Chapter 3

3. Eye Tracking Instrumentation

3.1 Overview

The introduction and background in the previous chapters provided context in which eye tracking systems have been used to study how people look at images. This chapter provides some detail about the eye tracking equipment used for this thesis and presents an overview of the typical accuracy achieved with a head-free eye tracking system. The final sections will describe the post-processing applied to the raw eye movement data in order to remove blink and saccade intervals, and to correct for offsets resulting from a shift or translation of the headgear.

3.2 Bright Pupil Configuration – Theory of Operation

The most common eye tracking technique uses bright pupil illumination in conjunction with an infrared video-based detector (Green, 1992; Williams and Hoekstra, 1994). This method is successful because the retina is highly reflective (but not sensitive) in the near infrared wavelengths. Light reflected from the retina is often exhibited in photographs where the camera’s flash is aimed at the subject’s line of sight. This produces the ill-favored “red eye.” Because the retina is a diffuse retro-reflector, long-wavelength light from the flash tends to reflect off the retina (and pigment epithelium),
and, upon exit, back-illuminates the pupil. This property gives the eye a reddish cast (Palmer 1999).

Bright-pupil eye tracking purposely illuminates the eye with infrared and relies on the retro-reflective properties of the retina. This technique also takes advantage of the first-surface corneal reflection, which is commonly referred to as the first Purkinje reflection, or P1, as shown in Figure 3.1 (Green, 1992). The separation between pupil and corneal reflection varies with eye rotation, but does not vary significantly with eye translation caused by movement of the headgear. Because the infrared source and eye camera are attached to the headgear, P1 serves as a reference point with respect to the image of the pupil (see Figure 3.2). Line of gaze is calculated by measuring the separation between the center of the pupil and the center of P1. As the eye moves, the change in line of gaze is approximately proportional to the line of change in this separation. The geometric relationship (in one-dimension) between line of gaze and the pupil-corneal reflection separation (PCR) is given in Equation 1:

\[ \text{PCR} = k \sin(\theta) \]  

(1)

\( \theta \) is the line of gaze angle with respect to the illumination source and camera; \( k \) is the distance between the iris and corneal center which is assumed to be spherical. In this configuration the eye can be tracked over 30-40 degrees (ASL manual, 1997).
3.3 Video-Based Eye Tracking

The Applied Science Laboratory Model 501 eye tracking system was used for all experiments in this thesis. The main component includes the head mounted optics (HMO), which houses the infrared LED illuminator, a miniature CMOS video camera (sensitive to IR), and a beam splitter (used to align the camera so that it is coaxial with
the illumination beam). An external infrared reflective mirror is positioned in front of the subject’s left eye as shown in Figure 3.3. This mirror simultaneously directs the IR source toward the pupil and reflects an image of the eye back to the video camera.

![Image of the Applied Science Laboratory Model 501 eye tracking system with labeled parts: Head-mounted optics (includes IR source and eye camera), Scene camera, Infrared reflective, visible passing, mirror, head-tracker receiver.]

Figure 3.3 – The video-based Applied Science Laboratory Model 501 eye tracking system.

A second miniature CMOS camera is mounted just above the left eye to record the scene from the subject’s perspective. This provides a frame of reference to superimpose a pair of crosshairs corresponding to the subject’s point of gaze (Figure 3.4). Above the scene camera a small semiconductor laser and a two-dimensional diffraction grating are used to project a grid of points in front of the observer. These points are used to calibrate the subject’s eye movements relative to the video image of the scene. Since the laser is attached to the headgear, the calibration plane is fixed with respect to the head. The laser points provide a reference for the subject when asked to keep the head still relative to a stationary plane such as a monitor.

Eye and scene video-out from the ASL control unit is piped through a picture-in-picture video-mixer so that the eye image can be superimposed onto the scene image.
(Figure 3.4). This reference provides important information regarding track losses, blinks, and extreme eye movements. The real-time eye and scene video images are recorded onto Hi8 videotapes using a Sony 9650 video editing deck.

Because the system is based on NTSC video signals, gaze position is calculated at 60 Hz (video field rate). The ASL software allows for variable field averaging to reduce signal noise. Since the experiments in this thesis were not designed to investigate the low-level dynamics of eye movements, gaze position values were averaged over eight video fields. This yielded an effective temporal resolution of 133 msec.
3.4 Integrated Eye and Head Tracking

Both horizontal and vertical eye position coordinates with respect to the display plane are recorded using the video-based tracker in conjunction with a Polhemus 3-Space Fastrak magnetic head tracker (MHT). Figure 3.5 shows an observer wearing the headgear illustrated in Figure 3.3.

![Figure 3.5 – Setup of the magnetic transmitter positioned behind the observer.](image)

Gaze position (integrated eye-in-head and head-position & orientation) is calculated by the ASL using the bright pupil image and a head position/orientation signal from the MHT. This system uses a fixed transmitter (mounted above and behind the subject in Figure 3.5) and a receiver attached to the eye tracker headband. The transmitter contains three orthogonal coils that are energized in turn. The receiver unit contains three orthogonal Hall-effect sensors that detect signals from the transmitter. Position and orientation of the receiver are determined from the absolute and relative strengths of the transmitter/receiver pairs measured on each cycle. The position of the sensor is reported
as the \((x, y, z)\) position with respect to the transmitter, and orientation as azimuth, elevation, and roll angles.

### 3.5 Defining the Display Plane Relative to the Magnetic Transmitter

Eye-head integration software reports gaze position as the X-Y intersection of the line-of-sight with a defined plane. In order to calculate the gaze intersection point on the display screen, the position and orientation of the display is measured with respect to the transmitter. This is done by entering the three-dimensional coordinates of three points (in this case, points A, B, and C on the 9 point calibration grid) on the plane into the ASL control unit as illustrated in Figure 3.6. Using the Fastrak transmitter as the origin, the distance to each of the three points is measured and entered manually. Observer’s real-time gaze intersection on the display is computed by the ASL and the coordinates are saved to a computer for off-line analysis.

![Figure 3.6](image)

Figure 3.6 – The viewing plane is defined by entering the three-dimensional coordinates of three points (in this case, points A, B, and C of calibration target) on the plane into the ASL control unit.
3.6 Eye-Head Calibration

The eye tracker was calibrated for each subject before each task. Calibrating the system requires three steps; 1) measuring the three reference points on the calibration plane as described in section 3.5, 2) defining the nine calibration points with respect to the video image, and 3) recording the subject's fixation for each point in the calibration target.

The accuracy of the track is assessed by viewing the video calibration sequence and by plotting the fixation coordinates with respect to the actual calibration image. Because the scene camera is not coaxial with the line of sight (leading to parallax errors), calibration of the video signal is strictly correct for only a single distance. For experiments in this thesis, gaze points fell on the plane of the display. Because viewers did not change their distance from the display substantially, parallax errors were not significant in the video record.

The gaze intersection point calculated by the ASL from the integrated eye-in-head and head position/orientation signals is not affected by parallax. After initial eye calibration, the gaze intersection is calculated by projecting the eye-in-head position onto the display, whose position and orientation were previously defined.

Figure 3.7 plots the X-Y position of a subject looking at a nine-point calibration target displayed on a 50” Pioneer Plasma display (more detail about the display is given in Chapter 4). The vector coordinates from the eye, which are reported in inches by the MHT/eye tracker, are converted to pixel coordinates relative to the image and screen resolution. Note that the upper-left point (point 1) shows an artifact resulting from a blink.
Figure 3.8 shows the fixations plotted on a 17-point target whose points fall between the initial 9-point calibration nodes. In viewing the 50” display, points near the edge of the screen require a large angle (greater than 20°) from the central axis. Points three and six demonstrate how accuracy is affected due to a track-loss of the first surface reflection. The 17-point fixation data for all subjects was recorded at the end of the experiment, which was typically one hour after initial calibration. In this example, the headgear has moved slightly during the experiment, resulting in a small offset toward the upper-right.


3.7 Fixation Accuracy

One disadvantage of using a head-free system is that the accuracy of the eye movement record can vary substantially from subject to subject. The differences are not systematic and vary from point to point since each observer’s cornea and retina are unique. To estimate the accuracy of the track across subjects, the average angular distance from the known calibration points and fixation record was calculated for both 9 and 17-point targets. Accuracy was examined on data acquired from two displays; a 50” Pioneer Plasma Display (PPD), and a 22” Apple Cinema Display (ACD). The PPD totaled 1280 x 768 pixels with a screen resolution of 30 pixels per inch. Viewers sat approximately 46 inches away from the display, yielding a visual angle of 50 x 30°. This distance results in approximately 26 pixels per degree. The ACD totaled 1600 x 1024 pixels with a screen resolution of 86 pixels per inch. Viewers sat approximately 30 inches
from the display, yielding a visual angle of 34 x 22°. This resulted in approximately 46 pixels per degree.

Figure 3.9 plots average angular deviation (in degrees) for 26 observers viewing the 9-point calibration grid on the PPD and 7 observers viewing the same target on the ACD. Center point 5 resulted in smaller error compared to corner points 1, 3, 7 and 9. The average angular deviation across all subjects and both displays for the 9-point target was 0.73°. Point 3 (upper-right) resulted in the lowest accuracy for targets displayed on the PPD. This error is likely due to a large, asymmetrical specular reflection that results from large eye movements. An example is illustrated in the eye-image above point 3.
Figure 3.10 plots average angular deviation (in degrees) for 36 observers viewing the 17-point calibration grid on the PPD and 17 observers viewing a 17-point grid on the ACD. Because points 1-9 in the 17-point grid are farther from the center than points 1-9 in the 9-point grid (compare Figures 3.7 & 3.8), larger errors often result. The average angular deviation across all subjects and both displays for the 17-point target was 1.17°.

It is typical for points near the edge of the display to result in poor accuracy. However, Figure 3.9 and 3.10 report the worst-case error since angular deviations were calculated on raw data eye movement data that include blink artifacts and offset due to movement or
translation of the headgear. Figure 3.11 plots a histogram of angular deviation across all subjects, both calibration targets, and both displays.

Figure 3.11 shows that, on-average, the accuracy of the eye tracker is roughly within 1 degree of the expected target, and that eye movements toward the extreme edges of the screen can produce deviations as large as 5.3°. An average error of 1° agrees with the accuracy reported in the ASL user manual (ASL manual, 1997, pg. 51). The reader should keep in mind that experiments in this thesis did not require subjects to spend much time looking near the edges of the screen. Most of the tasks required attention within the boundary of the smaller 9-point grid. The following sections describe some of the post-processing applied to the raw eye movement data in order to remove blink and saccade intervals, and to correct for offsets resulting from a shift or translation of the headgear.
3.8 Blink Removal

Along with horizontal and vertical eye position, the ASL also reports the size of the pupil for each field. This is useful because the pupil diameter can be used to detect and remove blink artifacts such as those shown in Figure 3.7. An algorithm was written in Matlab to parse out regions of the data where the pupil diameter was zero. Figure 3.12 plots a subject’s fixations over ~18 seconds before and after blink removal. Green lines indicate vertical eye position as a function of time. Blue lines indicate pupil diameter as reported from the ASL. Segments of the pupil record equal to zero were used as pointers to extract blink regions. Because of field averaging, a certain delay resulted before detecting the onset and end of a blink. The Matlab algorithm used the average width of all blinks within each trial to define the window of data to remove for each blink. Red markers at the base of the blink spikes indicate the onset of a blink as detected by the algorithm.

Figure 3.12 – The spikes in the left graph (green line) indicate regions in the vertical eye position record where blinks occurred. The blue lines indicate the pupil diameter. Red dots indicate the start of the blink as indicated by the algorithm. The graph to the right plots the data with blinks removed.
Figure 3.13 plots X and Y fixation coordinates before and after blink removal from the data shown in Figure 3.12. The cluster of blue dots indicates where the subject was looking. In this example the task was to adjust a small patch in the center of the image to appear achromatic, hence the large cluster of fixations in the center. More detail about this task is given in Chapter 6.

Figure 3.13 – Fixations plotted before (upper plot) and after (lower plot) blink removal.
3.9 Saccade Detection and Removal

As stated earlier, the ASL software allows for variable field averaging to reduce signal noise. While averaging over eight video fields is optimal for the video record, it does result in artifacts that can obscure the data when plotting fixation density or compiling a spatial histogram of fixation position across multiple subjects. Typically, the sampled data between fixations (during saccades) is unwanted because it obscures the actual eye position. A simple saccade removal algorithm was written to extract these unwanted data points. Figure 3.14 shows examples of fixation data plotted before and after saccade removal. The data removal is based on a moving window which compares the maximum Euclidian distance of three successive points to the maximum tolerance distance defined by the program. In this example, the maximum distance was 13 pixels. Again, this is an example taken from the patch adjustment task described in Chapter 6.

Figure 3.14 – The top image shows an example of the raw eye movement data. The bottom image shows the result with blinks and samples in-between fixations removed.
3.10 Offset Correction

Despite efforts to get an optimal calibration, the MHT accuracy can still drift over time due to the headgear settling or shifting. This often results in an additive offset as illustrated in Figure 3.8 and Figure 3.15. Ideally, a single offset correction would be applied to the entire data file. However, this does not always provide the best results since the headgear may shift more than once during the experiment. To get the most accurate correction, an offset should be applied relative to some known target in the viewing plane; such as a central fixation point. For the achromatic patch adjustment task (discussed in Chapter 6), an offset correction was applied with respect to the center of the adjustment patch for each of the 72 images across 17 observers. The following description illustrates how this was done.

For this example it is clear that the large cluster of fixations should fall over the central adjustment patch. However, because the headgear shifted during the experiment, the offset to the upper-left is evident in the MHT record. This error typically does not affect the video record since the separation between the eye and specular reflection do not

Figure 3.15 – Shows an example of the eye movement data where an offset occurred.
vary significantly when the headgear slips (discussed in section 3.2). However, when headgear is bumped, or moved, it shifts the MHT receiver and offsets the calculated eye position. Rather than stop the experiment to recalibrate, it was possible to continue on with the expectation of correcting for the offset later.

Since most a large number of fixations occurred on the central patch, a program was written to apply a correction on a per-image basis if an offset was necessary. First, the image was displayed with raw fixation data (in this example blink segments and saccade intervals were removed). Next a crosshair appeared in which the user selects the region of the fixation data intended to be located at the center of the image. The offset is then applied and re-plotted for verification, as shown in Figure 3.17.

Figure 3.16 – Shows an example of crosshairs used to identify the central fixation cluster, which should be located over the gray square in the center of the image.
Along with blink and saccade data removal, a similar method of offset correction was applied to the other experiments using fixation landmarks such as buttons and sliders as offset origins, or the offset was manually applied by referencing the video footage. In the achromatic patch selection task, all mouse movements were recorded, and the last mouse position (which the observer was sure to be fixating) was used as an offset marker.

3.11 Data Smoothing and Visualization

The Applied Vision Research Unit at the University of Derby has recently collected eye movement data from 5,638 observers looking at paintings on exhibit at the National Gallery in London. This exhibition is the world’s largest eye tracking experiment and has generated so much data that researchers were faced with the problem of trying to visualize subjects’ fixation data beyond conventional statistics such as fixation duration and number of fixations. Wooding (2002) has presented this data in the form of 3-D fixation maps which represent the observer’s regions of interest as a spatial
map of peaks and valleys. This thesis has expanded on Wooding’s visualization techniques to include a suite of Matlab tools aimed at plotting 3-D fixation surfaces over the 2-D image that was viewed. The following sections describe the visualization approach.

The ASL control unit reports the horizontal and vertical eye position projected onto the display in inches for each sampled point. These values are converted to pixel coordinates relative to the image. Fixation distribution across multiple observers (with blinks and saccade intervals removed) is converted into a 2D histogram (1 pixel bin size) where the height of the histogram represents the frequency of fixation samples for a particular spatial location. Because the number of pixels covered by the fovea varies as a function of viewing distance, the data is smoothed with a Gaussian convolution filter whose shape and size is determined by the pixels per degree for a display at a given viewing distance. Table 3.1 provides sample calculations used to compute pixels per degree for the two displays.

<table>
<thead>
<tr>
<th>Monitor</th>
<th>Pioneer Plasma</th>
<th>Apple Cinema</th>
</tr>
</thead>
<tbody>
<tr>
<td>viewing distance (inches)</td>
<td>46</td>
<td>30</td>
</tr>
<tr>
<td>Calculations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>screen dimensions (pixels)</td>
<td>1280 768</td>
<td>1600 1024</td>
</tr>
<tr>
<td>screen dimensions (inches)</td>
<td>43 25.8</td>
<td>18.5 11.84</td>
</tr>
<tr>
<td>pixels per inch</td>
<td>30 30</td>
<td>86 86</td>
</tr>
<tr>
<td>visual angle</td>
<td>50 30</td>
<td>34 22</td>
</tr>
<tr>
<td>pixels per degree</td>
<td>25 25</td>
<td>46 46</td>
</tr>
<tr>
<td>Gaussian width at half height (pixels)</td>
<td>16</td>
<td>34</td>
</tr>
</tbody>
</table>
The width of the Gaussian function at half-height is given in Table 3.1. The top images in Figure 3.18 show sample data from an image viewed on a Pioneer Plasma Display. These maps plot normalized frequency of fixation across 13 subjects before and after smoothing the 2D histogram. The bottom image shows a color contour plot of the smoothed data.

Figure 3.18 – Shows normalized frequency of fixation across 13 observers convolved with Gaussian filter whose width at half-height is 16 pixels. The filter corresponds to a 2 degree visual angle at 46 inches for a 50” Pioneer Plasma Display with a resolution of 30 pixels per inch.
3.12 Conclusions

This chapter provided description of the eye tracking equipment used for this thesis. The accuracy of the track across two displays was roughly within 1 degree of the expected target, and eye movements near the edges of the screen produced deviations as large as 5.3°. This result agrees with the tracking accuracy reported by the manufactures. A library of Matlab functions was developed to remove blinks and extract saccade intervals resulting from video field-averaging. While no rotational correction was applied, a simple offset was used to improve the accuracy of the eye movement data in cases where the headgear shifted during the experiment.