PIXEL

Eye Tracking Research from the MID LABS

University of the Arts, 2010

John Paul MacDuffie Woodburn

Matt Miller

How Can You See What I See through the overabundance of visual in today's world? Is there an advantage to blurring the visual noise in your peripheral view? Could we experience other peoples' focused point of view? Furthermore, is there a way to take this to the next step by seeing from other peoples' perspectives? Looking from the inside out, is this really possible?

The Tech: Eye Tracking

These questions of vision and perspective came from the technology we were researching: Eye tracking works when infrared (IR) light is absorbed by the pupil, recorded with a web camera, and processed with eye tracking software. IR light increases the contrast between the pupil and the rest of the eye without blinding the user. Software then traces the outline of the pupil and locks onto its movements. Calibration maps these movements to points on the computer screen.

The essence of eye tracking has been around since humans were able to look each other in the eyes. You see, without any fancy equipment you can tell, with a decent level of accuracy, where someone is looking. Eye contact in conversation provides clues to whether or not the other person is paying attention and their level of interest. The pupil size (the round black dot in your eye) can indicate an emotional state is present, although not necessarily what that emotion is [1].

As early as 1878, eyes were being tracked, albeit with very invasive methods. Small needles would be affixed to the eye to act as pointers (not stuck in the eyes, but adhered on), indicating the general direction of the subjects eye gaze. Slight advancements could be seen in 1901, when reflected light and mirrors were used to track horizontal movement only. Of course, with the advent of computers around 1970, eye tracking made great advancements. Real time eye tracking and data collection was available, and much of the efforts to this day focused on making the technology smaller, more compact, more accurate, and less invasive for the user [2].

The 1990's gave life to a number of home-brewed eye tracking software solutions. The availability of high quality webcams for a moderate cost made eye tracking hardware affordable to a larger audience (Compared to big research labs and institutions.) At the same time, complex software for computer vision and object tracking was being developed [3, 4, and 5] and released under GNU license agreements, meaning anyone was free to use it. The ITU gazetracker is one example of this type of software. Developed by the IT University of Copenhagen in Netherlands, their software is released free to the public and requires only a basic webcam to perform eye tracking.

Recently the EyeWriter project, http://www.EyeWriter.org, has launched a new wave of eye tracking, DIY and MIF (make it fun.) The EyeWriter initiative is a collaboration between Free Art and Technology (FAT), OpenFrameworks and the Graffiti Research Lab. It aims to give new life to graffiti artist Toby Quinn. Quinn, paralyzed by ALS, is able to use the software to create complex tags that are projected across NY. The software is built using openFrameworks, an open-source C++ library for creative coding.

Projections as to the future of eye tracking as a viable technology are solidified when you consider Apple's investment in the technology [6]. They recently entered into a business deal with major eye tracking manufacturer Tobii and have a number of eye tracking related patents. The new iPhone 4G, with its forward facing camera, could be a perfect device to launch this new technology. Surely an Apple eye tracking device would clear the way for widespread adoption, just as they did with the touch-screen interface.

Eye Tracking Calibration

Today much work is being done in the area of calibration, a process necessary to map eye movements to screen position. The process used by the majority of eye tracking systems requires the user to follow a series of points on the computer display. The eye position and other data such a head and glint locations are recorded and processed by software. There are limitations to this method, for example requiring the user to sit through an extra process before being able to use eye tracking capable devices. Some solutions we investigated were a method that places IR lights in a pattern under the screen [7]. Software can then recognize the pattern and since it originates from a constant location, head movement can be mapped and calibration made more accurate.

For mobile devices the calibration process is a greater issue, since it would need to be repeated each time the phone is removed and looked at by the user. A proposed solution would be to reduce the calibration points to two, one in each opposing corner [8]. This would reduce the time required, however it would still need to be repeated each time the device was used. Further problems could arise from glancing at interactive elements that you did not necessarily want to select, for example a contact in your address book. These misunderstood selections would be confirmed with a secondary confirmation screen, all but negating the efficiency of eye input.

However, even with proper calibration the eye is not an efficient input device. Quick and involuntary eye movements called saccades are always occurring but are mostly ignored by our brain when it processes visual information. In the case of a menu system, even if you are looking at menu option A it's likely your eyes move to options B and C without you even knowing. This greatly complicates using the eye as an input device. Shumin Zhai notes, in his paper that "...the eye, as one of our primary perceptual devices, has not evolved to be a control organ," [9]. In an interview with Dr. Dario D. Salvucci from Drexel University's Department of Computer Science, he noted that we should "exploit what the eyes do naturally," for example looking at a landscape and scanning for objects [10]. The eye does not afford selecting objects or acting as an input device.

I can show you what I see... but it's hard to convey exactly how I see it.

I can paint you a picture, take a photo, or show you a home movie -- all in an attempt to show you what I see. Yet it's not possible to show you this imagery the exact way I see it. Factors such as field of vision and the way our brain processes visual information make each individual visual experience unique. Field of vision is the range in which our eyes take in information and spans approximately 120 degrees of arc [11]. Most of this arc, however, is peripheral vision that our brain disregards. The fovea is an area approximately 6 degree in the center of this arc that has the sharpest detail of all that we see. Data from the fovea alone is processed using 50 percent of our visual cortex while the remaining half of our brainpower goes to processing all remaining areas of vision [12]. This raises the question: why is foveal vision so important? If we remove peripheral vision will more brainpower go to processing the foveal information?

Visual Noise: Everything is competing for your vision

What is visual noise? Urban areas are filled with noises that disturb the masses, both auditory and visual. Is this why urbanites tend to dream of the peaceful countryside? Perhaps, but why people flock to the intense sensory experiences such as Times Square in New York City? Individual States and cities in the United States have been banning digital billboards: Vermont, Maine, Montana, Durham North Carolina, Knoxville Tennessee, Denver Colorado, San Francisco California, St. Petersburg Florida, Pima County in Arizona, several cities in Texas, [13]. Even the fourth largest mega city, Sao Paulo in Brazil, in the world outlawed billboards [14].

Subliminal advertising does leave a mark on the brain [15]. Yet, just as quickly new digital roadside billboards are appearing everyday. The overflow of visual noise is impossible to avoid. What effect does this have on people, and what happens if you remove that peripheral distraction? Could blurring the peripheral vision of people allow people to not feel over whelmed in these visually populated cities? The peripheral vision is already naturally blurred. So, if the peripheral vision were further blurred, would this null the surrounding visual noise? There are physiological questions and qualms when mentioning the idea of pixilation

Our Solution: Pixelate peripheral noise, so you can better see my point of view.

So, essentially we have a problem (how can I see what you see?) and an observation (increasing digital noise). A solution to relieve the observation (by blocking peripheral noise) leads us to a solution for our problem (seeing what you see). The answer is that we can block out peripheral vision in order to communicate what we are looking at in a way similar to how our brain processes it.

The PIXEL connects people through their eyes. One is able to share their viewing perspectives with others in real time. The peripheral view is pixilated to help the other user focus on exactly what the original user to focusing on (i.e. their visual narrative). The one user will be able to reduce the cognitive load by being able to exactly see where the one user is focusing. This can drastically reduce the cognitive load of the one watcher user since he/she will be able to physically understand how the looker user is visually thinking. At the same time, the looker user will reduce their cognitive load by being alleviated of the process of describing their visual perspective to an audience. The genius of the PIXEL is in the social interactions it creates.

Finding an Audience

We wanted to make something exciting and interactive for the gallery exhibition. We wanted people to say, "yea, eye tracking is cool, I want to know more." Part of the challenge with a new tech, especially on with such big brother connotations as "I can see where you are looking" is (leveraging a bias??) finding a way to get people to accept it. Based on early feedback, this was successful (see appendix a).

Now that we have a working prototype of the PIXEL, the next step is to put it in the hands of people and find the ideal audience. We have a few in mind, but there are sure to be unexpected uses that we have not imagined. We believe that commercial applications are driven by the Pixel's ability to spark unique social interactions and inspire dialogue. For example, in a museum application the tour guide could share their view of the artworks with visitors in real time. Imagine seeing exactly what your guide is seeing while they explain a Picasso or Monet.

Imagine physically handicapped people who would not be able to physically experience various expects of life but still have vision, with the PIXEL a handicapped person would be able to visually experience the physical experience that is going on. Visually over stimulated people in urban areas may feel at ease with the pixel since the typical visual noise in the metropolis is now diminished with the pixel. Who knows this may be the next step into the telecommunication where people communicating vast distances will be able to view through each other's eyes, thus enhancing their experience. As we are now able to put the PIXEL in the hands of people, many new audiences and uses are sure to arise.

Design of the PIXEL

VERSION 1: Forms that Afford Picking Up and Looking Into

How can the PIXEL communicate that you can pick it up, look into it, and see an image? For the PIXEL to work this people needed to instantly realize this is how you interact with it. To help answer this question we looked to objects that have this same functionality and to leverage the pre-existing bias towards the form. The first that came to mind was the classic view Master. It's safe to say that almost everyone has used this iconic toy. Looking into a viewMaster, you expect to see an image. Click the button and see something new. This form was ideal for the functionality of the PIXEL, so much so that we hacked apart a viewMaster for to build PIXEL v1. By using the same design structure of the viewMaster, we are allowing the pre-existing bias to help society adjust to a new technology with out scaring them away from the new technology.

We also took apart a pair of binoculars. Similar to the viewMaster, you place them to your eyes. In contrast to the viewMaster, you look through binoculars rather then into. You expect to see the landscape in front of you, only magnified. While we didn't use the binoculars to prototype, we did find a nice collection of plastic lenses inside. The observation deck viewer, similar to binoculars, shows an enlarged version of the landscape in front of you. The from of these viewers, with their exaggerated eye holes and over sized enclosure, provided inspiration for the PIXEL's form.

VERSION 2: It's about the concept
The PIXEL v1 was very bulky but it worked. The technology required a large amount of space and lots of electrical connections. We debated using a network of Arduinos, making the hardware more compact, but high-tech was not the aim of our project. Instead, we decided to exaggerate the bulkiness. So, we found a suitcase and made the unit portable by adding a long extension cord. The suitcase exaggerates the bulkiness of the device just as the pixel exaggerates your vision.

The graphics on the PIXEL units and suitcase help indicate further the devices function. We tried to simplify them as much as possible, using minimal words. We wanted the graphics to communicate the function to everyone, regardless of language. These scaffolding elements are critical to helping new users understand the devices function, and are rarely noticed after that.

Developing the PIXEL Software

As we like to tell people, the PIXEL either works or appears to work. This bit of graceful degradation was added in the 11th hour (literally, the day of the show at 4am) by simplifying the software used to run PIXEL. Let's backtrack a bit to better understand. The original PIXEL software was a modified version of the EyeWriter. Zach Lieberman and his crew over at Parsons sent us the code to enslave the mouse immediately after the software was calibrated. We included this code and tweaked things a bit so the move mode could be toggled with a key press. We also toggled the EyeWriter window from full-screen to 1x1 pixels. This quick hack hid the EyeWriter interface while still leaving the window selected so mouse control would work.

Then we developed a flash application for the display part of the PIXEL. This app took the input from the outwards facing camera (camera 2) and displayed it on screen twice, side by side (this dual vision was necessary for v1, since the viewMaster lenses required two images that were brought together. In v2 only one large image was required.) Mouse movement was mapped from full screen coordinates into the small video window. Sound complicated? It was. V1 of the pixel required you to start two applications and adjust their settings.

Running two applications was not the main problem, however, it was the need to calibrate and move the mouse to move the display mask. While this system worked in testing, it became clear when we set up that no two people would have the same calibration -- contrary to our belief we could not even find a rough calibration that worked consistently for one person using, removing, and then placing the PIXEL back to their head. We needed to make things simpler.

We noticed that what did work remarkably well and consistent was the EyeWriters ability to track the pupil. This blob tracking only required an IR illuminated pupil to lock on. Never mind calibrating, why not use this pupil location to move the mask? This was a turning point... why hadn't we thought of it before? Matt began to tweak the software. The first step was bringing both camera feeds into the EyeWriter software. Then, the pupil position was mapped to a mask over the camera feed. The finished software was very accurate and simple. Even better, if it couldn't blob track your pupil, it simply kept the mask in the middle of the screen... and looked like it worked. The same persons will not be able to look through both (although this was tried by one person) so no one will ever notice... shhh, don't tell. ;)

Future versions of this software could enter into this mode on startup, rather then requiring the user to set the mode. This way the Mac Mini could launch that app on startup, and thing would work literally at the push of a button. Also, the software still does not take full advantage of the screen resolution. Future versions should detect screen resolution and adjust accordingly.

Prior Art

Seeing through other peoples' eyes is a very intangible concept for most. Consider the movie Being John Malkovich, where people can enter a room and into the head of the main character John Malkovich. While you can never really put on another persons eyes, the PIXEL does bring this intangible concept closer to reality. While we can question if another persons visual experience makes sense when taken out of context, the PIXEL units close proximity to each other helps ensure that both users are aware of what is happening. Since they are sharing a similar experience, the visuals are less out of context (then if two people were in different locations, unaware someone else was participating.)

Along the journey of development, after explaining our plan to pixelate visual noise Mike McAllister (one of our professors), Mike suggested we look into a project done by one of his prior students at Syracuse University. Steve Belovarich had created a short video series he called the Tourist. His video demonstrated the concept of head mounted camera and display system. This system, controlled by an artificially intelligent computer, would pixelate advertisements on request of the wearer. While the concept (that we had been toying with) was communicated brilliantly, the actual device did not work. All pixilation was rendered as post-processing. With our system the visual manipulation is done live, affording a rich interaction between two people.

When deciding on a screen to use for the PIXEL we found a number of head mounted displays, mainly for sale through eBay from China. These devices advertised creating the effect of a 50" projection TV just inches from your eyes. They consisted of two small Lads on a very futuristic headband. While these devices promise a huge screen experience, we wanted something more personal. Our decision to use a 7" widescreen LCD inside of a plastic enclosure affords looking into. Rather then just looking at a big screen users can look into the WATCH PIXEL, moving their heads and squinting slightly to see the corners of the screen. It's like looking into their minds.

Our decision to use open-source software and inexpensive webcams means the PIXEL technology was not only affordable to develop, but it’s affordable for the most people to create their own. Compared to Tobii systems that cost upwards of

$40,000 and the augmented reality systems used by Boeing and large corporations the PIXEL technology is even more affordable. The overall cost is around $576 ($100 flat screen monitor, $100 for 2 webcams, monitor converter box ($75), Plastic Mold Frame ($300, this is not necessary though), Open frameworks program ($0), suitcase ($1, used from a thrift store), plywood ($15). This does not include the laptop (or Mac mini).

Further Development Possibilities

Having a huge number of people using this at the opening allowed for good feedback. Also listed is feedback from our own observations.

Eyes are too close

Many people noted the eyes were too close when using the PIXEL. We also noticed this when fitting the camera: it didn't read well in one direction because the edge of the plastic eyehole cut off camera. In hindsight, the eyeholes are not the exact dimensions of the ViewMaster on which we based the form. This should be addressed in future iterations of the PIXEL.

LOOKer doesn't get rewarded.

Many people noted the WATCH unit was the cooler of the two; you get to see something new. How can we bring this cool factor to both units? One solution is to supplement regular vision by adding binoculars. This experience would make the LOOK pixel a useful artifact by giving the looker a close up view of what they were looking at.

Needs 2 people to function

Without someone looking through the LOOK PIXEL, the experience for the WATCHer is mediocre at best. Often the LOOK unit was positioned in the suitcase, camera down, and WATCHers would see complete darkness. This was fixed during the show by placing one pixel on the side on the ground, so it showed the wall. This at least let the WATCHer understand what was happening. However this must be addressed in future iterations: the PIXEL needs to degrade gracefully by showing something to the WATCHer even with only one person present. Or, there needs to be a way to better ensure two people will always be using the device. Perhaps it would light up (or turn on) once both units were held. Rearranging the wires so the two units were directly connected would also help indicate that two people should use the device at once.

REFERENCES

Dieter Koller[1],

1. Tombs, Selina., Silverman Irwin, “Pupillometry: A sexual selection approach,” Evolution and Human Behavior 25:4 (2004): 221-228.

2. Jacob, R.J.K., K.S. Karn, “Eye Tracking in Human-Computer Interaction and Usability Research: Ready to Deliver the Promises (Section Commentary),” Mind’s Eye: Cognitive and Applied Aspects of Eye Movement Research,” Amsterdam: Elsevier Science (2003).

3. Koller, Dieter., Joseph Weber, Jitendra Malik, “Robust multiple car tracking with occlusion reasoning,” “Lecture Notes in Computer Science,” Berlin: Springer (1994), (89-196).

4. Koller, Dieter., K. Daniilidis, T. Thorhallason, H. Nagel, “Model-based object tracking in traffic scenes,” “Lecture Notes in Computer Science,” Berlin: Springer (1992), (437-452).

5. Hartmann, Björn., Scott Klemmer, Leila Takayama, “How Bodies Matter: Five Themes for Interaction Design*”.* DIS June 26-28 2006. University Park, Pennsylvania, USA.

6. Chowdhry, Amit., “Apple Spending Cash on Eye Tracking Technology From Tobii,” Pulse 2, May 16, 2009, Web.

7. Vertegaal, et al., US Patent Application: US 2006/0110008 AI.

8. Drewes, Heiko., Alexander De Luca, Albrecht Schmidt, “Eye-gaze interaction for mobile phones,” “International Conference on Mobile Technology, Applications, and Systems,” (364-371), (2007).

9. Zhai, Shumin., Carlos Morimoto, Steven Ihde, “Manual and Gaze Input Cascaded (Magic Pointing),” In Proc. CHI’99: ACM Conference on Human Factors in Computing Systems, May 15-20 1999.

10. Salvucci, Dr. Dario D.. Interview performed by Matt Miller, Georgia Gunthrie, Kristen Freeze, and John MacDuffie Woodburn. Feb 2, 2010

11. Rayner, “The perceptual span and peripheral cues in reading,” Cognitive Psychology 7(1975):65-81.

12. Mason C., Kandel E. R. Central visual pathways. Principles of Neural Science, E. R. Kandel, J. H. Schwartz, T. M. Jessel. Appleton & Lange, Norwalk 1991; 420–439

13. Copeland, Larry., “More cities ban digital billboards,” *USA Today* March 4, 2010.

14. New York Times. Billboard ban in Sao Paulo angers advertisers. http://www.nytimes.com/2006/12/12/world/americas/12iht-brazil.html, Dec. 12, 2006.

15. ScienceDaily 9 March 2007. 9 May 2010. http://www.sciencedaily.com /releases/2007/03/070308121938.htm.

APPENDIX A: Critic's Responses

1. Make it more collaborative
    a. Idea of tethering the two physically together
    b. Idea of projecting the watcher onto a screen so then the user can actually watch
2. Have the looking out actually be able to view more
    a. What is really the intention
    b. Tony mentioned that its fun, bringing fun to the weird virtual world that people do not get
3. What are peoples reactions when using Pixel
4. Jonas, thought that this could be used to raise social issues or provoke a conversation
5. Good use of trying to find a use with eye tracking.
6. We are the two so close together
    a. The reason that the USB can only go certain distance is not a good answer
    b. How far do we really want them
    c. However, the USB makes them a connection. People want to switch back and forth
    d. Adding a string that connects the two pixels would help communicate that there is a bond.
    e. Adding audio would further push
7. Tony said that I brought joy to the ID world, that this project is fun and full of joy. Don't lose the joy.