2010

# EyesOn Mobile Eye Tracking

Thomas Jacques

Carnegie Mellon University

School Of Computer Science Senior Thesis

# Dedication

I would like to dedicate this thesis to Florence Doo, for inspiring the idea and supporting me throughout the effort of its implementation. Without her support I would not have been able to complete this work.

# Table of Contents

## List of Figures



## List of Tables

## Abstract

Eye tracking is an extremely valuable resource for behavioral research and the next generation of human computer interaction, especially for handicapped individuals. However, obtaining robust, high quality eye-tracking data can be enormously expensive, and low cost alternatives can be inaccurate and unable to be used as computer input devices. High end mobile eye tracking systems, such as those manufactured by *Tobii®*, can cost as much as forty thousand dollars. A very low cost, easy to assemble mobile eye tracking unit was constructed, using a USB interface. Robust software to analyze the video data streams in real time was creating using C#. A method for screen detection on the mobile eye-tracker through the use of a hot swappable scene camera filter and Infrared LEDs was also developed. This allows for real time eye-tracking as a method of computer input on a head mounted mobile eye-tracker. The device was thoroughly tested in experimental settings and found to be comparably precise to commercial systems. This work enables quality eye-tracking research to be performed at extremely low cost.

# Chapter 1 Overview

## Introduction

Eye tracking is capable of measuring where a person is gazing. Eye tracking can be used both as a scientific research tool and has many applications for Human Computer Interaction (HCI). There are many techniques which have been used to accomplish the task of eye tracking, but they all fall under two main categories: analog techniques, and video based techniques. Analog techniques for eye tracking, which use physical components placed on or near the eyeball, have been available since the 1950s (Yarbus 1967). However, they are uncomfortable and unwieldy for the users, and can often be worn only for short periods of time (Young and Sheena 1975). The other major method of eye tracking is through video based analysis. Video images of the eye are analyzed by algorithms to discern where the eye is looking. There are a variety of video based techniques and products which utilize them. More recently, off-the-shelf video-based eye trackers have become available as alternatives to these high priced specialized eye trackers because the price of high-quality digital cameras has dropped and because of the advent of image processing and computer-vision techniques for eye tracking.

## Eye Tracking Techniques and Methods

### Analog Techniques

Eye tracking technology has been available for many years using a variety of methods such as contact-lens based eye-coil systems, Purkinje-reflection based, and electro-oculography. (Young and Sheena 1975) The electro-oculography method measures the electrical changes in electrodes placed on the skin near the eyes, and uses this to track eye movements (Kaufman, Bandopadhay and Shaviv 1993). Contact-lens based eye coil systems places contact lens with affixed mechanical lever or magnetic coil into the eyes. And then eye movement can be tracked by tracking these affixed objects (Jacob 1995). A comparison of these eye tracking techniques can be found in (Glenstrup and Engell-Nielse 1995).

### Video Based Eye Tracking Techniques

Two types of imaging approaches are commonly used in video based eye tracking: visible and infrared spectrum imaging (Hansen and Pece 2005) (see Figure 1.1 and 1.2 for the eye images captured with visible and infrared spectrum imaging). The two approaches are very different because of the way they display the three most relevant features of the eye: the pupil - the aperture that lets light into the eye, the iris - the colored muscle group that controls the diameter of the pupil, and the sclera - the white protective tissue that covers the remainder of the eye. Visible spectrum imaging is a passive approach

that captures ambient light reflected from the eye. In these images, it is often the case that the best feature to track is the contour between the iris and the sclera known as the limbus.

**Figure 1.1**

**Figure 1.2**

Figure 1.1-2: (1) Image of the eye under visible light, with ambient reflections visible. (2) Image of the eye under infrared light illumination. Notice the clear contrast of the pupil against the iris.

Visible spectrum eye tracking is complicated by the fact that uncontrolled ambient light is used as the source of illumination, which can contain multiple specular and diffuse components. Infrared imaging eliminates uncontrolled specular reflection by actively illuminating the eye with a uniform and controlled infrared light not perceivable by the user. A further benefit of infrared imaging is that the pupil, rather than the limbus, is the strongest feature contour in the image (see Figure 1.2). Both the sclera and the iris strongly reflect infrared light while only the sclera strongly reflects visible light.

Tracking the pupil contour is preferable given that the pupil contour is smaller and more sharply defined than the limbus. Furthermore, due to its size, the pupil is less likely to be occluded by the eyelids and eyelashes. The primary disadvantage of infrared imaging techniques is that they cannot be used in all outdoor scenarios during daytime because the ambient infrared illumination may interfere with the imaging.

Infrared eye tracking typically utilizes either a bright-pupil, dark-pupil technique or both. The bright-pupil technique illuminates the eye with a source that is on or very near the axis of the camera. The result of such illumination is that the pupil is clearly demarcated as a bright region due to the photo-reflective nature of the back of the eye. Dark-pupil techniques illuminate the eye with an off-axis source such that the pupil is the darkest region in the image while the sclera, iris and eye lids all reflect relatively more illumination. In either method, the first-surface specular reflection of the illumination source off of the cornea (the outer-most optical element of the eye) is also visible. The vector between

the pupil center and the corneal reflection center is typically used as the dependent measure rather than the pupil center alone. This is because the vector difference is less sensitive to slippage of the head gear - both the camera and the source move simultaneously.

The combined use of both bright and dark pupil techniques can be seen in (Zhai, Morimoto and Ihde 1999), (Morimoto, et al. 2000), and (Zhu and Qiang 2005). Those imaging systems usually include the LEDs on and off the axis of the camera. The interlaced bright-pupil image and dark-pupil image are obtained by switching between either the LED on or off of the axis of the camera. The difference image obtained by subtracting these images is thresholded, which results in an image of the pupil.

### Remote and head-mounted systems

Eye-tracking systems can be divided into remote and head-mounted systems. Each type of system has its respective advantages. Both visible-spectrum and infrared-spectrum imaging techniques have been applied in the context of remote video-based eye tracking. The single most attractive reason for using a remote eye-tracking system is that its use can be completely unobtrusive. However, a limitation of a remote system is that it can only track eye movements when the user is within a relatively confined area of operation. Furthermore, the accuracy of remote eye-tracking systems is usually worse than the head-mounted eye-tracking systems due to the nature of the set up. In a remote system, the camera is placed further away from the eye, and does not stay in the same position relative to the head, meaning the view of the eye can change. Stereo cameras can be applied to achieve better eye-tracking accuracy (Shih and Liu 2004) (Newman, et al. 2000). The design of remote eye-tracking systems must consider the three way trade-off between cost, flexibility and quality. For example, the flexibility to track eye movements over a wide area can be improved by using a pan-tilt camera, but such cameras are quite expensive. Furthermore, the quality of eye tracking can be improved by capturing a high-resolution image of the eye using a zoom camera (Beymer and Flickner 2003), with the trade-off of a reduced operational area and higher cost.

Although, there are a number of promising remote eye tracking approaches such as those described by Morimoto (Morimoto, Amir and Flickner 2002) and Tian (Tian, Kanade and Cohn 2000), it currently appears that a head-mounted system has a greater potential to achieve a reasonable compromise between all of these factors. The major drawback of the head mounted system is the inability to use the device as a real-time method of computer input. In remote systems, there is a static screen reference known to the eye-tracker, and the eye tracker is able to detect the eye and head to

maintain a static model of the eye. Mobile eye trackers have a static reference to the eye; however, they do not currently have the ability to discern where exactly on a computer screen the user is gazing.

The innovative work of Jeff Pelz and colleagues (Pelz, et al. 2000) (Babcock and Pelz 2004) at the Rochester Institute of Technology (RIT) on the construction of low-cost minimally invasive head-mounted eye trackers is particularly noteworthy. In their system, analog cameras are mounted onto safety glasses (in a similar configuration as that shown in Figure 2.1) and video of the user's eye and the user's field of view are interleaved in a single interlaced video frame and recorded using a mini-DV camcorder stowed in a backpack. Point of gaze computation is then performed off-line using proprietary hardware and software purchased from a production house. Given the goal to integrate eye movement measurements into human computer interfaces and use the system as a method of computer input, this dependence on high-cost proprietary equipment is a serious limitation of their approach. Furthermore, the off-line nature of the system is another limitation as some degree of real-time performance will be necessary in many HCI applications. However, their innovation in head-gear design and low-cost approach is laudable and both are adopted for the task.

# Chapter 2 The EyesOn System

## EyesOn Eye Tracking Hardware

In this section, the design of the EyesOn eye-tracking hardware is described in a way that shows the evolution of the system to its final form. This approach provides insight into principles, decisions, benefits, and limitations of the system.

### Generation 1

Two Logitech Quickcam Pro 9000® cameras were utilized as the video devices. The USB cameras were stripped of all plastic and nonessential components so ensure they were as light as possible. An infrared LED was affixed using solder to the USB camera board along with a resistor which reduces the current through the LED to ensure the light intensity projected by the LED is not harmful. USB is specified to provide 5 volts, and the voltage drop across the LED is 1.2 volts. The resistor used was rated at 430 Ohms, so the resulting current through the LED is approximately 8.8 mA. This results in an input power of 10.56 mW. An irradiance level less than 10 $mW/cm^2$ is considered safe for chronic IR exposure in the 720-1400 nm range (Sliney and Wolbarst 1980) (ICNIRP 1997) (ICNIRP 2000). Since the LED diffuses over an area of around 3 $cm^2$, and the LED is nowhere near perfectly efficient, it is reasonably certain that the irradiance level is within the 10 $mW/cm^2$ threshold.

The infrared blocking filter in the USB camera was removed from the eye camera, and the filter was replaced with an infrared passing filter. This was accomplished by removing the lens from the camera board, and breaking the infrared blocking filter on the lens frame. A Kodak® 87 Wratten filter was used as the infrared passing filter on the new cameras, attached to the outside of the lens frame using hot glue.

The first design consideration after choosing to use a head-mounted system was the configuration of the head gear. The most significant issue was where to mount the cameras. The lens of the scene camera was placed as close to the physical eye as possible without occluding vision in order to minimize the induced parallax error. Because the camera was attached to the safety glasses frame using heat shrink, the smallest distance achievable from the affixed camera to the eye was approximately 1.5 inches (3.8 cm).

Some commercial units utilize hot mirrors, bits of glass or plastic which reflect infrared light but are completely passing to other wavelengths, in conjunction with their eye-trackers to reduce the error induced by camera sway. However, hot mirrors can be expensive and difficult to obtain in small quantities. In order to reduce costs, the USB camera was affixed using stiff steel wire to point with a

direct line of sight towards the eye. Since the camera is extremely light, and the wire is comparatively strong and rigid, there is not a significant amount of camera sway due to head movement, and what little sway there is quickly tapers off. The primary disadvantage of a boom arm design is that a portion of the visual field is blocked by the camera and the armature. Since there is only a small extent of visual occlusion and peripheral positioning of the camera, this is an acceptable compromise. In fact, because these components are attached to the head gear and thus static in the user's visual field, they are easily ignored just as the frames of normal eye glasses are ignored.

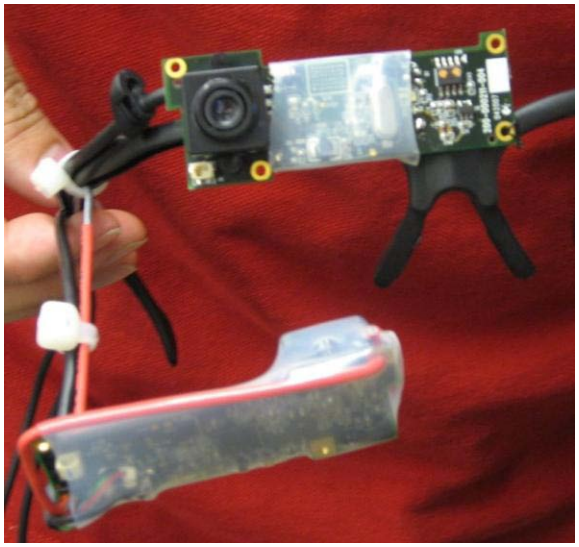

**Figure 2.1**

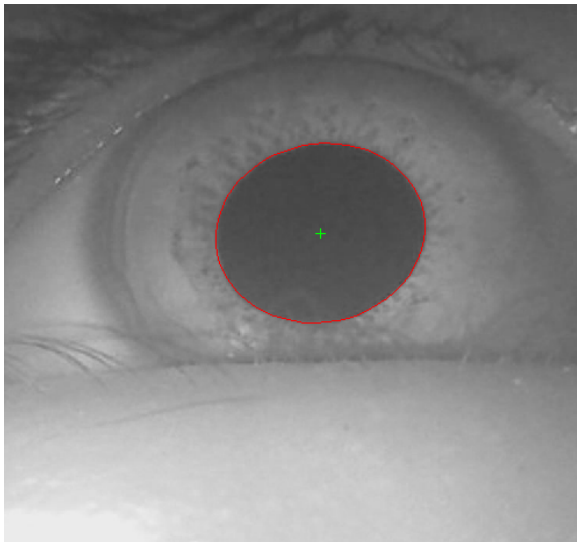

**Figure 2.2**

Figure 2.1-2: Eye tracker and the captured images. (1) Head-mounted eye tracker. (2) Image of the user's right eye illuminated with infrared light (taken from the first generation model). Note the clearly contrasted dark pupil and the best fitting ellipse corresponding to the algorithm's analysis of the image, with the pupil center marked in green.

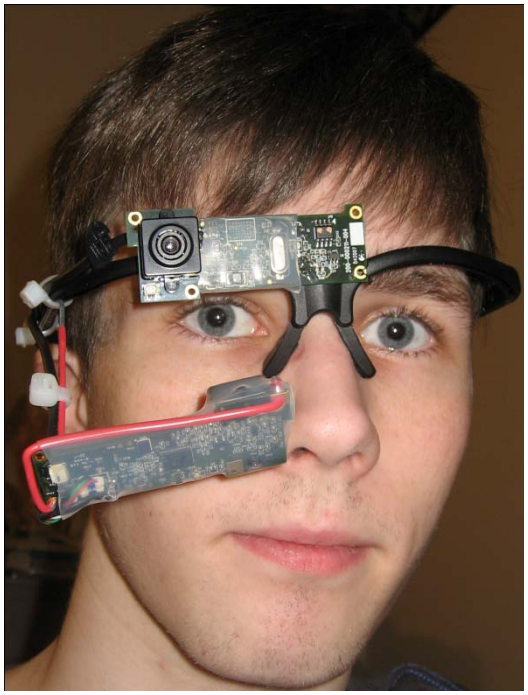

**Figure 2.3**

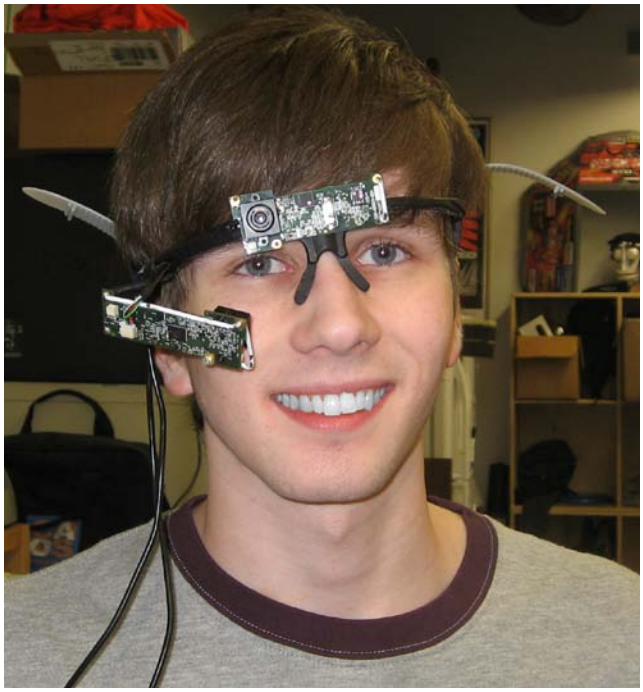

**Figure 2.4**

Figure 2.3-4: Pictures of the two generations of mobile eye trackers being worn by myself. (3) shows the first generation model, and (4) shows the second generation model.

## Generation 2

The second generation prototype was very similar. The most debilitating factor of the first generation model was the occurrence of slippage. If the device slipped or moved at all, enormous error resulted. The first generation model used an adjustable cloth strap to prevent slippage from occurring, but even when tightly secured slippage was still a significant issue when the device was used for more than several minutes. In order to address this issue a rubber goggle strap was used to snugly hold the device to the subjects head. The strap is adjustable and allows the device to be very firmly held in place. When adjusted properly, the problem of slippage was entirely eliminated in experimental settings.

The hot-mirror was removed from the scene camera in addition to the eye camera for the second generation model, and a hot-swappable filter was placed on the scene camera, to allow the camera to quickly switch between the visible and infrared light spectrums. The user simply places the correct filter over the lens and affixes it in place by snapping it into the steel wire frame. This is the key

feature allowing the device to be used as a method of computer input, as the corners of the monitor can be marked and identified using infrared LEDs.

The attachment of the scene camera to the glasses differed. Instead of using heat shrink to affix the scene camera, a steel wire frame was built onto the safety glasses. The holes of the camera board align with the steel wire frame to rigidly attach the scene camera to the safety glasses. This improved design increased the precision at which the scene camera could be placed. As a result the distance from the scene camera to the eye was reduced to approximately 1 inch (2.54 cm). The length of the boom arm attaching the eye camera to the frame was drastically reduced, because the cameras were upgraded to two Logitech Webcam Pro 9000s®. The Webcam Pro 9000® model has a significant enhancement over the Quickcam Pro 9000® in that it is able to focus at a much closer distance, allowing the eye camera to be brought closer to the eye. The first generation model suffered from a large loss of resolution on the image of the eye because there was no optical zoom. By placing the camera closer to the eye, this higher resolution is preserved. This higher resolution image allows for a more accurate extraction of the center of the pupil, hence increasing the precision of the device. An additional benefit to having a shorter boom arm is the reduced camera sway due to movement. The shorter steel wire is less susceptible to sway since it holds the eye camera more rigorously in place.

All of the experiments and tests were performed using the second generation model and all pictures from the eye and scene cameras were taken using the second generation model, unless otherwise specified.

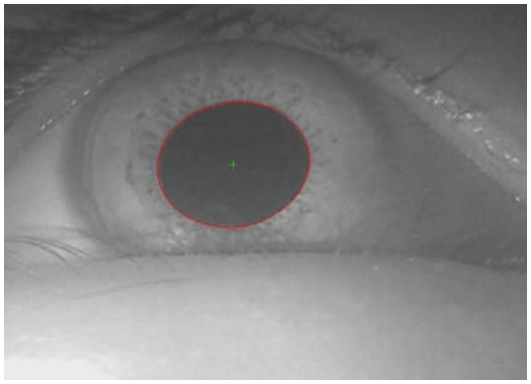

**Figure 2.5**

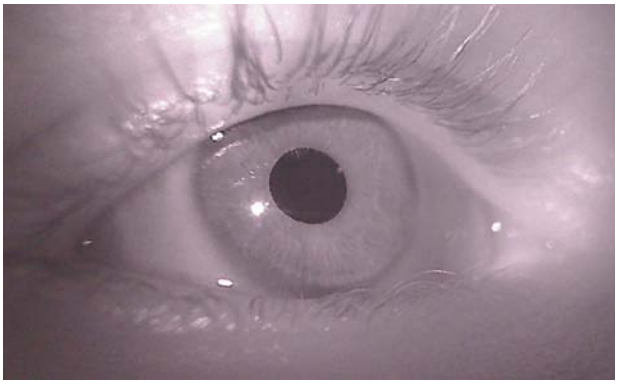

**Figure 2.6**

Figure 2.5-6: (5) Shows a picture of the eye from the first generation prototype, (6) shows a picture of the eye from the second generation prototype. Notice the large difference in resolution due to the second generation's closer proximity to the eye.

# Chapter 3 EyesOn Gaze Interpretation and Screen Detection

## Gaze Interpretation

Presented in this section is a robust eye-tracking algorithm, modified from the Starburst algorithm, which combines feature-based and model-based approaches to achieve a good trade-off between run-time performance and accuracy for dark-pupil infrared imagery. The goal of the algorithm is to extract the location of the pupil center and the corneal reflection so as to relate the vector difference between these measures to coordinates in the scene image. The algorithm does the following:

- Locate and remove the corneal reflection from the image using thresholding.
- Detect pupil edge points using an iterative feature-based technique.
- Identify and remove bad pupil edges using the Random Sample Consensus (RANSAC) algorithm.
- Calculate the best fitting ellipse to the remaining feature points to identify the pupil center.

### Corneal Reflection Detection

In infrared spectrum eye tracking using the dark-pupil technique, the corneal reflection corresponds to one of the brightest regions in the eye image. Thus the corneal reflection can be obtained through thresholding. However, a constant threshold across observers and even within observers is not optimal. Therefore an adaptive thresholding technique is used in each frame to localize the corneal reflection. Note that because the cornea extends approximately to the limbus, the search for the corneal reflection can be limited to a square region of interest. To begin, the maximum threshold is used to produce a binary image in which only values above this threshold are taken as corneal reflection candidates. It is likely that the largest candidate region is attributable to the corneal reflection, as any other specular reflections tend to be quite small and located off the cornea.

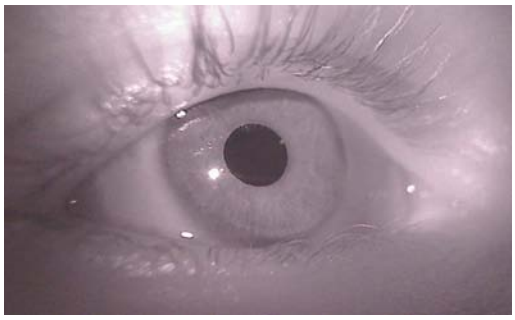

**Figure 3.1**

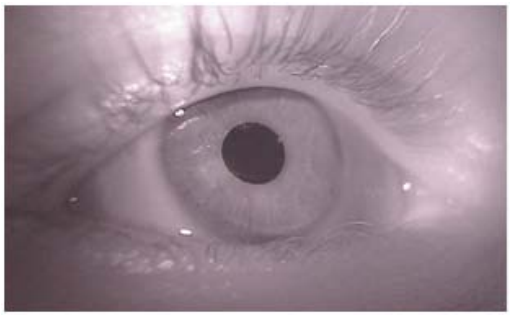

**Figure 3.2**

Figure 3.1-2: (1) Display the original image of the eye as captured by the eye camera. (2) Shows the eye image after the corneal reflection detection and removal procedure.

The ratio between the area of the largest candidate region and the average area of other regions is calculated as the threshold is lowered. At first, the ratio will increase because the corneal reflection will grow in size faster than other areas. Note that the intensity of the corneal reflection monotonically decreases towards its edges, explaining this growth. A lower threshold will, in general, also induce an increase in false candidates. The ratio will begin to drop as the false candidates become more prominent and the size of the corneal reflection region becomes large. We take the threshold that generates the highest ratio as optimal. The location of the corneal reflection is then given by the geometric center $(x_c, y_c)$ of the largest region in the image using the adaptively determined threshold.

Radial interpolation is then used to remove the corneal reflection. First, the central pixel of the identified corneal reflection region is set to the average of the intensities along the contour of the region. Then for each pixel between the center and the contour, the pixel intensity is determined via linear interpolation. An example of this process can be seen in Figure 3.1 and Figure 3.2

### Pupil Contour Detection

Pupil feature points are detected according to the following method:

- Initial starting point obtained through thresholding
- Pupil candidate feature points are iteratively detected by following rays extending outward from the starting point.
- At each step along the ray, the derivative of intensity is calculated at each point.
- If the derivative intensity exceeds the threshold, place a feature point and stop traversing that ray, if the threshold is not exceeded, do nothing.
- For each feature point placed: follow rays within 30 degrees of the line from the feature point to the starting point, and mark new feature points. The new starting point is the geometric center of all the points.
- Repeat until the starting point converges.

To better demonstrate the algorithm I will first show the iterations when the starting point it the pupil center of the previous frame, rather than the center obtained through thresholding. It can be observed that even when the starting point is a fairly poor estimate of the actual center, the algorithm quickly converges to the center of the eye.

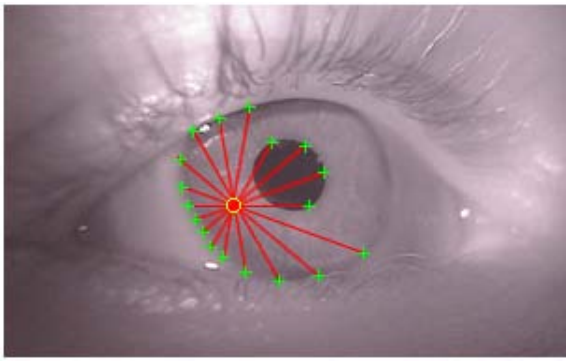

**Figure 3.3**

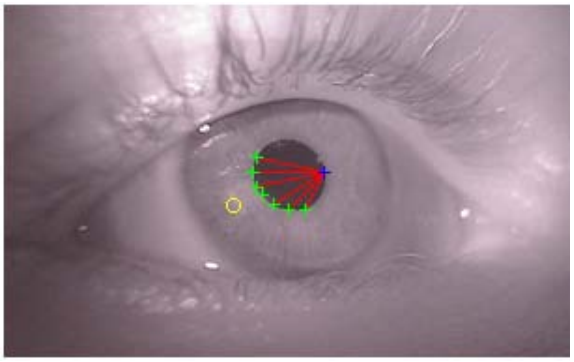

**Figure 3.4**

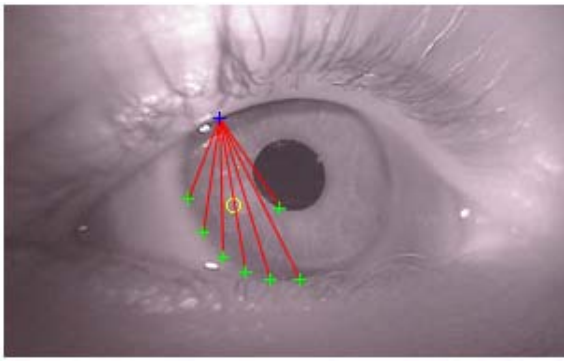

**Figure 3.5**

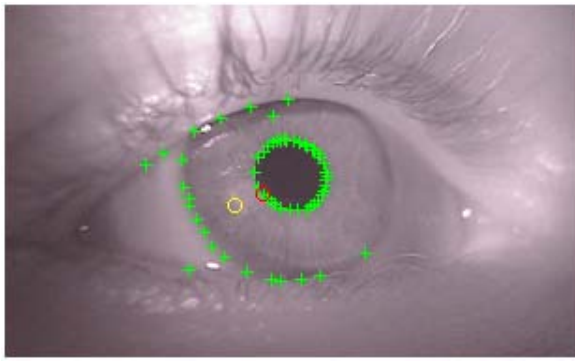

**Figure 3.6**

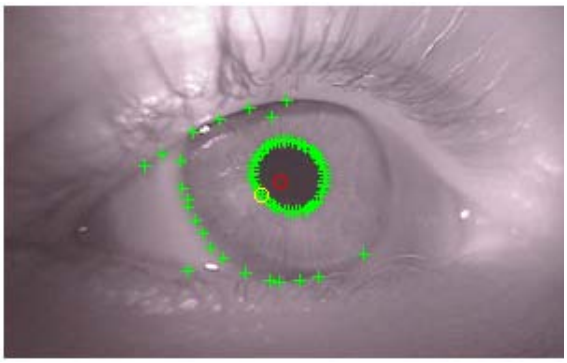

**Figure 3.7**

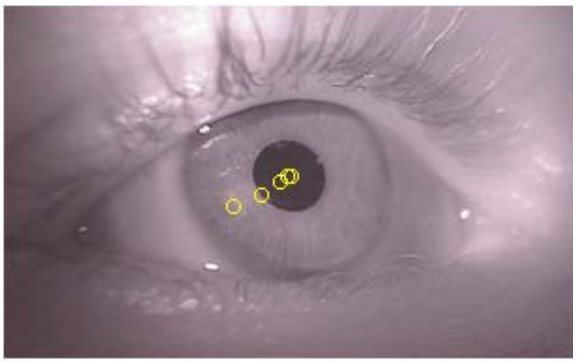

**Figure 3.8**

Figure 3.3-8: (3) shows the original starting point and the burst of rays in all directions, with green crosses indicating where a feature point was detected. The starting point of an iteration is shown as a yellow circle. (4-5) show the ray spread starting from a discovered feature point, shown as a blue cross, with new feature points detected shown in green. (6) shows the end of an iteration, with all feature points detected shown as green crosses, the yellow circle was the starting point of the iteration, and the red circle is the geometric center of the discovered points, used as the starting point in the next iteration. (7) shows the same information at the end of the second iteration. (8) shows the starting point progression until convergence.

Here is the same procedure repeated, instead using the implemented method of starting from the thresholded center.

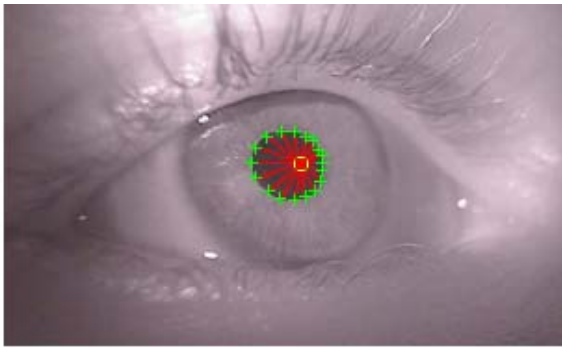

**Figure 3.9**

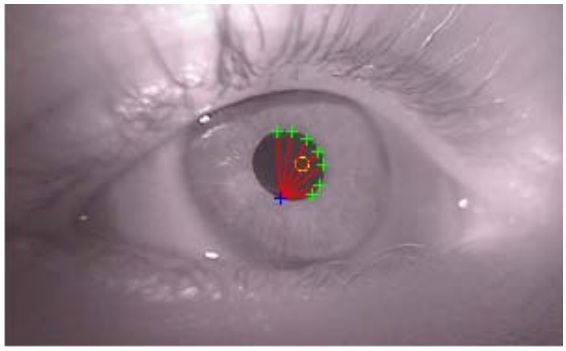

**Figure 3.10**

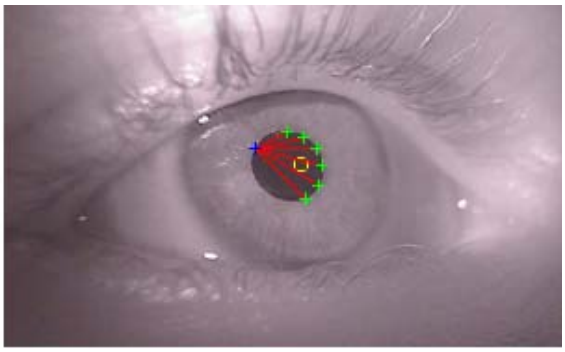

**Figure 3.11**

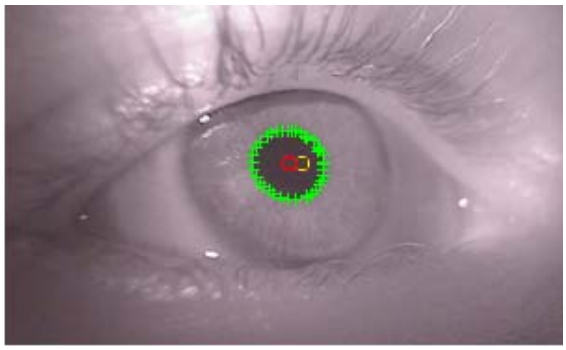

**Figure 3.12**

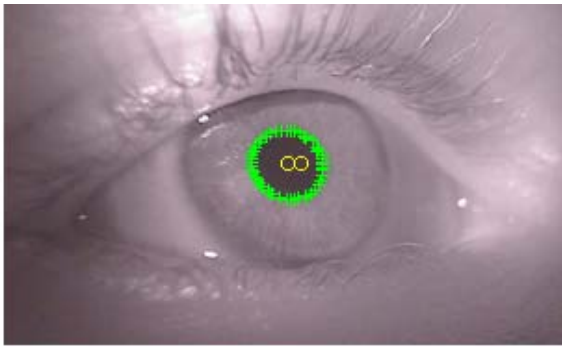

**Figure 3.13**

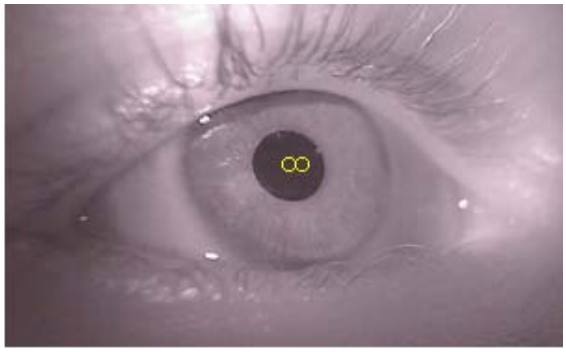

**Figure 3.14**

Figure 3.9-14: (9) shows the original starting point and the burst of rays in all directions, with green crosses indicating where a feature point was detected. The starting point of an iteration is shown as a yellow circle. (10-11) show the ray spread starting from a discovered feature point, shown as a blue cross, with new feature points detected shown in green. (12) shows the end of an iteration, with all feature points detected shown as green crosses, the yellow circle was the starting point of the iteration, and the red circle is the geometric center of the discovered points, used as the starting point in the next iteration. (13) shows the same information at the end of the second iteration. (14) shows the starting point progression until convergence.

## Ellipse Fitting

Given a set of candidate feature points, the next step of the algorithm is to find the best fitting ellipse. While other algorithms commonly use least-squares fitting of an ellipse to all the feature points, gross errors made in the feature-detection stage can strongly influence the accuracy of the results. Consider the detected feature points shown in the figure below, and the resulting best-fit ellipse using the least-squares techniques shown in the figure next to it.

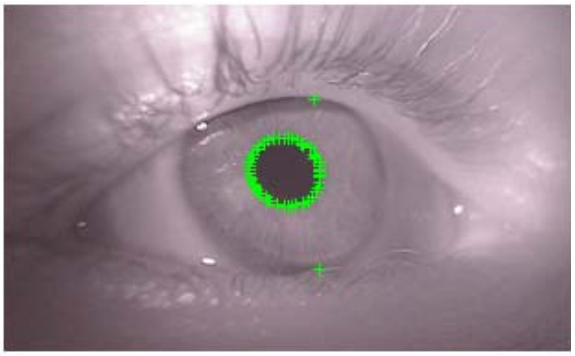

**Figure 3.15**

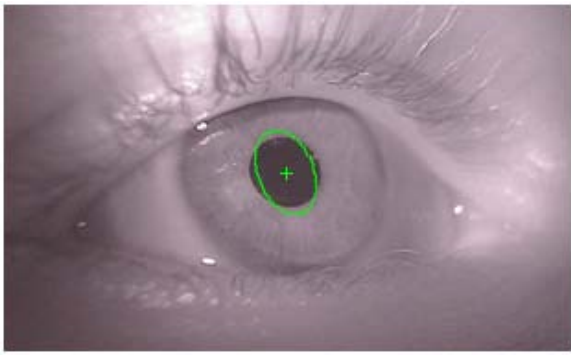

**Figure 3.16**

Figure 3.15-16: (15) Set of feature points detected by the algorithm.
(16) The geometric best fitting ellipse to these detected points.

Notice that a few feature points not on the pupil contour dramatically reduces the quality of the fit to an unacceptable level. To address this issue, the Random Sample Consensus (RANSAC) (Bolles 1981) paradigm for model fitting is applied. RANSAC is an effective technique for model fitting in the presence of a large but unknown percentage of outliers in a measurement sample. An inlier is a sample in the data attributable to the mechanism being modeled whereas an outlier is a sample generated through error and is attributable to another mechanism not under consideration. In our application, inliers are all of those detected feature points that correspond to the pupil contour and outliers are feature points that correspond to other contours, such as that between the eye lid and the eye. Least-squares methods use all available data to fit a model because it is assumed that all of the samples are inliers and that any error is attributable exclusively to measurement error. On the other hand, RANSAC admits the possibility of outliers and only uses a subset of the data to fit the model. In detail, RANSAC is an iterative procedure that selects many small but random subsets of the data, uses each subset to fit a model, and finds the model that has the most agreement with the data set as a whole.

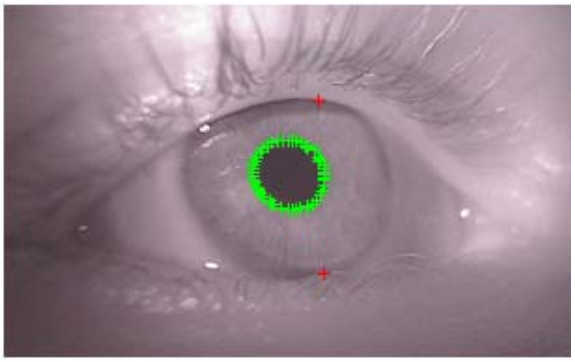

**Figure 3.17**

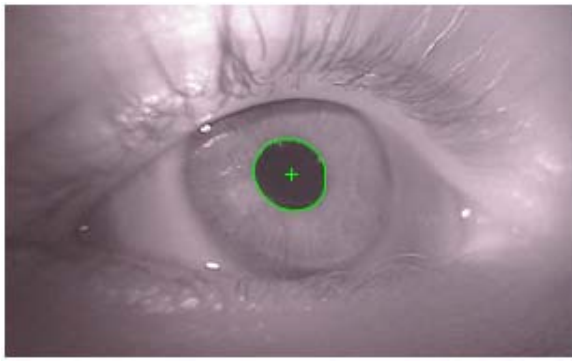

**Figure 3.18**

Figure 3.17-18: (17) Set of feature points detected by the algorithm, with inliers in green and outliers in red.
(18) The geometric best fitting ellipse to the inlier points.

## Calibration

In order to calculate the point of gaze in the scene image, a mapping must be constructed between eye-position coordinates and scene-image coordinates. The mapping can be initialized by relating known eye positions to known scene locations. The typical procedure in eye-tracking methodology is to measure this relationship through a calibration procedure (Stampe 1993). During calibration, the user is required to look at a 3 x 3 grid of scene points for which the positions in the scene image are known. While the user is fixating each at scene point $s_i = (x_{si}, y_{si})$, the eye position $e_i = (x_{ei}, y_{ei})$ is measured. For each correspondence between $s_i$ and $e_i$, two equations are generated that constrain the mapping:

$$x_{si} = a_{x0} + a_{x1}x_{ei} + a_{x2}y_{ei} \quad \text{(Equation 1)}$$
$$y_{si} = a_{y0} + a_{y1}x_{ei} + a_{y2}y_{ei} \quad \text{(Equation 2)}$$

The position of gaze for any new pupil center point is a quadratic transformation on the nine calibrated indices using

## Screen Detection

Screen detection is accomplished via the use of infrared LEDs affixed to the corners of the screen monitor. Using the infrared passing hot-swappable plastic filter on the eye tracking device, the corners of the computer monitor can be detected according to the same method we detected the corneal reflection in the eye. That is, by adaptive thresholding. This allows us to very quickly identify the monitor corners, which we then use internally to calculate whether or not the subjects gaze lies within the bounds of the computer screen, and if it does, we can project this gaze bearing directly onto the monitor with a standard linear transformation.

Unlike the corneal reflection, which stays in relatively the same area in each frame, the monitor corners almost never stay in the same location. Therefore, it is necessary to change the procedure for detecting the infrared LEDs through adaptive thresholding. Since the LEDs are virtually guaranteed to be the brightest source of infrared illumination, a binary filter is applied to intensity values within ten percent of the brightest value. If the intensity level at a particular point is in the top ten percent, then it appears as a one, otherwise it is a zero. From here, contiguous segments of ones are grouped together, and their geometric center and volume are calculated. The four points closest in volume are used as the four corners of the monitor. A naïve assignment of the identified LED positions is attributed to the corners of the actual screen. That is, the LED closest to the upper left corner of the scene camera is called the upper left corner, etc. A simple linear interpolation is utilized. This solution is robust enough for real world situations, where users are unlikely to have their heads turned at more than a 45 degree angle, or look at the screen from significantly oblique angles.

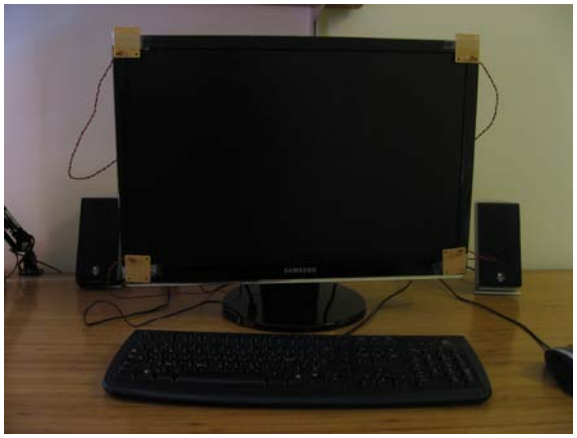

**Figure 3.19**

**Figure 3.20**

Figure 3.19-20: (19) Is a picture of the computer monitor as captured by the scene camera through the visible light filter, (20) is the same picture as captured by the scene camera through the infrared light filter.

# Chapter 4 Sources of Error

## Parallax Error

In the eye tracking device the scene camera and tracked eye are located on different optical paths, which introduces parallax error. If this problem is simplified to two dimensions, we can visualize it as in the Figure below. For example, if the system is calibrated for a plane at a given distance $d_c$ and the user fixates a plane at a further distance $d_f$, the system will not be able to compensate and the calibration will introduce a parallax error of distance $d_e$. This error depends on the difference between the calibrated distance and the fixated distance, as well as the distance $d_o$ between the optical axes of the tracked eye and the scene camera. We can solve for the relationship between these variables and the parallax error in degrees of visual angle relative to the scene camera ($\Theta_e$). Given the configuration in Figure 4.1, where the optical axes of the eye and scene camera are parallel, we know:

$$\Theta_f + \Theta_e = \tan^{-1}(d_e + d_o\, d_f) \qquad \text{(Equation 3)}$$

where $\Theta_f$ is the angle of the fixated point relative to the optical axis of the scene camera, which is

$$\Theta_f = \tan^{-1}(d_o\, d_f) \qquad \text{(Equation 4)}$$

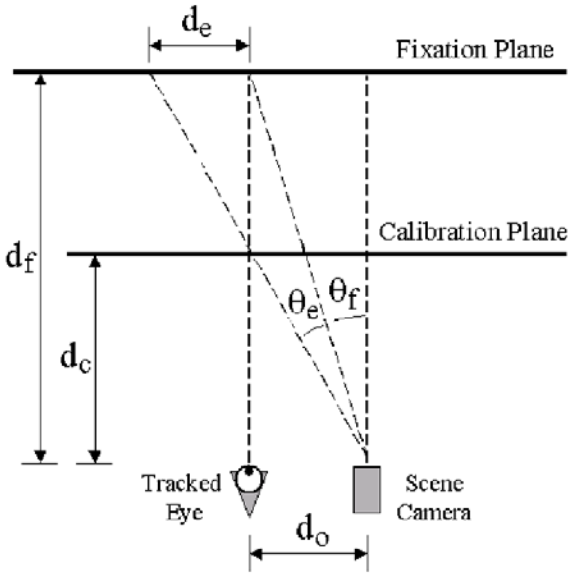


**Figure 4.1**

Figure 4.1: Figure depicting camera location with respect to the eye, the calibration plane, and the fixation plane.

The parallax error distance $d_e$ on the fixation plane can be determined via similar triangles:

$$\frac{d_e}{d_f - d_c} = \frac{d_o}{d_c} \quad \text{(Equation 5)}$$

$$d_e = \frac{d_o}{d_c}(d_f - d_c) \quad \text{(Equation 6)}$$

Therefore the visual angle of the parallax error $(\theta_e)$ in the scene camera is:

$$\theta_e = \tan^{-1}\left(\frac{\frac{d_o}{d_c}(d_f - d_c) + d_o}{d_f}\right) - \tan^{-1}\frac{d_o}{d_f} \quad \text{(Equation 7)}$$

The parallax error is plotted in Figure 4.2 as a function of the calibration and fixation distances for $d_o = 1.5$ inches to match the first generation prototype. It can be seen that the parallax error is zero when the fixated and calibrated distances are equal (on the diagonal) and then increases as they diverge. Interestingly, the parallax error rises faster as the calibration distance exceeds the fixation distance. This indicates that when the fixated and calibration distances are expected to diverge in a particular application, that the calibration should not be conducted in the middle of the working area as is common procedure, but rather it should be closer. If the minimum $d_{min}$ and maximum $d_{max}$ working distances are known, the optimal calibration distance is then $d_{min} + (d_{max} - d_{min})/3$. It is also clear from this result that the parallax error is inversely related to the calibration distance.

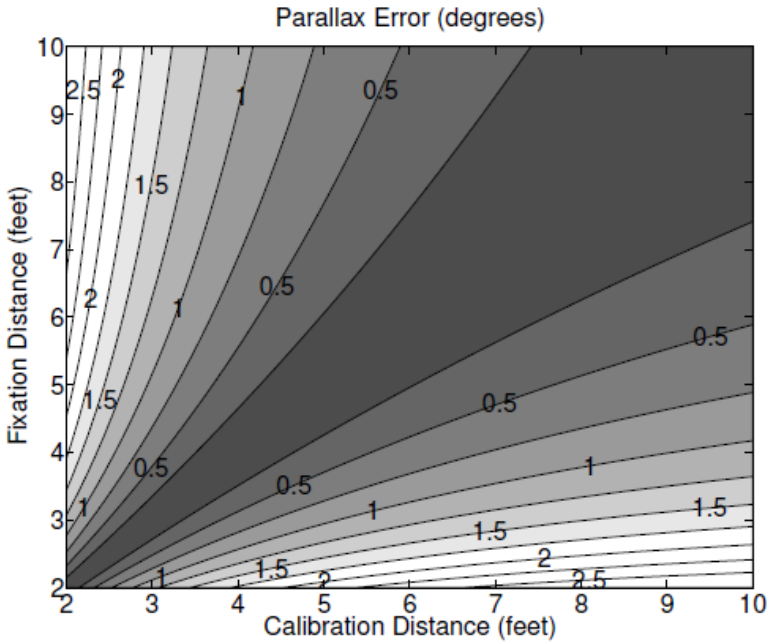


**Figure 4.2**

Figure 4.2: Graph of induced parallax error on given a calibration and fixation distance. Parallax error is equal along the labeled lines.

## Device Error

The device itself has several sources of error. There is error from the algorithm, the image quality, the lighting, the eye color of the subject and the movement of the camera on the steel wire (camera sway). Not all of these parameters were tested due to time restraints and the impracticality of testing certain conditions, such as the error induced by camera sway. The most useful metric for testing the accuracy of the device is to conduct controlled experiments on a large number of subjects, wherein as many parameters as possible are non-changing. A simple experiment was constructed in order to test the accuracy of the device, where the factors of lighting, eye color, camera sway, calibration plane, fixation plane, area of operation, and subject movement were controlled.

The accuracy tests were performed on subjects with brown eyes, in a controlled lighting environment. Parallax error is controlled by having each subject's head strapped into a chin rest at a fixed distance of six feet from the viewing screen. The screen was a projector screen onto which the display of a computer monitor was projected. The eye camera was positioned in approximately the same location for each of the subjects. For each trial, nine standard calibration points were used and kept constant across all subjects, with careful consideration to ensure the calibration points appeared in the same location in the scene camera across all subjects with as small a margin of error as possible. After calibration, a grid of sixteen points, at non-changing locations was displayed. The subject was asked to look at each of the points in succession, and to hit the space bar on a laptop in front of them while they were looking at the center of the point. Twenty trials were repeated for each of the ten subjects, giving a total of two hundred trials. Figure 4.3 shows the resulting error margins (in degree visual angle) at each of the points, with the standard deviation shown in parentheses.

Results showed that the larger the angle between the eye camera and the eye's point of gaze, the less accurate the device, suggesting that error is introduced by a skewed view of the eye. The eye camera was placed centered below the eye, aimed up, to avoid scene occlusion. Since the camera only records a single eye, there are two possible approaches for reducing the error introduced by a skewed eye image. Either the camera can be place directly on the axis of the eye, getting a direct head on view, or a hot mirror could be placed in front of the eye at a 45 degree angle downward. The eye camera could be placed below the hot mirror, and capture the on-axis view of the eye in the mirrors reflection. Both of these methods have a tradeoff between cost, ease of construction, and obtrusiveness of the device. The induced error is small enough that the approach used in this experiment is acceptable.

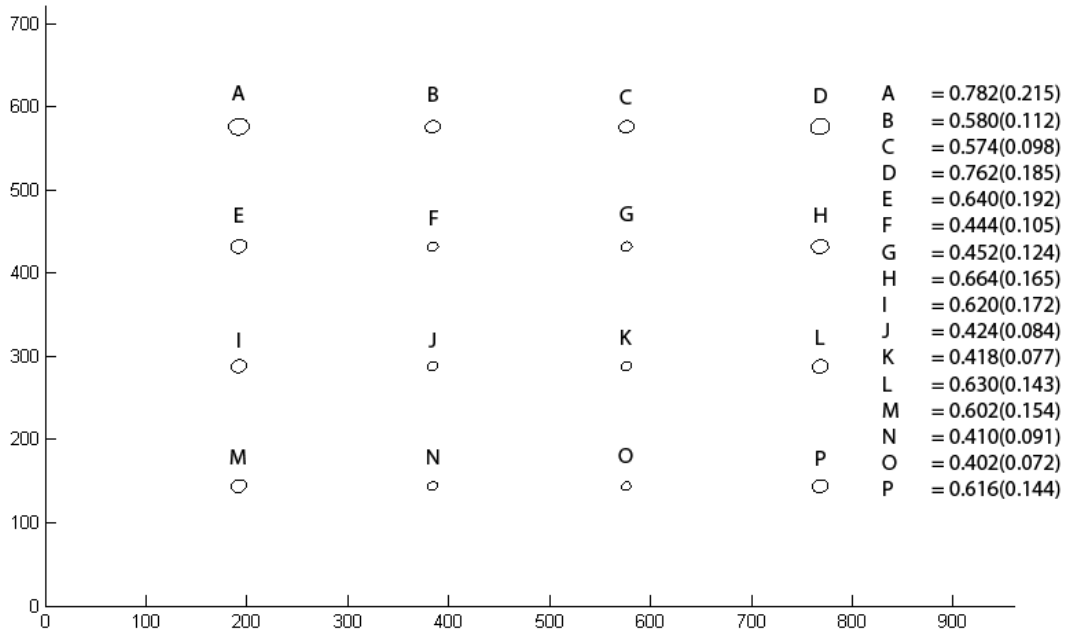


**Figure 4.3**

Figure 4.3: A graphical representation of the error margins at each of the 16 sampled points in the scene camera. The X and Y axes represent the location in the scene camera of the fixation point the error was calculated on, and are labeled in pixels, as the scene camera produces images which are 720x960 pixels. The error varied between 0.402-0.782 degrees of visual angle. The data shows that the further the eye is from directly pointed at the eye camera, the larger the induced error. Standard deviation values are shown in parentheses.

# Chapter 5 Discussion

There are many remaining factors to analyze; specifically, the calibration accuracy and the device accuracy analysis in different lighting settings and on subjects of differing eye color are two areas of note. Blue eyed subjects do not have as clear a contrast between the iris and the pupil as brown eyed subjects, making the device presumably more accurate on brown eyed individuals. This error, however, is not substantial, as many subjects with blue eyes have used the device with no more noticeable difficulty than the brown eyed subjects. There is also the possibility of discovering exact error on the device at different angles between the scene camera and the point of gaze (as in Figure 4.3).

It remains to be tested what error is introduced by the screen detection and interpolation, a very important measurement if the device is to be used in research using this feature. However, overall the data quite clearly shows that the device is extremely precise given its cost of construction. Commercial units claim a margin of error of about 0.5 degrees of visual angle. The EyesOn device, even at its greatest error, is still of comparable quality. A good approximation of the error induced in screen detection is magnifying the error by the ratio of the size of the quadrilateral identified by screen detection to the size of the screen's resolution (in pixels). In a perfect setup, where the screen resolution is identical to the scene camera, and the four corners of the monitor are the four corners of the scene image, the induced error by the screen detection algorithm would be zero. Another, more pressing matter is the fact that screen detection uses a naïve linear interpolation. This means that when a screen is not viewed head-on, significant error is induced because perspective is ignored. This can be overcome through the use of perspective correction, which will be implemented in future works.

There is also much room for improvement on the design of the hardware itself. Using a smaller scene camera would allow a closer placement to the axis of the eye, reducing the effect of parallax error. Additionally, as was mentioned in the device error analysis, error could be reduced through the use of a hot mirror placed at a 45 degree angle downward in front of the eye. This would allow the eye camera to capture the on-axis view of the eye and minimize the error induced from a skewed view, without obscuring the wearer's vision.

There are a plethora of applications for this device, from eye-typing for handicapped individuals to video-game immersion through the use of the eye as a method of input. By showing the viability of a low cost solution the hope is that eye tracking technology and usage will be proliferated in the HCI community.

## Cost

The second generation device cost a total of $318.04, an additional second generation device was created for $116.24. The second device was far less expensive because the majority of the Kodak® No. 87 Wratten filter (an $80.00 value) was not used to build a single device. In fact, only one eighth of the written filter is needed for the construction of one device. Additionally, purchasing the Logitech® Webcam 9000 Pro cameras through Amazon's® used good market place greatly reduced the cost. Additionally, money was saved purchasing the LEDs, tools, and steel wire by utilizing the Carnegie Mellon University Robotics Club's facilities. The retail value and links to purchase the materials are provided in the table below.

| Item | Name | Retail Price | Manufacturer |
|---|---|---|---|
| **1** | Logitech® Webcam Pro 9000 | $99.99 | Logitech |
| **2** | Logitech® Webcam Pro 9000 | $99.99 | Logitech |
| **3** | Kodak® No. 87 Wratten Filter 75mm x 75mm | $89.00 | Edmund Optics |
| **4** | Journey Clear Lens Safety Glasses | $1.56 | Radians |
| **5** | Hot Mirror 12.5mm x 12.5mm | $27.50 | Edmund Optics |
| | | Total | |
| | Expenses | $318.04 | |

**Table 1**

## Works Cited

Babcock, J., and J. Pelz. "Building a lightweight eyetracking headgear." *ACM Eye Tracking Research and Applications Symposium.* San Antonio, 2004. 109-114.

Beymer, D., and M. Flickner. "Eye gaze tracking using an active stereo head." *IEEE Computer Vision and Pattern Recognition*, 2003: 451-458.

Bolles, Martin A. Fischler and Robert C. *Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography.* 1981.

Glenstrup, A., and T. Engell-Nielse. "Eye controlled media: present and future state." *University of Copenhagen Thesis, Denmark*, 1995.

Hansen, D., and A. Pece. "Eye tracking in the wild." *Computer Vision and Image Understanding*, 2005: 155-181.

ICNIRP. *Guidelines on Limits of Exposure to Broad-Band Incoherent Optical Radiation (0.38 to 3µm). Health Physics.* 3. Vol. 73. 1997.

—. *Light-Emitting Diodes (LEDS) and Laser Diodes: Implications for Hazard Assessment. Health Physics.* 6. Vol. 78. 2000.

Jacob, R. *Eye tracking in advanced interface design in virtual environments and advanced.* Oxford University Press, 1995.

Kaufman, A., A. Bandopadhay, and B. Shaviv. "An eye tracking computer user interface." *IEEE Proceedings of Research Frontier in Virtual Reality Workshop* (IEEE Proceedings of Research Frontier in Virtual Reality Workshop), 1993.

Morimoto, C, A. Amir, and M. Flickner. "Detecting eye position and gaze from a single camera and two light sources." *16th International Conference on Pattern Recognition.* 2002. 314-317.

Morimoto, C., D. Koons, A. Amir, and M. Flickner. "Pupil detection and tracking using multiple light sources." *Image and Vision Computing*, 2000: 331-335.

Newman, R., Y. Matsumoto, S. Rougeaux, and A. Zelinsky. "Real-time stereo tracking for head pose and gaze estimation." *Fourth IEEE International Conference on Automatic Face and Gesture Recognition.* 2000. 122-128.

Pelz, J., R. Canosa, J. Babcock, D. Kucharczyk, A. Silver, and D. Konno. "Portable eyetracking: a study of natural eye movements." *SPIE, Human Vision and Electronic Imaging.* San Jose, 2000. 566-582.

Shih, S., and J. Liu. "A novel approach to 3-d gaze tracking using stereo cameras." *IEEE Transactions on Systems, Man and Cybernetics*, 2004: 234-245.

Sliney, D., and M. Wolbarst. *Safety with Lasers and Other Optical Sources.* New York: Plenum Press, 1980.

Stampe, D. M. "Heuristic filtering and reliable calibration methods for video-based pupiltracking systems." *Behavior Research Methods, Instruments, and Computers*, 1993: 137-142.

Tian, Y., T. Kanade, and J. Cohn. "Dual-state parametric eye tracking." *4th IEEE International Conference on Automatic Face and Gesture Recognition.* 2000. 110-115.

Yarbus, A. L. *Eye Movements and Vision.* New York: Plenum, 1967.

Young, L., and D. Sheena. *Survey of eye movement recording methods.* 1975.

Zhai, S., C. Morimoto, and S. Ihde. "Manual and gaze input cascaded (magic) pointing." *SIGCHI Conference on Human Factors in Computing Systems.* 1999. 246-253.

Zhu, Z., and J. Qiang. "Robust real-time eye detection and tracking under variable lighitng conditions and various face orientations." *Computer Vision and Image Understanding*, 2005: 124-154.

## Acknowledgements

I would like to take this opportunity to thank all of those who helped with conducting research, building the hardware, implementing the software, and writing the thesis, as well as those who helped in any and every meaningful way.

First and foremost, I would like to thank my thesis adviser, Dr. Tai Sing Lee, whose invaluable advice was paramount to the completion of the thesis. I would also like to thank him for accepting me as a thesis student even after the official deadlines for thesis proposals had passed. Thanks also to Mark Stehlik, whose support throughout the years has been an essential component leading up to this thesis, and for directing me to my adviser.

The two most essential contributors to the thesis outside of my adviser have been Florence Doo and Christopher Lu. I would like to thank Florence Doo for her constant support from the conception of the project, until its completion. Without her I would not have had a thesis to complete, nor would I have ever completed it.

Similarly, I cannot thank enough my good friend and roommate Christopher Lu for his constant support on both the creation of the hardware and the implementation of the software. I would also like to take this opportunity to thank the Carnegie Mellon University Robotics Club, for being kind enough to host and assist with the construction of the devices. Without him many key insights and improvements would have remained undiscovered.

I would like to thank Dan Schafer for his advice on applying for thesis candidacy, and his contributions to the construction of the thesis documents.

Thanks to: Anson Wang, Eddie Ng, Tamir Sen, and Lisa Sun. Their strong encouragement and support kept me motivated and allowed me to keep my sanity throughout the year.

Finally, thank you to the rest of my friends and family, for supporting be throughout my life and making this achievement possible.

The following individuals participated as experimental subjects for the testing of the device:

- Akira Yashiro
- Anson Wang
- Christopher Lu
- Elaine Lee
- Hyun-Soo Lee
- Idora Sopin-Vilme
- Joshua Wise
- Justin Chen
- Philip Shirey
- Wuming Xie