show the voice-overs that were triggered while focusing on a particular detail.

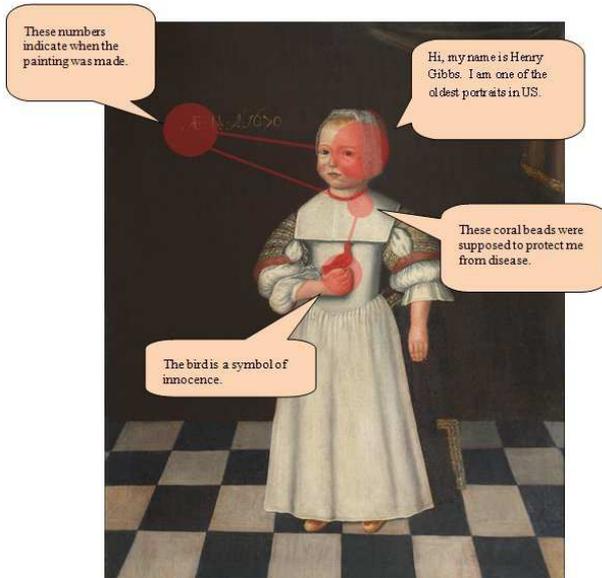


Figure 5: A recorded visual exploration path superimposed on the painting

4. EYE-TRACKING & MEDICAL VISUALISATION

In general, interacting with a visual display using eye- and gaze tracking can be of enormous practical value, especially in situations that require hands-free access to vital information. To the author of this paper, an artist who also has a medical degree, it was natural to explore the potential uses of eye-tracking in the field of medical visualisation. Nowadays, most of the medical diagnostic tools that produce images (CAT scan, MRI, fMRI, PET scan) actually create series of *digital images* that can be (and often are) displayed on a flat screen monitor. These images are also used for planning of major surgical procedures, and surgeons study them before an operation and often have them displayed during the actual procedure. In order to do this, they have to rely on assistants and request a certain view of the scan to be displayed. Now, imagine if a neurosurgeon could, just by looking at a display, pick and choose the most relevant image. This was exactly the goal of the described prototype, which eventually became a US patent (US patent 7561143).

4.1 Eye gestures and visual control

As described before, one of the main problems in using gaze-tracking for interacting with a visual display is distinguishing between merely looking at something (observation) and looking at something with the intent to initiate action. Traditionally the problem was solved either by using time mechanism (if one focuses on certain control/button for a period of time, usually 200-400 milliseconds, it triggers an action), or by using location (similar to choosing different “tools” from a

tool palette in Adobe Photoshop). These approaches work well if the cost/benefit ratio is positive. For example, a person with amyotrophic lateral sclerosis (ALS) or a similar condition, who are paralysed and can only move their eyes would find it extremely beneficial to be able to control the cursor on the screen with their eyes and type on a virtual keyboard. However, for able bodied person having to wait even a fraction of a second before any action is triggered is excruciating and quickly abandoned.

Trying to solve this problem, in an earlier publication (Milekic 2003), I proposed the use of eye gestures as a “clutch” mechanism for switching between the “observation” and the “control” mode. The idea for this solution came from two observations:

- Eyes are the fastest moving body part we have. Eyes move in a series of “jumps” called saccades and the velocity of saccades can be as large as 1000 deg/sec.
- Eye gestures are naturally used in everyday communication. In group communication we almost exclusively use gaze-direction as a pointing mechanism.

Thus, the proposed solution was to use simple eye-gestures (briefly shifting the gaze to the left, right, up or down) as a “control vocabulary” (Figure 6, Milekic 2003).

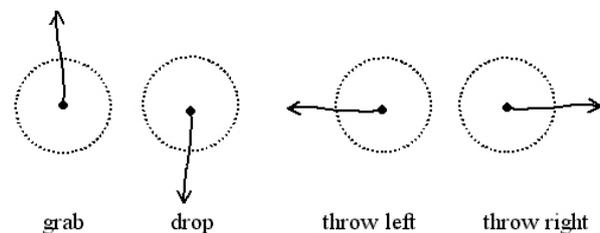


Figure 6: Eye-gestures mapped to different actions

Of course, depending on context, it is entirely possible to map these gestures to different actions. For example, in one context, looking briefly to the left (or right) would “flip” the page of the book on the screen. In another context the same mechanism can be used to zoom in or zoom out of particular area.

4.2 Medical application prototype

The following prototype describes an application of eye-tracking that would allow a medical professional to interact with a visual display hands-free using only eye-gestures.

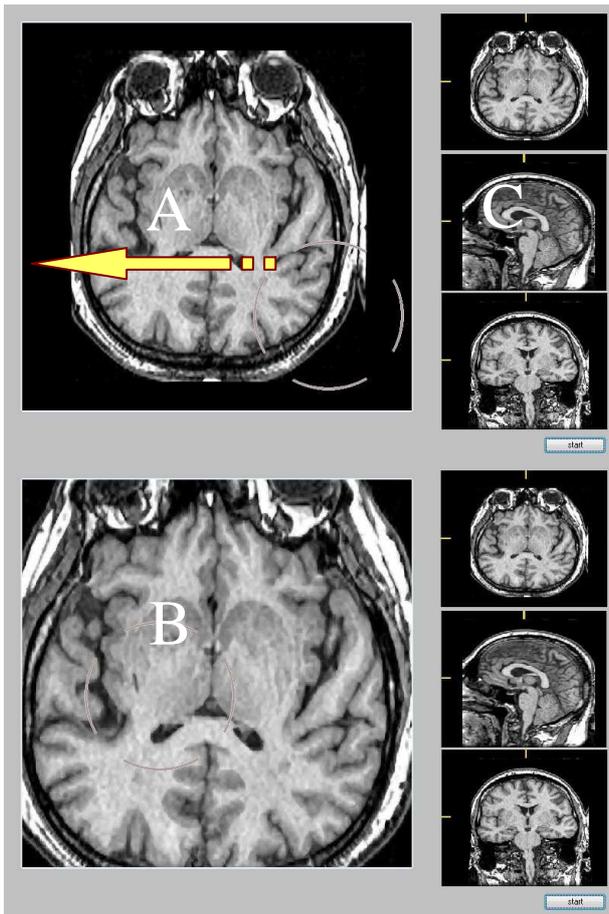


Figure 7: Successive monitor screenshots illustrating eye-gesture looking briefly to the left, which is mapped to zooming into the image

Figure 7 illustrates hands-free, gaze controlled application used in a medical context. When a surgeon looks at the screen a visual feedback is provided (dashed circle at the base of arrow A) to indicate that the image is “gaze-aware”. Looking briefly to the left, a surgeon is able to zoom into the image (B). Consequently, shifting the gaze quickly to the right, zooms out of the image. Additionally (not illustrated) looking briefly upwards “grabs” the image and moves it with the gaze direction making it possible to reposition it, while quickly shifting the gaze downwards “drops” the image at the new location. The interface uses the fact that while focusing on the details of an image, fixation points are relatively clustered, so the software is capable of recognizing a rapid shift in gaze direction as a mode changing trigger. Therefore, a normal observation of the image would not trigger any actions.

The proposed solutions combine gesture- and location-based control mechanisms. The panel on the right of the screen contains different views (frontal, lateral, sagittal). If, at any point in time, a surgeon needs different point of view, s/he would just have to focus on any of the icons on the right for a brief period of time, and the large image would

display the same area from a different view. Having to “dwell” on an icon to trigger an action, makes them immune to normal observation pattern.

5. CONCLUSION

Although it has been around for a long time, due to recent technological advancements, the use of eye- and gaze-tracking has spread across wide variety of fields. It started as a specialized technique for visual perception studies, then it found its use in marketing and Web design, and more recently as an aid to special needs populations. The use of this technology in art projects and interactive installations indicates that it has “matured” and became accessible to wider segment of population.

Current laboratory demonstrations and described applications indicate that we are coming closer and closer to the future where everyday objects and our environment will be practically able to read our minds and intentions, hopefully for greater good. I am looking forward to this future.

6. REFERENCES

- Amnesty International eye-tracking poster. <http://www.brandinfection.com/2009/07/24/it-happens-when-nobody-is-watching-amnesty-international/> (retrieved April 3, 2012)
- Basset, B., Milekic, S. (2001) Recreating the Gallery Experience in Interactive Media, *NAEA Convention*, March 14-18, 2001, New York
- Bulling, A., Gellersen, H. (2010) Toward Mobile Eye-Based Human-Computer Interaction, *Pervasive Computing, IEEE*, 9(4), pp. 8-12
- Challis, B.H. & Velichovsky, B.M. (eds.) (1999) *Stratification in Cognition and Consciousness*, Amsterdam and Philadelphia: John Benjamins
- Knep, B. (2003) Big Smile, available at: <http://www.blep.com/bigSmile/index.htm>
- MacDuffie Woodburn, J., Miller, M. (2010) Pixel: Seeing through other Peoples' Eyes, MID Studio project, University of the Arts, Philadelphia, available at: <http://www.mid-uarts.org/node/251>
- Milekic, S. (2003) The More You Look the More You Get: Intention-based Interface using Gaze-tracking In Bearman, D., Trant, J. (eds.) *Museums and the Web 2003: Selected papers from an international conference*, Archives and Museum Informatics, Pittsburgh
- Miluzzo, E., Wang, T., Campbell, A. (2010) EyePhone: Activating Mobile Phones With Your Eyes, *Second ACM SIGCOMM Workshop on Networking, Systems, and Applications on Mobile Handhelds (MobiHeld'10)*, New Delhi, August 30, 2010

United States Patent 7561143
<http://www.freepatentsonline.com/7561143.html>
(retrieved April 3, 2012)

United States Patent 20060007222 2006
<http://www.freepatentsonline.com/y2006/0007222.html>
(retrieved April 1, 2012)

United States Patent 20120036433 2012
<http://www.freepatentsonline.com/y2012/0036433.html>
(retrieved April 1, 2012)

Yarbus, A.L. (1967) Eye movements during perception of complex objects, in L. A. Riggs, (ed.), *Eye Movements and Vision*, Plenum Press, New York