

EyesOn Mobile Eye Tracking

A low cost alternative to high end commercial eye tracking units

by Tom Jacques



Table of Contents

| | |
|-------------------------------|----|
| Abstract | 3 |
| Infrared Eye Tracker Hardware | 4 |
| Generation 1 | 4 |
| Generation 2 | 4 |
| Gaze Interpretation | 5 |
| Noise reduction | 5 |
| Corneal Reflection Detection | 5 |
| Pupil contour detection | 6 |
| Ellipse fitting | 7 |
| Calibration | 7 |
| Screen detection | 7 |
| Sources of Error | 8 |
| Parallax analysis | 8 |
| Device analysis | 9 |
| Discussion | 10 |

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 twenty thousand dollars or more. I hope to deliver a method for creating a very low cost, easy to assemble mobile eye tracking unit which uses a USB interface, and to create robust software to analyze the video data streams. The novel contribution is a method for screen detection on the mobile eye-tracker by the use of a hot swappable filter and Infrared LEDs, and rigorous analysis of the accuracy of the tracker in various lighting conditions. This will allow real time eye-tracking as a method of computer input on a head mounted mobile eye-tracker, and create the opportunity for quality research at a low, affordable cost.

Infrared Eye Tracker 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.

The first design consideration after having chosen to use a head-mounted system was the configuration of the head gear. The most significant issue was where to mount the cameras. Two Logitech 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 to the USB camera board with the help of a resistor, which reduces the current through the LED to ensure the light intensity projected by the LED is not harmful. The infrared blocking filter in the USB camera was removed from the eye camera, and the filter was replaced with an infrared passing filter. 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. The smallest distance achievable from the affixed camera to the eye was approximately 1.2 inches (3 cm). Some commercial units utilize hot mirrors, bits of 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, these products can be expensive and difficult to obtain in small quantities. In order to reduce costs, the USB camera was affixed using steel reinforcement wire to point with a direct line of sight towards the eye. Since the camera is extremely light, and the steel reinforcement cable is comparatively strong and rigid, there is

not a significant amount of camera sway due to head movement, and what 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. Because 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.

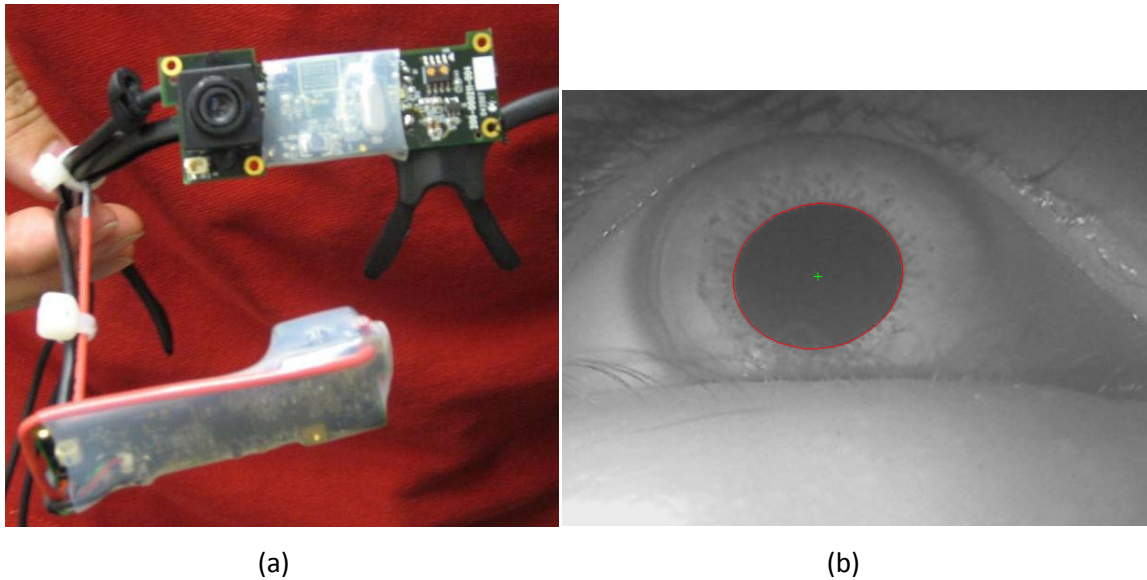


Figure 1: Eye tracker and the captured images. (a) Head-mounted eye tracker. (b) Image of the user's right eye illuminated with infrared light. Note the clearly contrasted dark pupil and the best fitting ellipse corresponding to the algorithms analysis of the image.

The second generation prototype was very similar. A rubber goggled strap was used to snugly hold the device to the subjects head, the hot-mirror was removed from both the scene and the eye camera, 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. A detailed step by step guide on the construction of the device will be covered in appendix B, and a detailed step by step guide on the use of the device will be covered in appendix C.

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.
- Remove bad pupil edges using the Random Sample Consensus (RANSAC) algorithm.
- Find the best fitting ellipse to the remaining feature points, the center of the ellipse is the pupil center.

Noise Reduction

I wanted to design my software system to be robust and work with any web cameras. However, because of the wildly different USB web cameras on the market, I wanted to make sure that even with a very low quality web camera, imaging could be used for eye tracking. I reduce image noise by applying a 5 x 5 Gaussian filter with a standard deviation of 2 pixels.

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.

Pupil Contour Detection

- Pupil candidate feature points are iteratively detected by following rays extending outward from a 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.

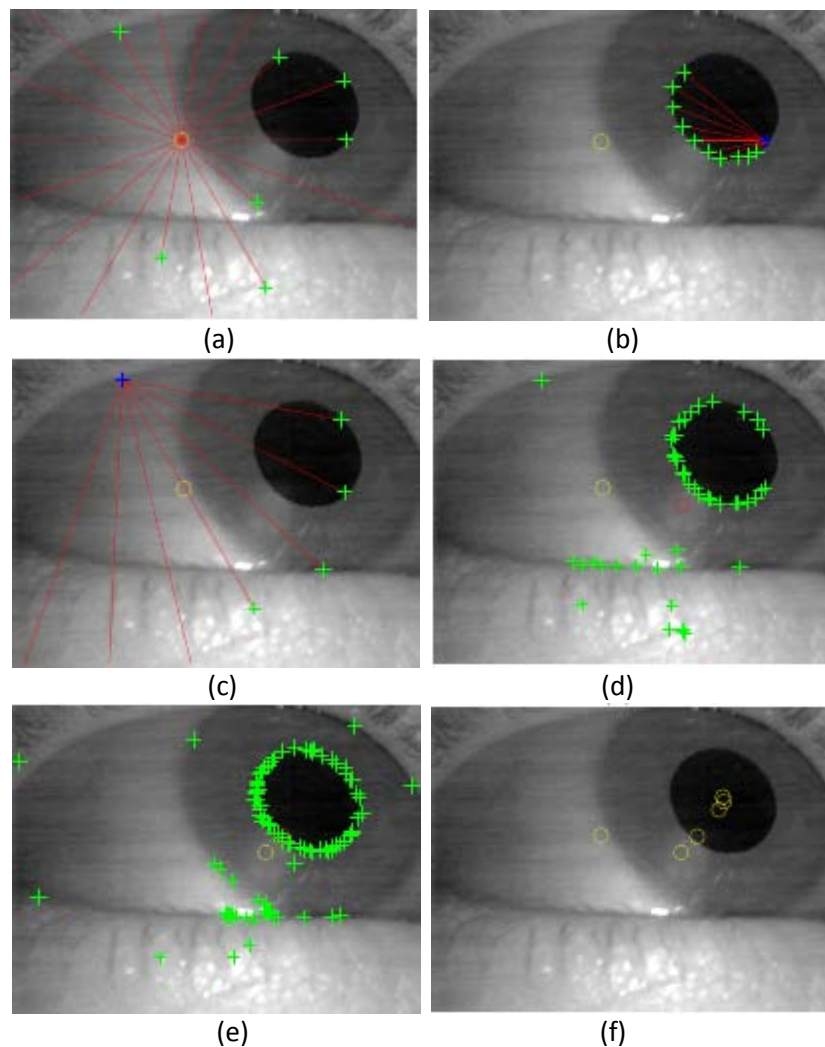


Figure 2: (a) shows the original starting point, and the burst of rays in all directions, with green +'s indicating where a feature point was detected. The starting point of an iteration is shown as a yellow circle. (b) and (c) show the ray spread starting from a discovered feature point (a blue +), with new feature points detected. (d) shows the end of an iteration, with all feature points detected shown as green +'s, the yellow circle was the starting point, and the red circle is the new starting point. (e) shows the same information at the end of the second iteration. (f) 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.

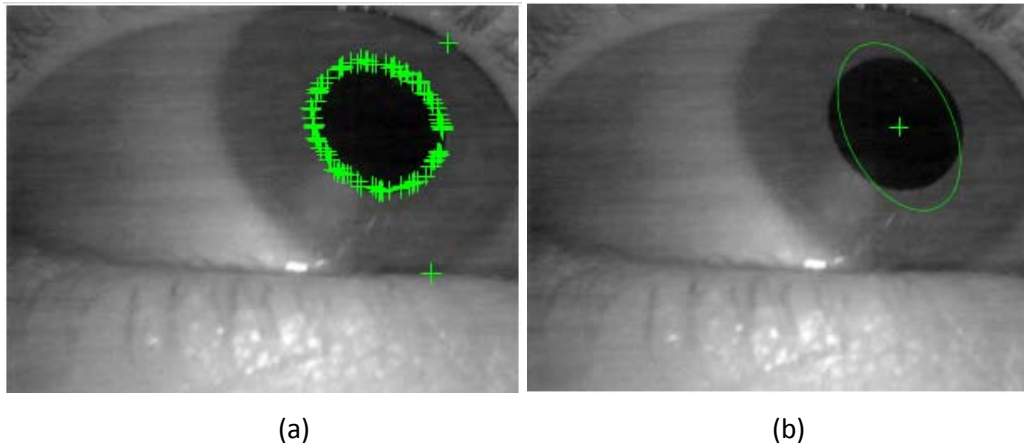


Figure 3: (a) Set of feature points detected by the algorithm,
(b) The geometric best fitting ellipse to these detected points.

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, we apply the Random Sample Consensus (RANSAC)¹ paradigm for model fitting. 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.

¹ Martin A. Fischler and Robert C. Bolles (June 1981). "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography"

Calibration

Calibration is accomplished by having nine fixed entries. That is, nine points within the scene camera are correlated to the center of the pupil in nine eye frames. The position of gaze for any new pupil center point is a quadratic transformation on the nine calibrated indices.

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, we can detect the corners of the computer monitor 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.

Sources of Error

Parallax Analysis

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 (θ_e). Given the configuration in Figure 4(a), where the optical axes of the eye and scene camera are parallel, we know:

$$\theta_f + \theta_e = \tan^{-1}(d_e + d_o d_f) \quad (\text{Equation 1})$$

where θ_f is the angle of the fixated point relative to the optical axis of the scene camera, which is

$$\theta_f = \tan^{-1}(d_o d_f) \quad (\text{Equation 2})$$

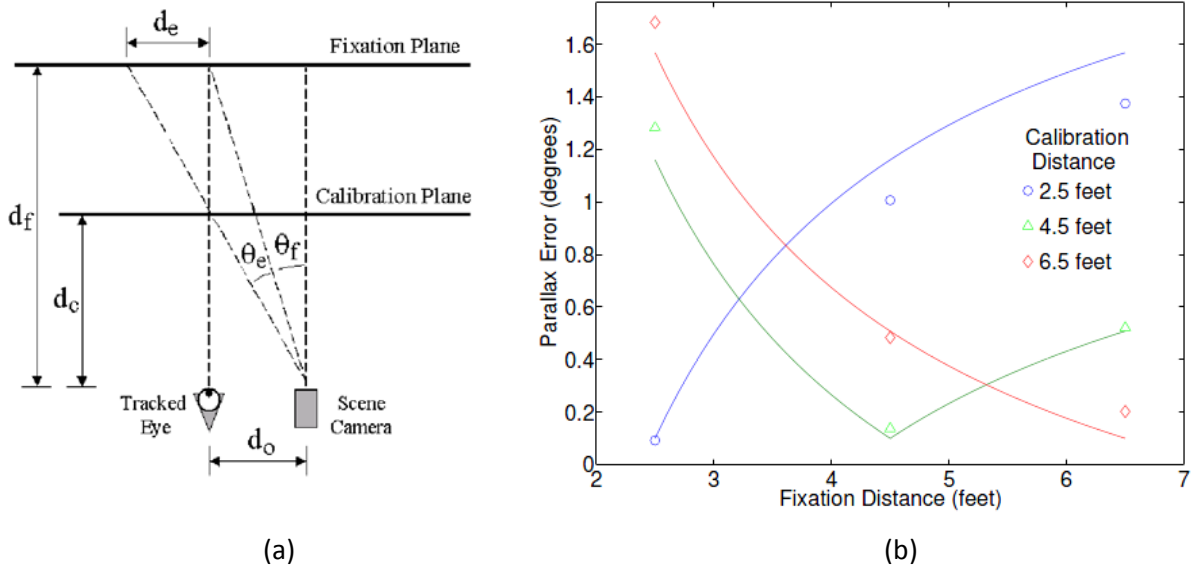


Figure 4: (a) Figure depicting camera location with respect to the eye, the calibration place, and the fixation place. (b) Chart of induced parallax error at a fixation distance given a set calibration distance.

Device Analysis

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. 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 calibration points were chosen by the following process. Split the screen into nine equal segments, and randomly choose a point from each of the nine segments. 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 two hundred trials. Figure 5 shows the resulting error margins (in degree visual angle) at each of the points.

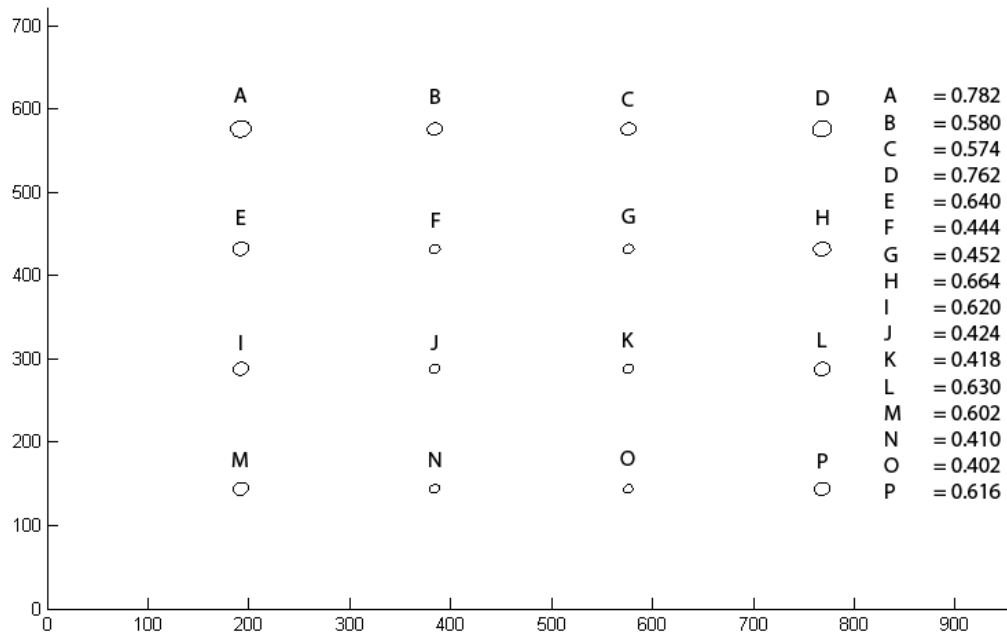


Figure 5: This figure shows a graphical representation of the error margins at each of the 16 sampled points in the scene camera. The error varied between 0.402°-.782° viewing angle

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.

There is also much room for improvement on the design of the hardware itself. The incorporation of an optical zoom lens would ensure a higher quality image of the eye, however optical lenses are expensive and can be hard to affix to the eye tracker. A second improvement to the device would be to use a hot mirror instead of attaching the camera with steel wire. This would reduce camera sway significantly and allow the camera to be used in more mobile settings, such as playing sports, or jogging. As it stands now, any significant or quick head movement will greatly affect the accuracy of the device, making those frames unusable.