

Eye-Gaze Tracking With Free Head Motion

by

Craig Hennessey

B.A.Sc., Simon Fraser University, 2001

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

Master of Applied Science

in

THE FACULTY OF GRADUATE STUDIES

(Electrical and Computer Engineering)

The University of British Columbia

August 2005

© Craig Hennessey, 2005

Abstract

Knowledge of the eye-gaze position of a subject may allow machines to interact with humans in a more intuitive and natural fashion. The goal of this thesis is to develop a minimally restrictive eye-gaze tracking system using a single camera, that is 3D model based, uses multiple glint light sources and that can re-acquire the eye position rapidly after head movements.

The implemented system estimates the gaze position on a computer screen of a subject, solely by tracking features in images of their face and eye. The system is capable of estimating the point of gaze independent of head position with an update rate of 15 Hz. The processing time required for a full sized image frame is 110 ms which reduces to 28 ms when the region of interest, used to reduce the size of image to be processed, is locked on the eye. A delay of 3 frames before reprocessing the full image was added to avoid losing lock due to eye blinks. After determining the region of interest has lost the eye due to translation, the system is able to reacquire the eye within 110 ms. A novel eye-glass reflection compensation algorithm allows people who wear corrective lenses to use the system. Insensitivity to ambient lighting conditions is achieved through infrared system lighting and optical filters.

The system was tested with the eye located in six different positions resulting in average accuracies ranging from 0.54° to 1.1° of visual angle. The system was also tested on twelve different subjects of various ethnicities, gender and eye-glass use, with an average error 0.87° and an average maximum error of 1.85° .

The system we have developed is a fully functional eye-gaze tracker which is based on a single camera, compensates for reflections off eye-glasses, and handles varying ambient lighting conditions. Our current system may be used as a testbed for further eye-gaze tracking improvements and as a platform for developing eye-gaze aware applications.

Table of Contents

Abstract	ii
Table of Contents	iii
List of Tables	vi
List of Figures	vii
Acknowledgments.....	x
1. Background.....	1
1.1. Introduction	1
1.2. Eye Gaze Tracking Applications.....	1
1.3. Key Concepts in Eye Gaze Tracking.....	2
1.3.1. <i>Anatomy and Physiology of the Human Eye</i>	2
1.3.2. <i>Eye-gaze Tracking Systems</i>	5
1.4. Current Implementations and Limitations.....	5
1.4.1. <i>Non-video based methods</i>	5
1.4.2. <i>Video Based Methods - Research Systems</i>	6
1.4.3. <i>Video Based Methods - Commercial Systems</i>	11
1.5. Motivation and Objectives of Thesis	15
1.6. Organization of Thesis	16
2. System Architecture	17
2.1. Overview of System Design.....	17
2.2. Image Acquisition	18
2.2.1. <i>Eye-gaze Tracking Scene</i>	18
2.2.2. <i>Scene Lighting</i>	19
2.2.3. <i>Camera System</i>	23
2.2.4. <i>Electronics</i>	26
2.2.5. <i>Computer System</i>	27
2.2.6. <i>Physical Layout</i>	27
3. Eye and Feature Tracking.....	29
3.1. Software System Architecture.....	29
3.2. Image Processing Sequence of Operations.....	30
3.3. Rough Pupil Processing	34
3.4. Pupil Glint Detection.....	39
3.5. Fine Pupil Detection.....	41
3.6. Dual Glint Detection	46
4. Optical Geometry for Point of Gaze Estimation	49
4.1. Computing the Center of the Corneal Sphere Model	51
4.2. Computing the Center of the Pupil.....	62

4.3.	Computing the POG	69
4.4.	POG Correction.....	71
5.	System Testing and Results	75
5.1.	Accuracy Metrics.....	75
5.2.	Eye Tracking	77
5.3.	Feature Extraction	77
5.4.	Feature Extraction Sensitivity	80
5.5.	Single User Accuracy Testing	81
5.5.1.	<i>With and Without Headrest</i>	82
5.5.2.	<i>Shutter speed</i>	82
5.5.3.	<i>Sunlight Compensation</i>	83
5.5.4.	<i>IR LED Wavelength</i>	86
5.5.5.	<i>Eye-Glasses</i>	86
5.5.6.	<i>Ambient IR Levels</i>	87
5.5.7.	<i>Head Locations</i>	88
5.6.	Multi User Testing	91
6.	Conclusions	98
6.1.	Discussion and System Comparison	98
6.2.	Future Refinements	101
7.	References	103
8.	Appendix A - Infrared LED Safety	108
8.1.	System Diodes.....	108
8.2.	Maximum Permissible Exposure	108
8.3.	System Radiance	110
9.	Appendix B – Camera Search	111
10.	Appendix C – Camera Calibration.....	113
11.	Appendix D - Electronics Schematic	115
12.	Appendix E - Image Processing Parameters.....	116
13.	Appendix F – Auxiliary / World Rotation Matrixes	117
14.	Appendix G – Intersection of a Line with a Sphere.....	120
15.	Appendix H – Intersection of a Line with a Plane	122
16.	Appendix I – System Testing Figures.....	123
16.1.	Head Rest	123
16.2.	Shutter Speed.....	124
16.3.	LED Wavelength.....	126
16.4.	Eye-Glasses	127

16.5.	Ambient Lighting	128
16.6.	Free Head	129
16.7.	Multi User Trials	134

List of Tables

Table 1-1 - Research Based Eye-Gaze Tracker Comparison	11
Table 1-2 - Commercial Eye-Gaze Tracker Comparison.....	14
Table 1-3 - Limitations of the Leading Eye-Gaze Tracking Systems	15
Table 2-1 - Positions of Lights and Monitor	28
Table 4-1- Schematic Values for the Eye.....	51
Table 5-1- Results of Feature Extraction Test on Synthetic Images	79
Table 5-2- Feature Extraction Sensitivity Differences.....	80
Table 5-3 – With and Without Headrest	82
Table 5-4 – Varying Shutter Speeds.....	83
Table 5-5 - Sunlight Compensation Test Conditions	84
Table 5-6 – Different LED Wavelengths	86
Table 5-7 – Eye Glass Reflection Compensation Routine	87
Table 5-8 – Ambient IR Levels	88
Table 5-9 - miniBird Head Positions.....	89
Table 5-10 - Head Position Results 1 – 3.....	90
Table 5-11 - Head Position Results 4 – 6.....	90
Table 5-12 - Head Position Results 7 – 8.....	90
Table 5-13 - Multi User Trials 1 - 2	92
Table 5-14 - Multi User Trials 3 - 4	93
Table 5-15 - Multi User Trials 5 - 6	93
Table 5-16 - Multi User Trials 7 - 8	94
Table 5-17 - Multi User Trials 9 - 10	94
Table 5-18 - Multi User Trials 11 - 12	95
Table 5-19 – Multi User Trial 1 and 2 ANOVA Comparison Results.....	97
Table 6-1- Comparison of Leading Systems with Our Design	99
Table 8-1 – HSDL-4220 LED Specification	108
Table 9-1 - Camera Search.....	112
Table 10-1 - Camera Calibration Parameters	114

List of Figures

Figure 1-1 - Schematic of the Human Eye	3
Figure 1-2 – Spectral sensitivity of Rods and Cones	4
Figure 2-1 – Eye-Gaze Tracking System Overview	18
Figure 2-2 - Infrared Filter Spectrum	20
Figure 2-3 - Bright Pupil Image	22
Figure 2-4 - Dark Pupil Image.....	22
Figure 2-5 - Camera Sensitivity	24
Figure 2-6 - Pinhole Camera Model.....	25
Figure 2-7 - IR Lighting Waveforms.....	26
Figure 2-8 - Overall Physical System.....	28
Figure 3-1 - Image Processing Flowchart.....	30
Figure 3-2 - Bright Pupil Thresholded for Glasses	32
Figure 3-3 - Eye and Cornea ROI.....	32
Figure 3-4 - Dual Glint Detection Failure	34
Figure 3-5 - Dual Glint Recovery.....	34
Figure 3-6 - Rough Pupil Detection	35
Figure 3-7 - Gaussian Filter Applied to Raw Images	36
Figure 3-8 - Differencing Images of Eye.....	36
Figure 3-9 – Rough Pupil Image Histogram	37
Figure 3-10 – Binary Result of Thresholding for the Rough Pupil.....	38
Figure 3-11 - Pupil Glint Detection Flowchart	39
Figure 3-12 - Pupil Glint Mask	40
Figure 3-13 - Glint Threshold Extraction.....	40
Figure 3-14 - Dual Glint Threshold.....	41
Figure 3-15 - Fine Pupil Ellipse Estimation Flowchart.....	42
Figure 3-16 - Fine Pupil Extraction Masks	43
Figure 3-17 - Dilated Pupil Glint Contour	43
Figure 3-18 - Inverted Pupil Glint Contour.....	43
Figure 3-19 - Pupil Glint Mask AND Pupil Contour	44
Figure 3-20 - Masked Bright Pupil.....	44
Figure 3-21 - Fine Pupil Contour Thresholded	45
Figure 3-22 - Fine Pupil Contour ANDed with Inverted Pupil Glint.....	45
Figure 3-23 - Fine Pupil Convex Hull.....	46
Figure 3-24 - Identified Fine Pupil Contour.....	46
Figure 3-25 - Dual Glint Detection Flowchart	47
Figure 3-26 - All Off-Axis Glint Light Sources Enabled.....	47
Figure 3-27 - Identified Dual Glint Contours.....	48
Figure 4-1 - POG Estimation Algorithm Flowchart.....	49
Figure 4-2 – POG Estimation System Overview	50
Figure 4-3 - Eye Model	51
Figure 4-4 - Cornea Center Estimation World Coordinate System.....	52
Figure 4-5 - Cornea Center Calculation Flowchart	54

Figure 4-6 – Auxiliary Coordinate System for Cornea Center Estimation (as viewed along the world Z axis by a user)	55
Figure 4-7 - 2D Auxiliary Coordinate System	58
Figure 4-8 - Pupil Center Calculation Flowchart	63
Figure 4-9 - Computing the Center of the Pupil.....	65
Figure 4-10 - Pupil Sphere Intersection.....	68
Figure 4-11 - Computing the POG	70
Figure 4-12 - Calibration Example.....	72
Figure 4-13 - POG Correction Algorithm Flowchart.....	73
Figure 5-1 - Pixel Error to Visual Angle Error	76
Figure 5-2 – Image Test Set 1 for Feature Extraction Testing	78
Figure 5-3 - Image Test Set 2 for Feature Extraction Testing	79
Figure 5-4 - Narrowband Filter	84
Figure 5-5 – Sunlight Compensation Test Images	85
Figure 5-6 – 3 Planar Views of the Head Position Locations	89
Figure 5-7 - Average X and Y Errors for 8 Head Positions	91
Figure 5-8 - Average Error Comparison Between Subjects for Trial 1 and Trial 2	96
Figure 8-1 - Angular Subtense	109
Figure 10-1 – Checkerboard Images Used to Calibrate the Camera	113
Figure 10-2 – Identified Corners in a Checkerboard Image	114
Figure 11-1 - Infrared Diode Schematic.....	115
Figure 16-1 - Without Headrest.....	123
Figure 16-2 - With Headrest.....	123
Figure 16-3 – 8.33 ms Shutter.....	124
Figure 16-4 – 25.00 ms Shutter.....	124
Figure 16-5 – 58.33 ms Shutter.....	125
Figure 16-6 – 66.66 ms Shutter.....	125
Figure 16-7 – 760 nm LEDS	126
Figure 16-8 – 875 nm LEDs.....	126
Figure 16-9 – No Glasses and No Correction	127
Figure 16-10 – Glasses and Correction	127
Figure 16-11 – No Sunlight.....	128
Figure 16-12 – Cloudy Day	128
Figure 16-13 – Indirect Sunlight	129
Figure 16-14 – Position 1	129
Figure 16-15 – Position 2	130
Figure 16-16 – Position 3	130
Figure 16-17 – Position 4	131
Figure 16-18 – Position 5	131
Figure 16-19 – Position 6	132
Figure 16-20 – Position 7	132
Figure 16-21 – Position 8	133
Figure 16-22 – Subject 1 – Trial 1 and 2.....	134
Figure 16-23 – Subject 2 – Trial 1 and 2.....	135
Figure 16-24 – Subject 3 – Trial 1 and 2.....	136

Figure 16-25 – Subject 4 – Trial 1 and 2.....	137
Figure 16-26 – Subject 5 – Trial 1 and 2.....	138
Figure 16-27 – Subject 6 – Trial 1 and 2.....	139
Figure 16-28 – Subject 7 – Trial 1 and 2.....	140
Figure 16-29 – Subject 8 – Trial 1 and 2.....	141
Figure 16-30 – Subject 9 – Trial 1 and 2.....	142
Figure 16-31 – Subject 10 – Trial 1 and 2.....	143
Figure 16-32 – Subject 11 – Trial 1 and 2.....	144
Figure 16-33 – Subject 12 – Trial 1 and 2.....	145

Acknowledgments

I would foremost like to thank Dr. Peter Lawrence for his supervision over the course of this thesis. The discussions we had on potential solutions to the many issues encountered provided a great deal of inspiration for this work. The freedom to pursue the many possible solutions and discover which would work and which would not was also greatly appreciated. The feedback provided on the thesis document proved invaluable in improving the written work. I would also like to gratefully acknowledge his financial support for both my time as a researcher in the Robotics and Control Laboratory and also for the materials required in the development of the thesis.

I would like to thank Borna Nouredin whose work on the Gaze Tracking Device laid the groundwork for the eye-gaze tracking system developed here. I would also like to thank the various support staff in the Electrical and Computer Engineering Department for their assistance.

Finally I would like to thank my family and close friends for their support and friendship over the years spent working on this thesis. In particular I would like to thank Julie for making the final year all the more memorable and for her understanding and patience as I completed my work.

1. Background

1.1. Introduction

Eye-gaze tracking has the potential to greatly influence the way we interact with machines as a new form of human machine interface. The point of gaze of a user is closely related to their intention. By tracking the eye-gaze of a user, machines may gain valuable insight into what the user is thinking of doing, resulting in more intuitive interfaces and the ability to react to the users intentions rather than explicit commands.

Traditional methods for tracking eye-gaze involve either direct contact with the subject, or time consuming post-processing of recorded video [1]. Recent advances in electronics and computing technology have made possible non-contact and real-time video based eye-gaze tracking systems. These systems are replacing the traditional methods used for eye-gaze tracking in many applications due to their increased ease of use, reliability, accuracy and comfort for the subject. While the recent advances in eye-gaze tracking technologies have improved their effectiveness, a number of issues remain which may be limiting their widespread use [1].

The following sections of this chapter will introduce some of the many applications for eye-gaze tracking and an outline of basic concepts useful for better understanding the issues in eye-gaze tracking. A review of the current leading eye-gaze tracking systems and an outline of their limitations are identified. The objectives for this thesis are described and finally an overview of the remainder of the thesis document is provided.

1.2. Eye Gaze Tracking Applications

Eye-gaze information has proven useful in a diverse number of applications from physiological and psychological studies [2], to usability studies in driving [3] and aviation [4] and analysis of layout effectiveness in advertising [5]. These applications do not typically require real-time performance as the information of interest may be extracted by post-processing recorded data.

Applications using eye-gaze information as a control tool include target sighting for the military [6], robotic control [7] [8], gaze contingent displays [9], computer mouse augmentation and control [10] and eye typing [11] for the physically disabled. These applications require more advanced eye-gaze tracking devices that are capable of operating in real-time.

One of the most common environments for developing eye-gaze tracking devices is in estimating the eye-gaze position of a subject sitting in front of a computer and looking at the computer screen [12]. System accuracy is easily determined by asking the subject to look at a grid of points on the screen and record the computed eye-gaze position at each. The difference between the reference point and the estimated eye-gaze position, averaged over the entire screen, provides a measure of system accuracy.

A number of assumptions are made in the development of an eye-gaze tracking device for an environment such as this. Eye-gaze estimations are valid only when the head is held relatively still. While the head is in motion, the recorded data is often corrupted by motion blur. Use of the eye-gaze tracking device should not detrimentally affect how the user operates the computer nor restrict how the user sits in front of the screen. The allowable range of motion of the eye-gaze tracking device should be at least as great as the allowable range of motion of an average computer user. It is also assumed that the eye-gaze tracking device will be robust to ambient lighting conditions typically found in a computing environment.

1.3. Key Concepts in Eye Gaze Tracking

1.3.1. Anatomy and Physiology of the Human Eye

A cross section of the human eye is shown below in Figure 1-1. The figure shown is a simplified version of the eye illustrating the elements of greatest significance for eye-gaze tracking.

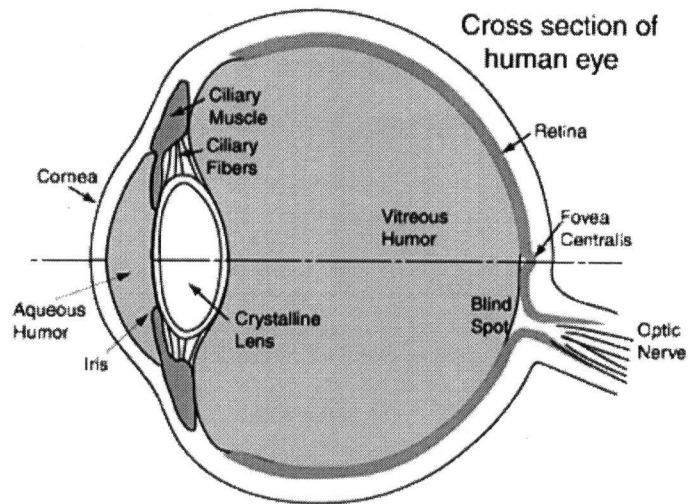


Figure 1-1 - Schematic of the Human Eye¹

The **cornea** is the outermost surface of the eye and is filled with a transparent fluid called the **aqueous humor**. The **pupil** is the circular area through which light enters the interior of the eye and is surrounded by the **iris**. The pupil appears to be black as most of the light entering the eye is absorbed by the **retina**, or posterior surface of the eye. The **lens** is an optical surface which focuses the incoming light onto a small patch of the retina called the **fovea centralis**. The fovea is the portion of the retina responsible for our sharpest vision, the remaining surface of the retina is used for peripheral vision. The interior of the eye is filled with a transparent fluid called the **vitreous humor**. The surface of the retina is covered with **rods** and **cones** which are the sensors used for converting light into electrical impulses which are then transmitted to the brain. The spectral sensitivity of the cones (for colour) and rods (for intensity) are shown in Figure 1-2.

¹ Figure adapted from the following website:
<http://hyperphysics.phy-astr.gsu.edu/hbase/vision/eye.html>

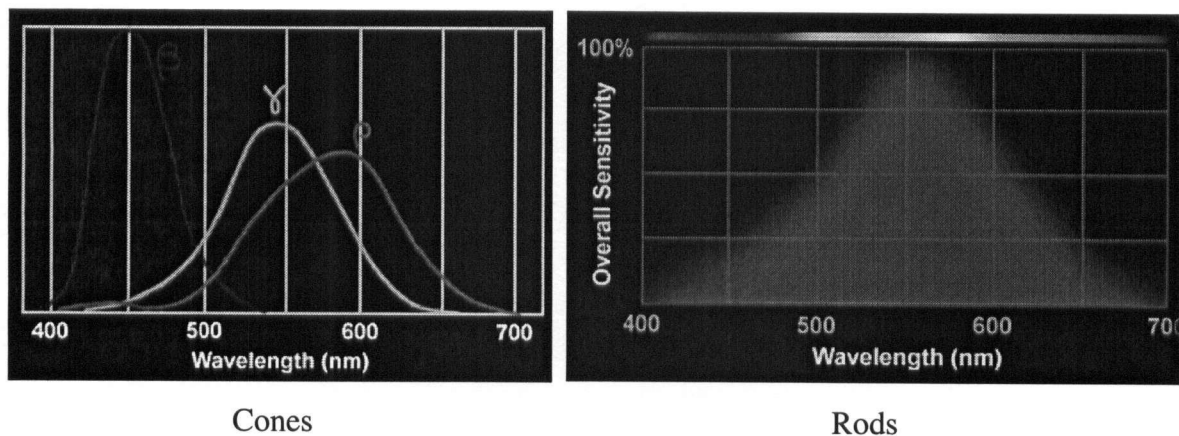


Figure 1-2 – Spectral sensitivity of Rods and Cones¹

The reflections of light off the outer surface of the cornea are known as the first **Purkinje** images, named after the Czech physiologist Jan Purkinje who first described them [13]. These reflections are more commonly known as **glints**. The second, third and fourth Purkinje images are the reflections of light off interior interfaces in the eye and are not nearly as bright as the first Purkinje images.

The retina exhibits a particular optical phenomenon called **retro-reflectivity**. Retro-reflectivity is the property by which light reflects off a surface directly back along the incoming path. This is commonly seen as the red-eye effect in flash photography. A review of the variations in retro-reflective response of the retina over a number of factors such as age, gender, and ethnicity can be found in Nguyen et al [14].

There are three main types of motion involved in reorienting the eye within the head, **saccades**, **smooth pursuits** and **fixations** [16]. A saccade is a very rapid change in visual angle, thought to be ballistic in nature, in which the desired visual angle is pre-determined with the subsequent eye motion occurring without correction until completed. Visual perception is suppressed during a saccade which lasts 10 to 100 ms and can re-orient the eye at speeds up to 800 °/s [15]. Smooth pursuit is the mechanism by which the eye tracks a moving object within the visual field, or tracks a stationary object when the body is in motion. Fixations occur when the eye stabilizes the image of the object in view on the retina through micro-saccades, drift and tremors. These small motions are required

¹ Figure adapted from the following website:
<http://www.photo.net/photo/edscott/vis00010.htm>

to refresh the optical sensors (the rods and cones in the eye), as a stationary image with respect to the retina will simply fade away. These small changes in eye orientation during a fixation occur at random and can result in fixation fluctuations of up to 5° of visual angle [16].

Finally, motion of the head occurs both to stabilize and to re-orient the eye on an object of interest. Rotation of the head may occur at speeds of up to 100 °/s [15].

1.3.2. Eye-gaze Tracking Systems

The **line-of-sight** (LOS) is defined as the vector from a point located at the center of the cornea extending through the center of the pupil and out of the eye into the external world. The location of the fovea shown in Figure 1-1 differs from person to person and can be found up to 5 degrees from the LOS [13]. This difference may result in an offset in the estimated **point-of-gaze** (POG), or point in space at which the LOS intersects with the object the eye is looking at.

Accuracy of eye-gaze tracking devices is typically reported in terms of visual angle error, which is the angle in LOS between the actual POG and estimated POG. The error measured in the computer application outlined above is recorded in units of screen pixels which can be converted to degrees of visual angle with knowledge of the size of the screen, the resolution of the screen and the average distance from the screen to the eye. Visual angle is the preferred accuracy metric as it allows for easy comparisons between systems, avoiding the need to report the details of the screen and distance to eye.

1.4. Current Implementations and Limitations

1.4.1. Non-video based methods

Two of the most common non-video based methods for estimating the LOS are **electro-oculography** (EOG) and the **scleral search coil**. The EOG and search coil methods measure the rotation of the eye within the head.

EOG systems measure small DC potentials on the skin using electrodes attached around the eye [16]. EOG is a reasonably accurate technique typically from 1-2° of visual angle [17], however, in addition to an intrusive mounting system it also suffers from

significant drift and is susceptible to interference from electromyography (EMG) and electroencephalograph (EEG) signals [18].

The scleral search coil method uses a coil mounted in a contact lens which is applied to the eye of a subject [16]. Measurements are taken as the eye moves the coil through an externally applied magnetic field and the position of the lens and consequently of the eye is determined. The search coil method has reported accuracies of 0.017° of visual angle [17], however it is also quite invasive and requires a number of difficult calibrations [18].

Both of these non-video based methods are fairly intrusive, requiring direct contact with the users face or eye. Far less intrusive video based systems have replaced these devices in many applications.

1.4.2. Video Based Methods - Research Systems

Video based eye-gaze tracking devices use images of the eye and head to track the subjects eye and determine the LOS and POG. One of the most common methods for estimating the POG in video based methods is the **pupil-glnt** (PG) vector method [19]. The PG vector is the vector formed between the center of the image of the pupil and the center of the image of a glint reflected off the surface of the cornea. When the eye rotates within the head the position of the pupil changes, while the position of the glint remains stationary as the glint is a reflection off the surface of a spherical object. The change in PG vector is roughly proportional to the change in the POG of the subject when the eye re-orientes.

The accuracy of the PG method is quite good, down to 0.5° of visual angle [20] [21] [22]. The major limitation of this method is that when the eye translates from the calibrated position, the accuracy of the POG estimation rapidly begins to degrade [23]. To compensate for this, the user is asked to keep their head very still, or to use a head rest or bite bar. An alternative solution is to mount the entire system on the head of the user which then would remain fixed relative to the head regardless of the motion of the head [22], provided no slippage occurs. Head mounted systems are often bulky and heavy and can cause discomfort to the users who wear them, making them impractical for extended use [24]. The requirement for a stationary, fixed head or a head mounted system for the

PG vector method to work well has led to the development of alternative but more complex free head methods for estimating the POG.

Zhu and Yang [25] used a **wide angle** (WA) lens to capture a large **field of view** (FOV) allowing for a large range of head positions. They have developed an iris tracking method using subpixel ellipse fitting to compensate for the low image resolution of 320x240 pixels, but a large FOV. POG estimation is similar to the PG vector method with the center of the iris used instead of the center of the pupil and the corners of the eye replacing the stationary pupil glint. They reported an average accuracy of 1.4° over 3 sequences but do not mention what was changed between sequences. The similarity to the PG vector method, which cannot handle free head motion, indicates that this system may have trouble maintaining the accuracy reported when the head changes position.

Wang, Sung, and Venkateswarlu [26] used a single camera with a **narrow angle** (NA) lens to record high resolution images of the eye. They fit an ellipse to the image of the iris contour which was then back projected into 3D space. Using the normal to the back projected ellipse and an estimate for the center of the eye they were able to estimate the POG on a screen. Their iris fitting algorithm compensates for occluded portions of the iris due to the upper and lower eyelids. They reported an accuracy error of less than 1° for 10 subjects, including males and females, children and the use of eye-glasses. To allow for free head motion the authors noted that they will need to add a WA camera system to direct the high resolution NA camera. They do not mention if their system operates in real time.

Morimoto, Amir, and Flickner [27] developed a free head motion algorithm for estimating the POG with a single camera. Their method is based on the theory of spherical mirrors and a model of the eye in which the surface of the cornea is assumed to be spherical. They determine the center of the cornea using multiple glints and triangulation and determine the center of the pupil using ray tracing and refractive geometry. In [27] they claimed their system is calibration free, achieved by using an eye model with population norms as parameter values. They do not compensate for the possibility that the fovea does not lie on the LOS. They reported an accuracy of 2.5° , but have only tested on simulated data.

Yoo and Chung [28] have developed a novel method for estimating eye gaze using a cross-ratio metric. The image locations formed from point light sources located at the four corners of the monitor are projected back to the vanishing point to determine the center of the cornea. With their method they are able to estimate the POG without camera calibration, though personal calibration is still required. A WA camera is used to direct a NA camera for high resolution images of the eye. The cameras have image resolutions of 640x480 pixels each. The system runs on a Pentium IV 1.4 GHz computer at a speed of 15 Hz. The authors report an accuracy of 0.98° in horizontal error and 0.82° in vertical error over a range of head positions. One major problem envisioned with their system was encountered in the development of our own system. Placing point light sources at the four corners of the monitor can cause some of the glint reflections to fall off the surface of the cornea and onto the sclera causing the reflection to distort. If only 2 or 3 reflections are required this would not be a problem but their method requires all four reflections to be located on the cornea at the same time.

Noureddin, Lawrence, and Man [29] developed a system utilizing a WA lens for tracking the location of the eye and a NA lens for tracking the POG. The WA camera system determines the location of the eye and controls the orientation of the NA camera. The NA camera is oriented using a motorized mirror and base which can achieve faster slew rates than the pan-tilt mechanisms typically used for NA camera orientation. The feature parameters extracted from the NA images are used with a method of functional approximation to estimate the POG. Their system was capable of tracking eye motion up to $100^\circ/\text{s}$ and to compute the POG with an accuracy of 2.9° at a rate of 9 Hz.

Some of the more promising advances in free head POG estimation routines have been in developing 3D models of the camera, eye and scene to directly calculate the position of the eye, the position of the center of the pupil, the LOS and ultimately the POG. These methods allow for the eye to be located at any point in space while still accurately estimating the POG, though at the cost of more complex systems.

The system by Beymer and Flickner [30] uses four analog NTSC cameras of which two have NA lenses and two have WA lenses. The system is run on a dual Pentium 933 MHz computer and achieves an update rate of 10 Hz. The stereo WA cameras are

used to record images in which the position of the eyes are tracked using eigenfeatures. The knowledge of the position of the eyes is used to direct the orientation of the dual NA cameras. The NA cameras are used for taking high resolution images of the eye from which the POG will be estimated. To determine the POG depth from focus, stereo constraints and a non-linear estimation technique are used. Depth from focus provides a depth estimate by varying the focal length until the image with the highest frequency image components is found. The focal length is then used to estimate the distance from the eye to the camera.

The NA camera view is oriented using mirrors mounted on high performance galvanometer motors which can rotate and settle in under 2 ms. Galvo motors are also used to control the position of the camera lenses, allowing a variable focal distance. The authors report a horizontal FOV of 1 meter and have a large DOF due to their computer controlled focal length. An accuracy of 0.6° is reported for a one sample trial on the author D. Beymer in which he looked at a grid with 22 points.

The system is quite complex and uses a large amount of hardware, including 2 pairs of stereo cameras, rotating mirrors mounted on galvo motors and galvo controlled focusing. The system requires considerable calibration, especially considering the camera calibration must take into account the varying mirror positions and focal lengths. The test sample set is unfortunately too small to draw a conclusion as to the accuracy of the system considering head motion and differing test subjects.

Shih and Liu [31] developed a free head system that uses 3D geometric modeling. They use a direct triangulation method utilizing dual glints and stereo constraints to determine the center of the eye and use ray tracing to determine the center of the pupil. Their system uses two glint light sources and two analog NTSC stereo cameras located 70 mm from the eyes. They attempt to estimate almost all the required parameters of the eye to avoid per user calibration, however, per user calibration is still required to estimate the offset in visual angle due to the variation in fovea offset from the optical axis.

Their system runs at 30 Hz on a Pentium III 500 MHz computer with 128 megabytes of RAM and uses only 3% of the CPU processing time. The cameras have a resolution of 640x240 with a FOV of 4x4 cm and a narrow DOF. They report an accuracy

of better than 1° on a trial of 6 subjects looking at 4x4 grids on the computer screen. None of the subjects tested were wearing eye-glasses. The small FOV and narrow DOF make it difficult to evaluate the 'free head' component of this design. The system is fairly restrictive because the cameras are located so close to the subject's eye, unfortunately a requirement due to the low camera resolution.

Ohno and Mukawa [32] have developed a free head camera system which uses two stereo analog NTSC cameras and a NA analog NTSC camera mounted on a pan-tilt mechanism. This system extends their previous work [33] on a single camera based system which consisted of just a NA camera. For their previous system the camera direction was fixed and the head was restricted to a 4x4 cm area at a distance of 60 cm from the screen. One possible reason for the move to a 3 camera system was to allow for a greater range of head positions. The WA stereo cameras of their most recent system direct the pan-tilt mechanism to orient the NA camera on the eye. The NA camera uses the depth from focus method to determine the distance to the eye. Ray tracing is used to estimate the center of the pupil and ultimately the POG.

The FOV is 10x10 cm with a large DOF due to the controlled focus. The stereo camera resolutions are 320x120 pixels and the NA camera is 640x480 pixels. To process the stereo camera images a dual AMD Athlon 1.7 GHz computer is used and to process the NA images a dual Pentium III 1 GHz computer is used. The overall system update rate is 30 Hz.

On testing with 9 subjects, 2 of whom wore glasses, the average system accuracy just after calibration was 0.68° and after each subject got up and then reseated themselves in front of the device had degraded to 1° . The author mentions that the depth from focus method is not very sensitive to small displacements of the head which can cause errors in the POG estimation.

Shown in Table 1-1 is a comparison of the important features found in the leading research based eye-gaze tracking devices. Not all information was available for all systems.

Table 1-1 - Research Based Eye-Gaze Tracker Comparison

	Beymer and Flickner	Shih and Liu	Ohno
Accuracy (average error)	0.6°	0.5° - 1°	0.68-1°
Cost	Possibly high, large amounts of hardware	Possibly moderate, only 2 cameras	Possibly high, large amounts of hardware
Camera type	Analog with Frame Grabber	Analog with Frame Grabber	Analog with Frame Grabber
Update rate	10 Hz	30 Hz	30 Hz
Tracking recovery time	Not specified	Not specified	Not specified
Personal Calibration	Not clearly defined	1 point calibration	2 point calibration
Non contact	Yes	Yes	Yes
Non restrictive (XxYxZ)	100x100 cm FOV with a large DOF	4x4 cm FOV with a narrow DOF	10x10 cm with a large DOF at 60 cm
Ability to handle ambient light	Not mentioned	Long pass filter used	Not mentioned
Processor Overhead	High	Low	High
POG estimation method	Multiple stereo cameras, depth from focus, non-linear estimation	Stereo cameras, ray tracing using dual glints	Stereo cameras, depth from focus, ray tracing
Number of subjects tested	1 subject tested	6 subjects tested	9 subjects tested
Tested with numerous head locations	No	No	2 positions
Works in presence of eye-glasses & contacts	Not mentioned	Not tested with glasses	Yes

1.4.3. Video Based Methods - Commercial Systems

Most commercial eye-gaze tracking systems have used the Pupil Glint (PG) vector method for determining the POG. These commercial systems often claim that their system allows for free head motion by which they mean the head can be located anywhere within the FOV, but once calibrated the head must remain in the current location. There are two commercial systems which stand out as having achieved the goal of true free head

motion, the Tobii 1750 system by Tobii Technologies [34] and faceLab 4.1 by Seeing Machines [35]. These two companies make a number of claims in their brochures but provide few details on their implementation. The eye-gaze tracking system designed by Matsumoto and Zelinsky [36] is the basis for the faceLab system. Their published papers can be used to provide some insight into the faceLab system design. A patent filed for the Tobii system [37] provides information into the methods used for their design.

Matsumoto and Zelinsky [36] developed a free head eye-gaze tracking system based on determining head pose and the orientation of the eye within the head. Head pose is determined by correlating template images of facial features such as eye and mouth corners, then using stereo matching to fit the identified feature locations to a model of the face. They use the Hough transform to identify the circular iris located between each of the identified eye corners. The gaze vector is determined from the head pose and their eye model. Their system uses stereo analog cameras with a wide FOV allowing for large head motions, which results in small images of the eye and consequently low resolution images of the eye. They report eye-gaze accuracy of 3° with their system operating at 15 Hz using a dedicated image processing board to perform the intensive correlation operations.

The current faceLab system has two operational modes, classic mode and high precision mode. The classic mode is very similar to the system designed by Matsumoto and Zelinsky where the head pose is tracked within a large FOV with fairly low accuracy. The precision eye-gaze tracking mode appears much more similar in operation to the Tobii system, with a more restrictive FOV and higher resolution images of the eye. A specification is provided for both modes, though the details are interspersed and it is difficult to determine which claims are for the low accuracy classic mode or the higher accuracy precision mode. The precision mode of the faceLab system claims an accuracy of under 1° with a FOV of 25x15x30 cm at 65 cm. It is recommended that the precision mode be used in an indoor environment to operate well.

The Tobii system uses a single digital Firewire camera for image capture and does not have any moving parts in the design. When the eye tracking system loses the eye it can be reacquired in less than 100 ms when the eye returns to the FOV. They claim the system tracks virtually all subjects, including those who use glasses and contact lenses.

The system operates in all lighting conditions except direct sunlight. The minimum system requirements are a P4 2.4 GHz computer with 256 megabytes of RAM.

The FOV is 20x15x20 cm at a distance of 60 cm from the screen. Both eyes can be tracked which effectively increases the FOV to 30x15x20 cm. Over head motion the average error is under 1° of visual angle. On a test of 10 subjects looking at 8x8 grids an average accuracy of 0.5° is reported.

The patent filed by Elvesjoe, Skogoe, and Gunnar [37] provides some details on the inner workings of their system. The distance to the eye is determined by shining a pattern into the eye in which the relative dimensions of the points in the pattern indicate the distance of the eye to the sensor. They then use the PG vector method to estimate the direction of gaze. The distance measurement is used to compensate for the change in PG vector due to head motion. To handle the large high-resolution images they first decimate the data when detecting the position of the eye in full sized images. Once the eye has been detected a **region-of-interest (ROI)** is used to reduce the size of the image to a manageable size, only processing the portion of the image containing the eye. To aid in the detection of the eye, the bright pupil / dark pupil method is used which will be described in Section 2.2.2.3.

A comparison of the leading commercial eye-gaze trackers is shown in Table 1-2. Not all information was available for all systems.

Table 1-2 - Commercial Eye-Gaze Tracker Comparison

	faceLab 4.1	Tobii 1750
Accuracy (average error)	1°	0.5° - 1°
Cost	\$30,000-\$40,000 USD	\$25,000 USD
Camera type	Digital Firewire or Analog	Digital Firewire
Update rate	60 Hz	50 Hz
Tracking recovery time	< 200 ms	< 100 ms
Personal Calibration	Calibration required but not described	Yes, 2 or more points
Non contact	Yes	Yes
Non restrictive (XxYxZ)	25x15x30 cm at 65 cm	20x15x20 cm at 60 cm or 30x15x20 with both eyes
Ability to handle ambient light	Recommended for indoor environments	Works with all but direct sun
Processor Overhead	Not specified	Low
POG estimation method	Not specified	PG vector, multiple glints and ray tracing
Number of subjects tested	Not specified	10 subjects tested
Tested with numerous head locations	Not specified	Yes
Works in presence of eye-glasses & contacts	Yes, but possibly only classic mode?	Yes

A summary of the limitations we feel remain to be improved in these leading systems are outlined below in Table 1-3.

Table 1-3 - Limitations of the Leading Eye-Gaze Tracking Systems

Cost	Cost is a major factor, the commercial systems specified above are from \$25,000 to \$40,000 USD. Many of the research systems use multiple cameras, computers and numerous hardware elements.
Camera	The analog cameras used by the research systems have limited resolution and because of the frame grabber required for digitizing analog signals, are more susceptible to noise.
FOV	The FOV is fairly large for the systems by Ohno, Beymer, faceLab and Tobii. Ohno and Beymer achieve the large FOV using mechanical orientation hardware. The moving parts decrease reliability and increase the time to reacquire the eye when the NA cameras lose their lock on the eye.
Ambient Lighting	The Tobii system claims to work in all but direct sunlight. The faceLab precision mode system recommends an indoor environment. The other systems do not mention or test this system aspect.
Processor Overhead	The systems by Beymer and Ohno require large amounts of processing power (multiple computers).
Testing	The testing methods reported for accuracy over multiple head locations were not explicitly specified for any of the systems.
Eye-Glasses	The systems by Ohno and Tobii both claim to handle eye-glasses but do not specify how they compensate for reflections off the lenses. It is possible they require the user to reorient their head in such a way as to cause the reflections of the lenses to not appear, which violates the goal of a non-intrusive system.

1.5. Motivation and Objectives of Thesis

An improved eye-gaze tracker may help to increase the acceptance of eye-gaze as a form of human machine interface. The objective of this thesis is to develop and evaluate a single camera, 3D model based, eye-gaze tracking system that improves upon existing designs. The 3D modeling method removes the requirement for fixing the user's head and allows for free head motion. Using a single camera reduces the complexity of the

associated system hardware and quite possibly reduces the computational load when compared with multi-camera systems.

Multiple light sources and the dual glint method were employed in the estimation of the POG. We also design the system in such a way as to permit the use of corrective lenses such as contact lenses and eye-glasses. As well, infrared lighting and filters were utilized to reduce the sensitivity of the system to ambient lighting conditions. Finally we strove to achieve the simplest system possible with a single camera, no moving parts, and minimal external hardware and custom circuitry. Such a system may reduce the difficulties encounter by users of eye-gaze tracking systems and result in a device that is more appealing to a wider audience.

1.6. *Organization of Thesis*

Chapter 2 provides an overview of the eye-gaze tracking system as well as details on the design and implementation of the system architecture. The imaged scene, scene lighting, camera system, electronics, computer and physical layout are all described in Chapter 2. Chapter 3 details the software system architecture and the algorithms used to extract the pupil and glint feature information from the images. Covered in chapter 4 is a detailed description of the POG estimation algorithm as well as the implementation methods for determining the center of the cornea, the center of the pupil, the POG, and the corrected POG. Chapter 5 outlines the system tests performed to evaluate the system as well as the results of the testing. System testing covered eye tracking, feature tracking, single user and multi user testing. Single user testing tested the system a number of different system parameters with a consistent user. Multi user testing tested the system accuracy over a number of users. In chapter 6 the conclusions of our work are presented, providing a comparison between our system and the leading research and commercial based systems. Our results and areas of possible improvement are also discussed.

2. System Architecture

2.1. *Overview of System Design*

There are three main systems to the eye-gaze tracking system; image acquisition, eye and feature tracking and LOS and POG estimation. Image acquisition involves recording images of the user in which the features of interest can be readily extracted. The recorded images are passed on to the eye and feature tracking system which tracks the location of the eye and precisely determines the location of features required for estimating the LOS and POG. The extracted data is then used in the final step to estimate the LOS and POG. The elements of each of the three main tasks are shown in Figure 2-1 and are fully described in the following sections.

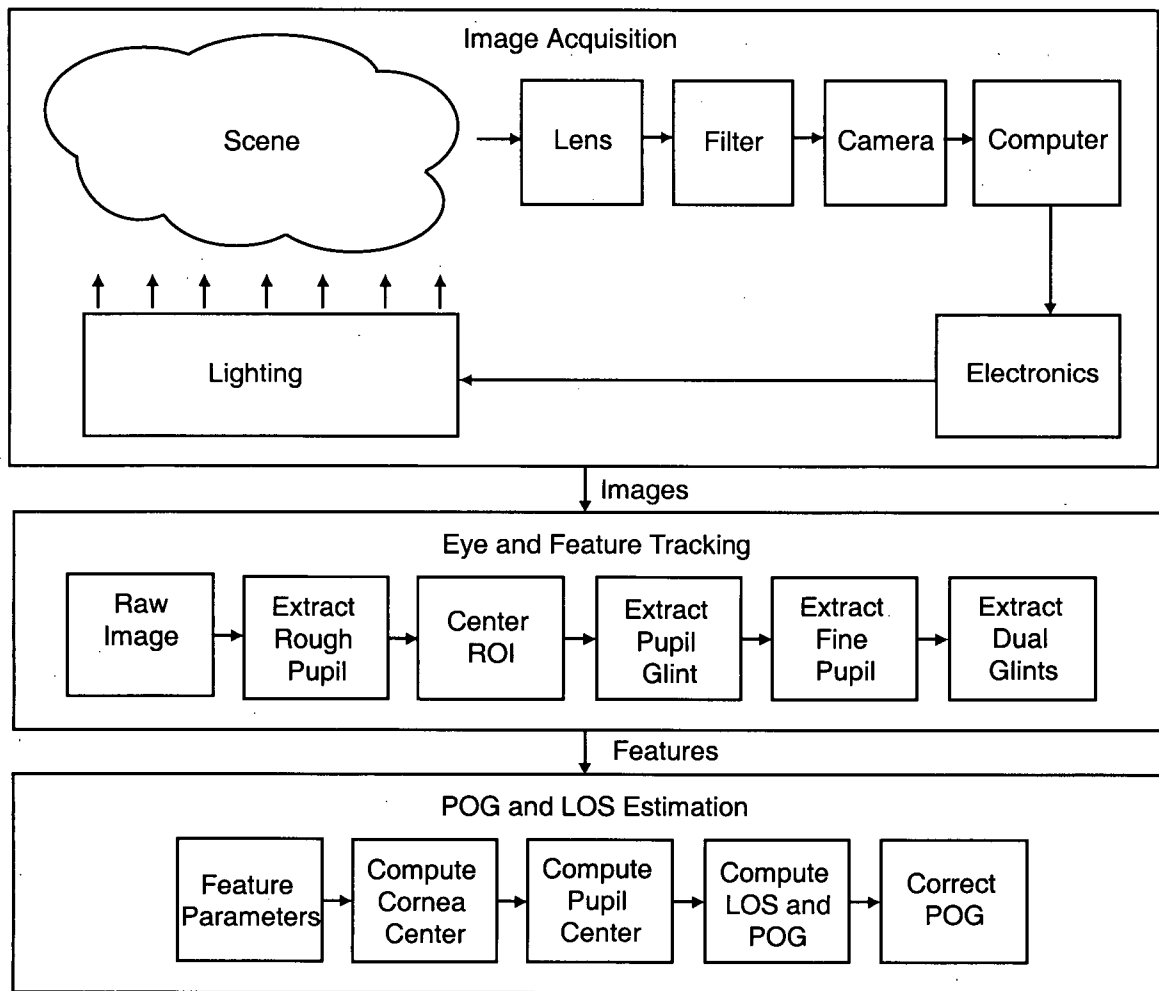


Figure 2-1 – Eye-Gaze Tracking System Overview

2.2. Image Acquisition

The image acquisition system is responsible for delivering images of the scene to the eye and feature tracking system. The features of interest such as the pupil and glints must be clearly discernable in the images in order for the eye and feature tracking system to operate properly. The image acquisition system is comprised of the scene, the lens, optical filter and camera for capturing the images, the computer running the system software, and the electronics and lighting system for illuminating the scene.

2.2.1. Eye-gaze Tracking Scene

The imaging scene consists of the head of the user looking towards the monitor. The user should be able to move their head to any natural position in front of the screen

while still remaining in view of the camera. The system will operate correctly as long as the user is looking somewhere on the screen with one of their eyes within the FOV of the camera. At the closest distance a user would typically sit in front of the computer screen (60 cm), the field of view is 13 cm in the horizontal direction and 10 cm in the vertical direction with a depth of field of 20 cm. The allowable range of head positions within this FOV is acceptable for most seated situations, though excessive slouching or leaning closely towards the screen will move the eye outside the FOV.

2.2.2. Scene Lighting

The lighting system for illuminating the scene is described in this section. The optical filter used will be discussed and the filter spectra shown. The locations of the light sources which illuminate the scene will also be described. The safety of the system is discussed while the safety factor calculations are found in Appendix A.

2.2.2.1. Infrared LED Lighting

The scene is illuminated using Infrared Light Emitting Diodes (IR LEDs) and imaged with an IR sensitive camera. IR lighting is used to reduce the sensitivity of the system to ambient lighting conditions as fluorescent lighting contains very low levels of light in the infrared wavelengths. The LEDs used were Agilent HDSL-4220 diodes which produce light at a wavelength of 875 nm. These diodes were chosen as they are inexpensive and readily available.

IR light is not visible to humans and therefore should not distract the user while the system is operating. Under normal operation the IR lights turn on and off rapidly, producing a flashing effect that could be distracting if it were visible. To produce enough light to illuminate the scene, sets of 4 LEDs are packed closely together to approximate point light sources.

2.2.2.2. Optical Filter

An optical filter that only transmits long wavelength light is used to filter out visible light and pass to the camera sensor only the light that was generated by the system. The long (low)-pass filter used was the Ilford SFX-A. The exact filter spectrum is not

provided by the manufacturer but was specified to be similar to the Wratten 89B high pass filter shown in Figure 2-2 [38].

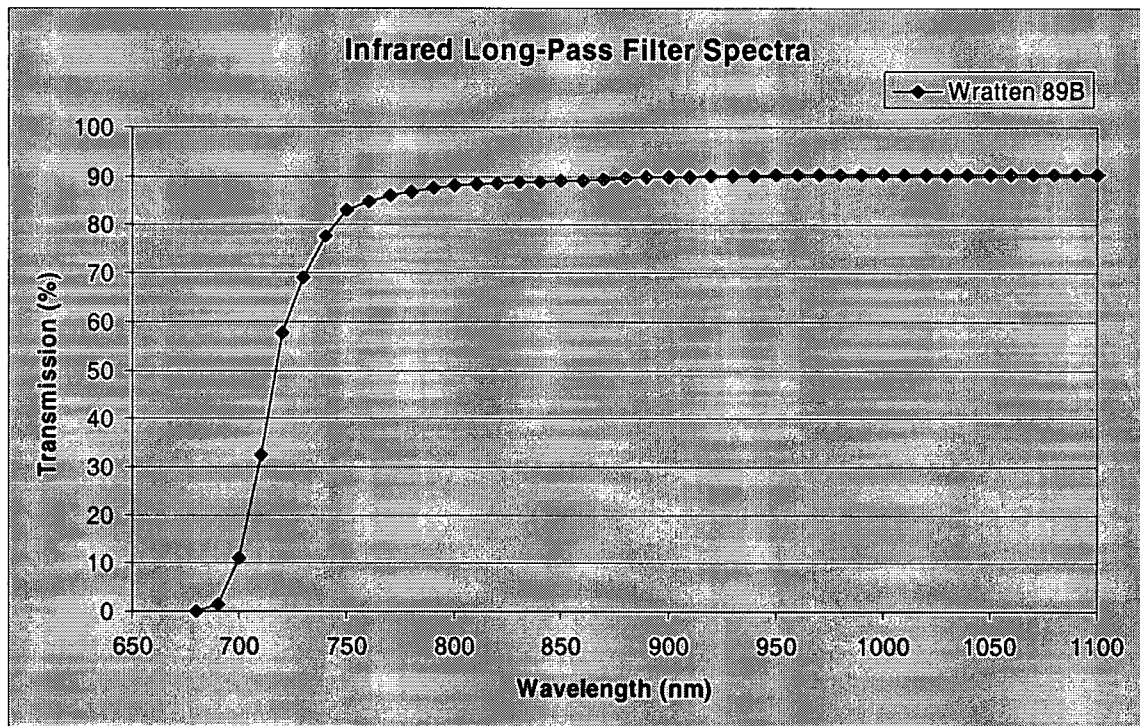


Figure 2-2 - Infrared Filter Spectrum¹

2.2.2.3. Light Source Placement

The placement of the system lighting is important for generating images in which the features of interest are easily detected. Tracking the eye requires an image in which the pupil contour is clearly defined. For individuals with dark irises, segmenting the black pupil can be difficult. The image differencing technique first described by Y. Ebisawa [39] is used to increase the contrast of the pupil. The image difference technique involves recording two images, one with the pupil brightly illuminated and one in which the pupil is dark. Provided the two images are illuminated to a similar average intensity, subtracting the dark pupil image from the bright pupil image results in an image with the pupil clearly highlighted.

¹ Figure adapted from data from the following website:
<http://www.a1.nl/phomepag/markerink/irfilter.htm>

To generate the bright pupil, a ring of lights are placed as close as possible to the optical axis of the camera. Due to the retro-reflectivity property of the retina, light that enters the eye and strikes the retina will reflect directly back out towards the source of the light. The on-axis lighting results in much of the transmitted light being reflected off the retina and back into the camera lens generating a bright pupil.

Point source lights are located around the monitor to generate the dark pupil image. These off-axis lights illuminate the face to roughly the equivalent average intensity level as the on-axis lighting but because they are not oriented along the optical axis of the camera the lights do not cause the pupil to illuminate.

The dual glint method which is used by the POG estimation algorithms requires an image containing two glint reflections off the surface of the cornea. By using two light sources for the off-axis lighting the dark pupil image will also contain the dual glint reflections. There are actually six off-axis light sources located around the monitor of which only two are used at one time. If any one of the lights causes reflections off the surface of eye-glasses to interfere with the eye tracking and feature extraction systems, the offending light may be turned off and an alternate light turned on.

An example of the bright pupil and dark pupil images are shown in Figure 2-3 and Figure 2-4 respectively. Note the contrast difference between the pupil in the bright and dark pupil images and the dual glint reflections in the dark pupil image.



Figure 2-3 - Bright Pupil Image



Figure 2-4 - Dark Pupil Image

2.2.2.4. Lighting System Safety

Infrared diodes are used in many household remote controls and are considered safe for human use [40] [41]. Infrared LED lighting should not be confused with infrared LASER lighting which is highly collimated and can be concentrated by the lens in the eye

to a very small spot on the retina. This concentrated beam can have a very high energy density which could cause injury to the eye. The infrared diodes used in this EGT are not collimated and have an angle of divergence of 30°.

The human eye does not have an aversion response to high levels of infrared lighting as it does to high levels of visible light; therefore the level of infrared light used must be kept below a certain safety threshold. The safety guidelines outlined in the International Electrotechnical Commission (IEC) Technical Report 60825-9 were followed [42]. Appendix A shows the calculations used to determine the maximum permissible exposure and calculate the exposure generated by our system.

2.2.3. Camera System

The camera used for this project is a monochrome digital camera from Point Grey Research. The camera model is the Dragonfly, which uses a Sony ICX204AL CCD sensor which has no built-in infrared filter. The sensor has a resolution of 1024x768 pixels and is capable of imaging at 15 **frames per second** (fps). The Firewire protocol is used to transmit images to the PC at data rates of up to 400 megabits per second. The camera has 4 **general-purpose input output** (GPIO) pins, which are used for external connectivity with the lighting system. The Dragonfly was chosen for its moderate resolution, compact size, low cost and availability. A search for an improved camera was performed, as outlined in Appendix B. Unfortunately no acceptable replacement camera was found that would meet our desired specifications.

A sensitivity plot for the sensor is shown below in Figure 2-5. Recall that our lighting system is located at 875 nm and that the human visual system has very low and decreasing sensitivity at wavelengths longer than 700 nm (see Figure 1-2).

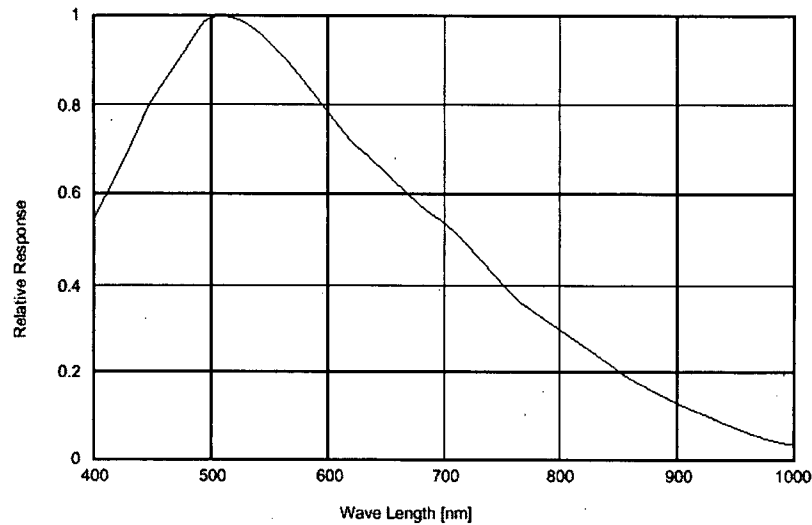


Figure 2-5 - Camera Sensitivity¹

The digital camera solution provides a number of advantages over the analog camera systems. A simpler, more compact, and lower noise design is achieved by avoiding the image frame grabber. In addition, improvements in resolution and frame rates are expected as digital imaging technology improves, while the NTSC standard for analog cameras will not.

2.2.3.1. Camera Lens

A Pelco 13VA5-50 variable focus lens with an adjustable focal depth (5-50 mm) was used for the camera. The flexible focal length allowed for testing different fields of view during system development. A focal length of 32 mm was chosen to provide the largest FOV while still maintaining a high resolution image of the eye.

2.2.3.2. Camera Model and Calibration

To develop a mathematical model of the system, the pinhole model for the camera and lens is used as shown in Figure 2-6. The intrinsic parameters for the pin-hole camera model includes the effective focal length, the critical point (or center of the image plane), skew, and the distortion parameters. Distortion was not compensated for as the images recorded do not exhibit significant curvature. The skew factor is zero because the sensor

¹ Figure adapted from Sony ICX204AL datasheet

has square pixels. A standard camera calibration routine was used to measure the necessary intrinsic camera parameters as outlined in Appendix C.

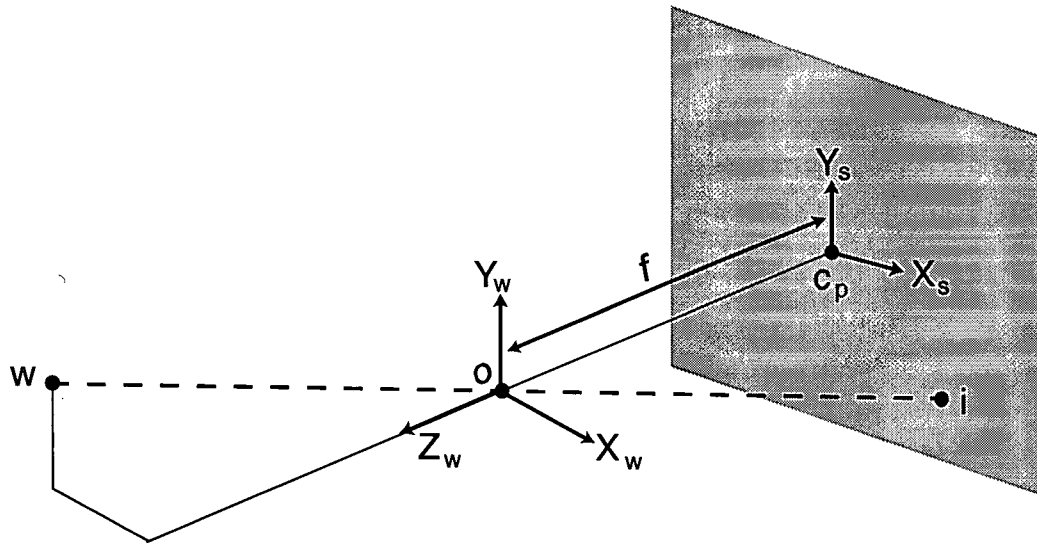


Figure 2-6 - Pinhole Camera Model

- w : Point in the external world (3D)
- i : Projected point on image plane (2D)
- o : Origin of world coordinate system
- f : Focal distance of camera
- c_p : Critical point
- X_s : X coordinate of sensor axis
- Y_s : Y coordinate of sensor axis
- X_w : X coordinate of world axis
- Y_w : Y coordinate of world axis
- Z_w : Z coordinate of world axis

Using the pinhole camera model a point w in the world coordinate system projects onto the camera plane at the point i using the relation shown in equation 1.

$$\begin{bmatrix} i_x \\ i_y \end{bmatrix} = \frac{f}{w_z} \begin{bmatrix} w_x \\ w_y \end{bmatrix} \quad (1)$$

2.2.4. Electronics

An electronic system is required for synchronizing the camera shutter with the lighting system, as well as for controlling which of the off-axis lights are activated. The electronics are explained in the following sections with a schematic of the circuitry shown in Appendix D.

2.2.4.1. Synchronization Pulse

One of the four GPIO pins on the camera is tied to the camera shutter and configured to output a short pulse at the start of each frame. The shutter pulse is used to toggle between activating the on-axis and off-axis light sources as shown below in Figure 2-7.

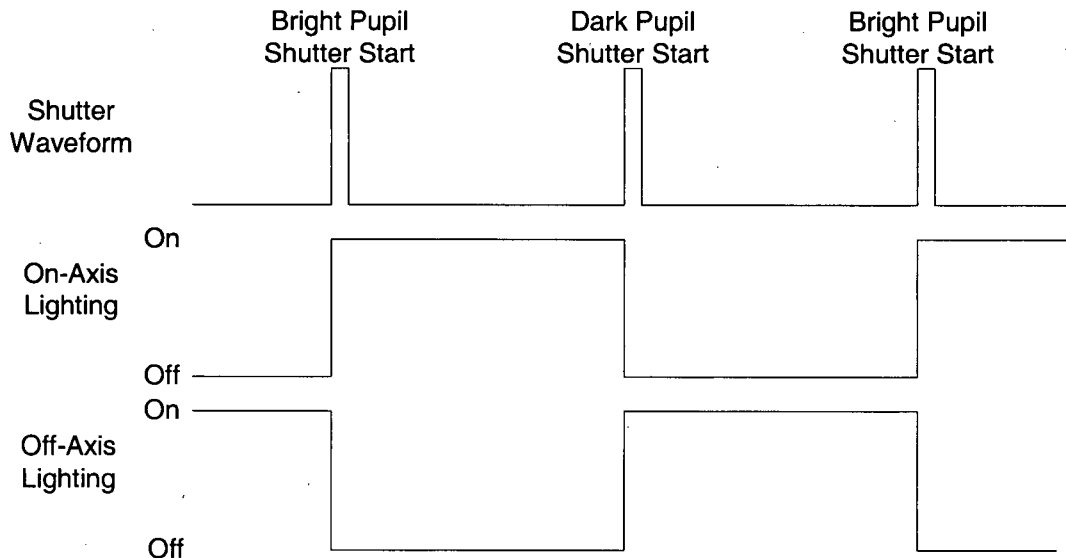


Figure 2-7 - IR Lighting Waveforms

2.2.4.2. Off-axis Light Selection

The dynamic glint source lighting requires the ability to activate any 2 of the 6 different off-axis light sources. The 3 remaining GPIO pins of the camera control are used to enable or disable any of the individual system lights.

2.2.5. Computer System

The computer used for operating the eye-gaze tracking system is a standard personal computer with 512 megabytes of RAM and a 1.4 GHz AMD CPU. The operating system running on the computer is Windows XP. An internal IEEE 1394 Firewire PCI card is used to interface with the camera. The monitor upon which the users POG is estimated is a flat screen LCD operating at a resolution of 1280x1024 pixels and has dimensions of 35 cm wide and 28 cm tall.

2.2.6. Physical Layout

An image of the overall system is shown in Figure 2-8. The entire system is mounted on a set of extruded aluminum bars to fix the location of each component with respect to one another. This prevents the need for system re-calibration in the event that someone bumps the table and the location of one part was to shift with respect to another. The lights surrounding the monitor are mounted on aluminum rails for simplicity; in a commercial design they could be mounted into the monitor housing. The locations of the point light sources as well as 3 of the 4 monitor corners were measured manually with respect to the world coordinate system (located at the focal point of the camera) and are listed in Table 2-1.

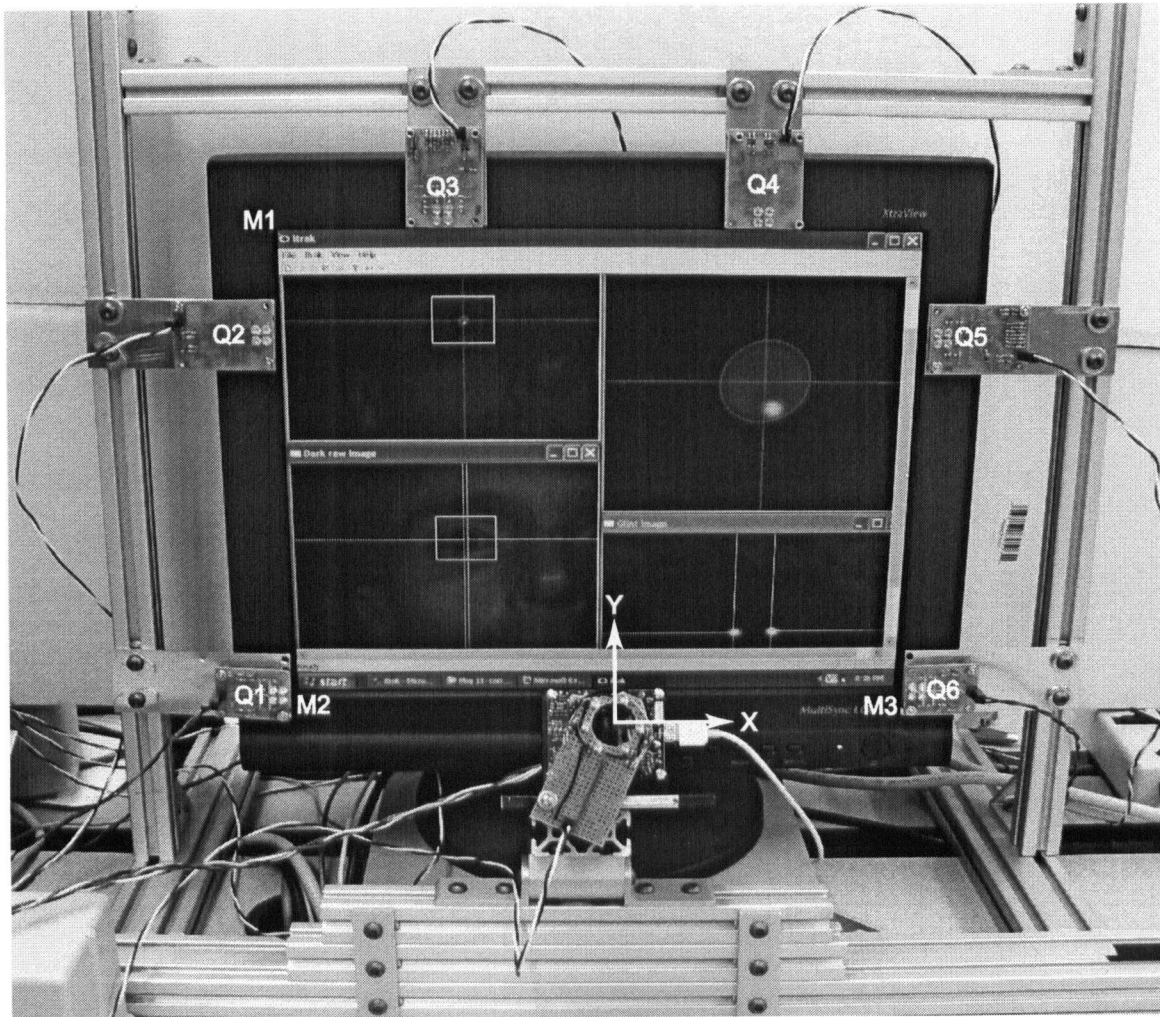


Figure 2-8 - Overall Physical System

Table 2-1 - Positions of Lights and Monitor

	Coordinates (cm)		
	X	Y	Z
Glint Source 1 (Q1)	-19.0	3.0	-12.5
Glint Source 2 (Q2)	-19.0	21.0	-6.5
Glint Source 3 (Q3)	-9.5	27.5	-2.0
Glint Source 4 (Q4)	8.5	27.5	-2.0
Glint Source 5 (Q5)	18.5	21.0	-6.5
Glint Source 6 (Q6)	18.5	3.0	-12.5
Top Left Screen Corner (M1)	-17.5	29.2	-1.5
Bottom Left Screen Corner (M2)	-17.5	4.0	-13.4
Bottom Right Screen Corner (M3)	17.7	4.1	-13.4

3. Eye and Feature Tracking

The eye and feature tracking systems are responsible for searching the entire FOV to determine the location of the eye and then to determine as accurately as possible, the locations of the pupil and the dual glints. The following sections outline the software system architecture and details the procedures through which eye and feature tracking takes place. The images shown in this chapter were selected to best illustrate the particular topic of discussion and do not all come from the same image sequence. A number of experimentally determined values are used in the following algorithms. These values are tabulated in Appendix E.

3.1. Software System Architecture

The eye-gaze tracking software system was developed in Matlab for ease of development and then ported to C++ for the faster execution time need for real-time performance. A Microsoft Windows based graphical user interface based on the Microsoft Foundation Classes was developed for user interaction. A custom matrix library was written to provide easy portability between the Matlab and C++ routines. An open source computer vision library called OpenCV [43] was used for the high speed image processing operations.

The entire system is scheduled with a timer running at 15 Hz (67 ms) which allows other processes to run in the downtime between captured images, reducing the load on the processor. When an image has been read by the camera, a system loop occurs in which the pupil and glint features are determined and sent to the POG estimation system for calculations. One entire system loop takes approximately 110 ms to complete when processing the entire image, which is slower than the 67 ms frame rate of the camera and requires 100 % of the processing power. The processing time for one system loop drops to 28 ms once the image processing system has locked the region-of-interest (ROI) onto the eye. With the ROI locked on the eye the system requires approximately 42 % of the processing power when operating at 15 Hz.

3.2. Image Processing Sequence of Operations

The end result of the image processing system is an equation of an ellipse for the pupil image contour and for both dual glint image contours. The control flow for the image processing system is shown in Figure 3-1 and each step is described in detail below. As part of the image processing system the ROI positioning and eyeglass reflection avoidance systems are also executed.

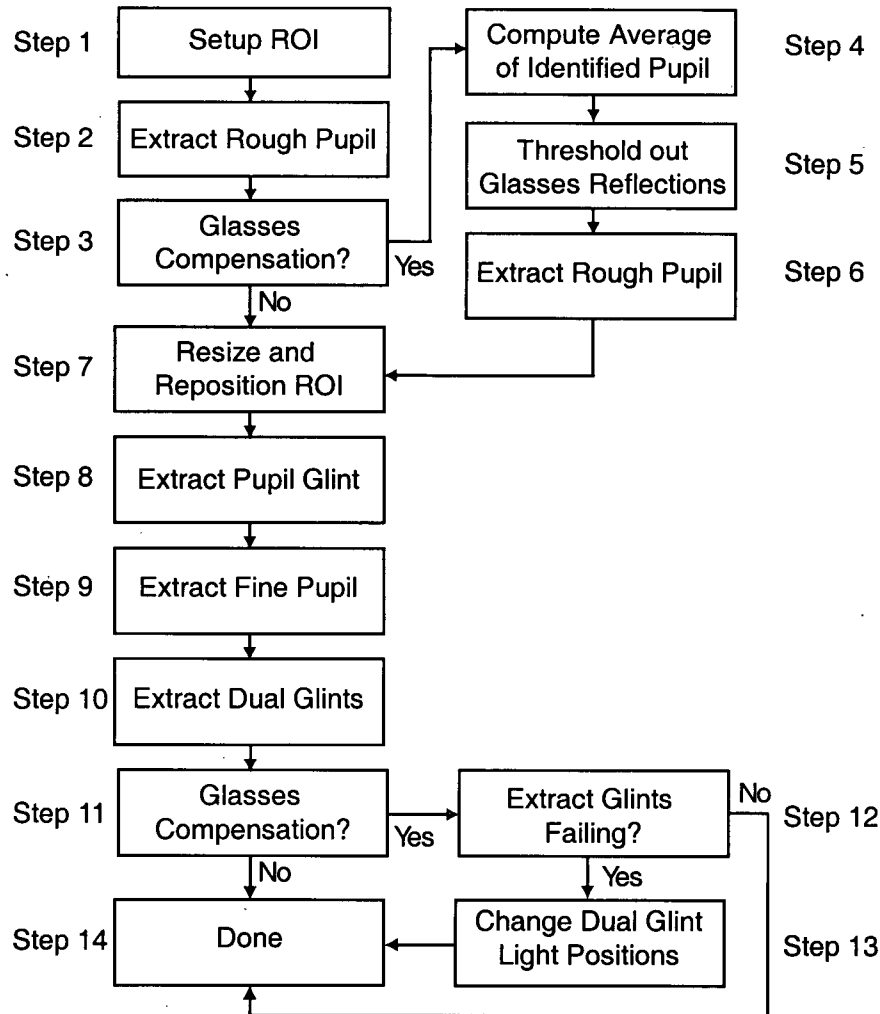


Figure 3-1 - Image Processing Flowchart

Step 1 - Setup ROI

The ROI is used to decrease the amount of image to be processed by subsequent operations. The first time the system analyzes the bright and dark pupil images, the entire image must be searched to determine the location of the eye using the rough pupil

identification algorithm. Once the location of the eye has been identified a ROI sized to fit just the eye is centered on the pupil. Subsequent image processing operations will only operate on the image information within the ROI. Future image processing loops will use the previously detected pupil center as the ROI center. If the eye has not moved outside of the ROI between image frames, the system can continue to re-center the ROI on the pupil without ever having to reprocess the entire image again.

Step 2 - Extract Rough Pupil

The rough pupil location is determined first as it the most likely feature to be reliably detected. The rough pupil extraction process is much faster than the fine pupil extraction but may result in a poorly formed pupil contour with ragged edges due to the method of extraction. The rough pupil extraction process is described in Chapter 3.3.

Step 3 - Glasses Compensation?

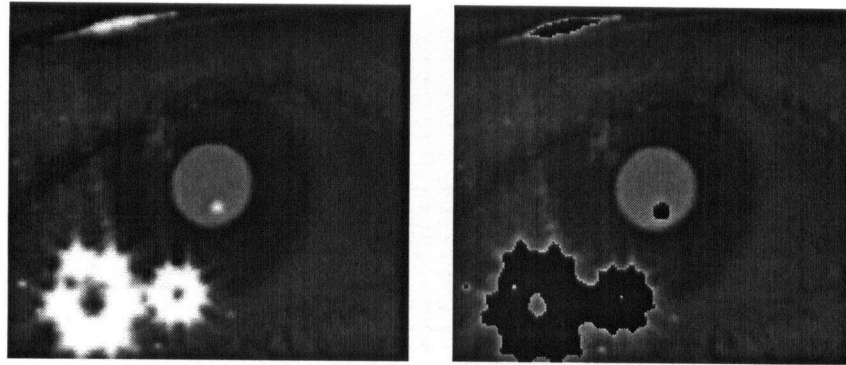
If the user is wearing eye-glasses the glasses compensation algorithm should be turned on. The glasses compensation algorithm helps prevent bright reflections off eye-glasses from interfering with the pupil identification process. If glasses compensation is turn on proceed to Step 4, otherwise skip to Step 7.

Step 4 - Compute Average of Identified Pupil

To determine if eye-glass reflections are present, the average pupil contour intensity is compared with a predetermined threshold. Reflections off eye-glasses are far brighter than the reflection off the surface of the retina. If the average pupil value is above a predetermined threshold then the identified "pupil" is considered to be a reflection from the eye-glasses and not the actual pupil.

Step 5 - Threshold Glasses Reflections

To remove the bright reflections from further consideration as part of the pupil contour, all pixels with intensity values greater than the threshold are set to zero. The inverted threshold to zero operation also removes the glint located on the bright pupil caused by the on-axis light source. This is not a problem as in future steps the pupil glint is removed anyway. An example of the thresholding operation is shown in Figure 3-2.



Before Thresholding

After Thresholding

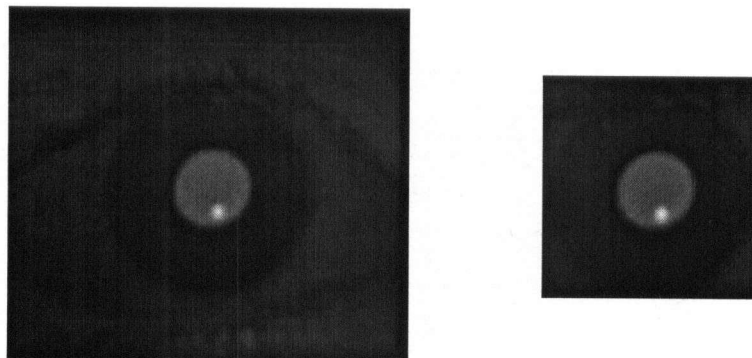
Figure 3-2 - Bright Pupil Thresholded for Glasses

Step 6 - Extract Rough Pupil

The rough pupil extraction algorithm is run once more on the image with the reflections removed. Provided the reflection had not overlapped the true pupil contour, the identification of the pupil should succeed.

Step 7 - Resize and Reposition ROI

Once the center of the rough pupil is identified the ROI is re-centered to the new pupil center and reduced in size to cover just the area of the cornea as shown in Figure 3-3. Reducing the ROI further increases the speed of execution.



Eye ROI

Cornea ROI

Figure 3-3 - Eye and Cornea ROI

Step 8 - Pupil Glint Extraction

The pupil glint is detected in the bright pupil image after the rough pupil center is identified. The pupil glint information is used to refine the detected pupil contour. The method used to detect the pupil glint is described in Section 3.4.

Step 9 - Fine Pupil Extraction

The fine pupil extraction searches the bright pupil image for the pupil contour. An ellipse equation is fit to the perimeter of the pupil contour which will be used in the POG estimation algorithm. The fine pupil extraction method described in Section 3.5 compensates for a number of common sources of error when determining the pupil contour such as obscuration by the pupil glint, eyelashes, and eyelids [44]. The center of the detected fine pupil is used as the ROI center for the rough pupil search in the next image processing loop.

Step 10 - Dual Glint Extraction

The locations of the dual glints are determined from the dark pupil image. Ellipse equations are fit to each of the dual glint contours for further use in the POG estimation algorithm. The method for determining the dual glints is described in Section 3.6.

Step 11 - Glasses Compensation?

The glasses compensation algorithm also compensates for reflections of eye-glasses that may interfere with the dual glint identification process. If glasses compensation is turned on, proceed to Step 12, otherwise skip to Step 14.

Step 12 - Extract Glints Failing?

If the average intensity of the search area in which the glints should be located is greater than a certain threshold then it is likely that there are larger glints from the glasses corrupting the images. After a number of such failures a control signal is sent to change the location of the active glint light sources. An example of glasses reflections causing a dual glint detection failure is shown in Figure 3-4. The two small vertical shapes are the correct glint reflections off the surface of the cornea, the large starburst shaped contours are the corrupting reflections off the surface of the eye-glasses.

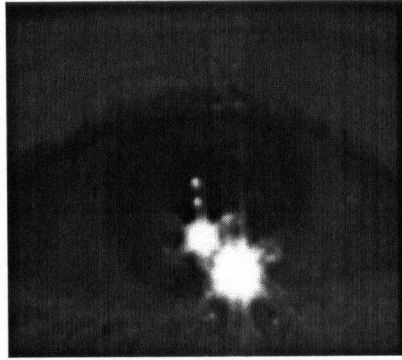


Figure 3-4 - Dual Glint Detection Failure

Step 13 - Change Dual Glint Light Positions

If glint extraction has failed due to eye-glass reflections, the current dual glint light sources are turned off and an opposite set of lights are activated. The dark pupil image is shown again in Figure 3-5, a few frames after the active glint light sources have changed position. The new glint light sources have changed both the glint locations as well as avoiding the corrupting eye-glass reflections.

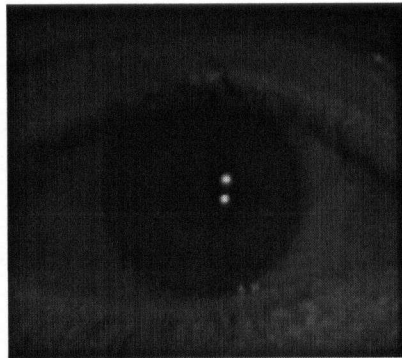


Figure 3-5 - Dual Glint Recovery

Step 14 - Done

Image processing is complete and the POG calculations begin.

3.3. Rough Pupil Processing

The rough pupil detection system is based on the method developed by Nouredin [15] which estimates the location of the pupil anywhere in the FOV of the camera. A flowchart of the rough pupil detection algorithm is shown in Figure 3-6. In addition to the various features of the face and head there is possibly a great deal of variation in the

background of the image making it difficult to segment the small dark pupil. The image difference method previously discussed is used to aid in identifying the location of the pupil. The rough pupil algorithm excels at detecting the location of the pupil in the scene but does not always identify the exact perimeter of the pupil contour required for the POG estimation routine. Refining the pupil contour is the purpose of the fine pupil processing system.

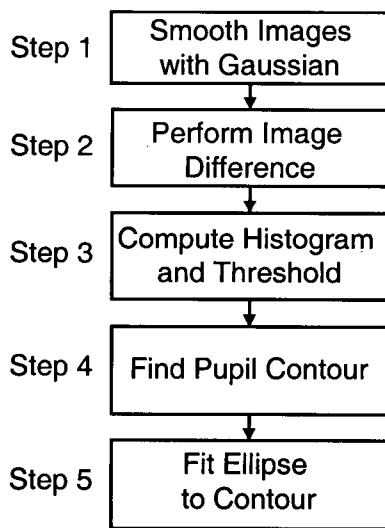
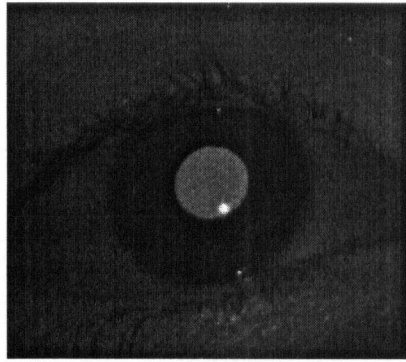


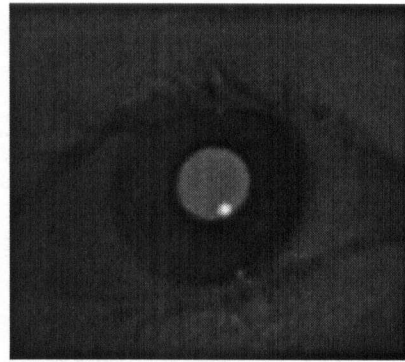
Figure 3-6 - Rough Pupil Detection

Step 1 - Image Smoothing with Gaussian

The bright and dark pupil images are preprocessed with a Gaussian smoothing filter, as shown in Figure 3-7 for the bright pupil image, to attenuate random noise in the images which could be exacerbated by the subsequent difference operation. There is considerable salt and pepper noise due to a large gain setting used on the camera. The high gain is required to make the image bright enough to clearly distinguish features.



Pre-Gaussian

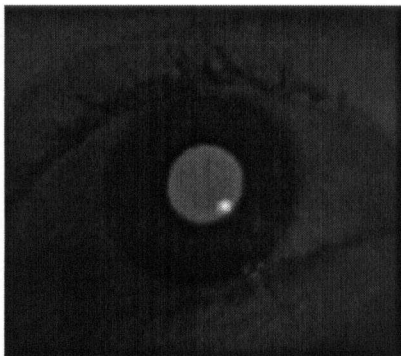


Post-Gaussian

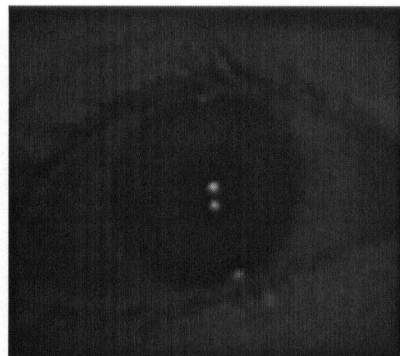
Figure 3-7 - Gaussian Filter Applied to Raw Images

Step 2 - Image difference

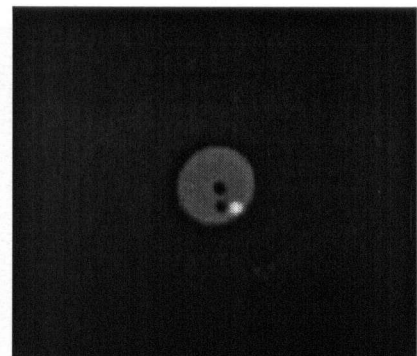
The dark pupil image is subtracted from the bright pupil image to produce a difference image. As mentioned previously, the lighting scheme was designed to closely illuminate the face between the on-axis and off-axis illumination. The pupil is bright in one image and dark in the next, which is highlighted by the difference operation as shown in Figure 3-8. In the difference image note the bright glint from the on-axis light source and the two dark circular areas which resulted from the subtraction of the two bright glints in the dark pupil image generated by the off axis light sources.



Bright Pupil



Dark Pupil



Difference

Figure 3-8 - Differencing Images of Eye

Step 3 - Histogram and Threshold Calculation

A histogram is used to determine the pupil / background threshold level. The histogram is computed from the difference image, an example of which is shown in Figure 3-9. The histogram appears bi-modal in which the brighter mode corresponds to

the pupil and the dark mode corresponds to the background. The threshold between the two modes is determined roughly by summing the histogram bins backwards from brightest to darkest until a fixed number of pixels have been counted [45]. The total number of pixels summed is an experimentally determined value, chosen to provide a sufficient threshold estimation for identifying the rough pupil while including the minimum number of artifacts. In the example shown the estimated threshold level determined is 30.

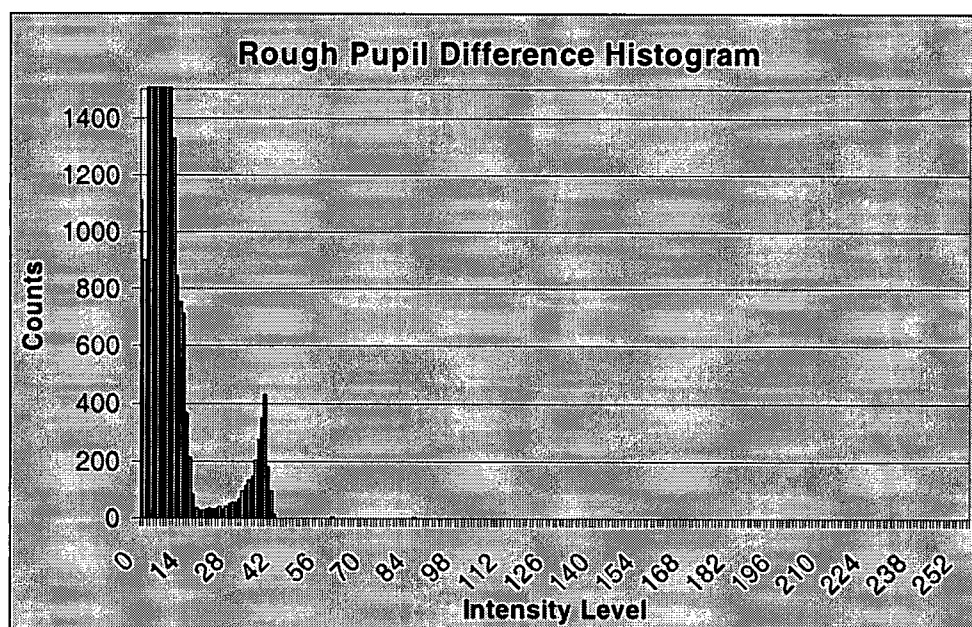


Figure 3-9 – Rough Pupil Image Histogram

The difference image is thresholded at the determined level to create a binary image in which the pupil should be the largest contour. An example of the binary difference image is shown in Figure 3-10. Note that the thresholding operation has left two holes in the contour from the two bright glints in the dark pupil image and an attached contour from the pupil glint in the bright pupil image. These artifacts are compensated for in the fine pupil extraction method.

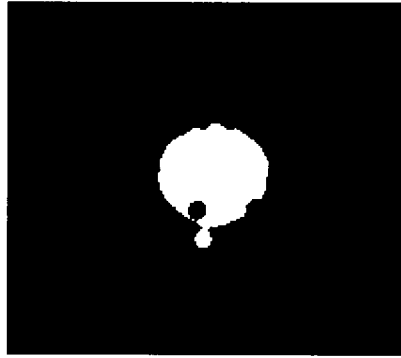


Figure 3-10 – Binary Result of Thresholding for the Rough Pupil

Step 4 - Find Pupil Contour

The largest contour in the image is most likely the pupil. Identification is easily performed by identifying the contour with the largest number of pixels that is within a pre-determined maximum and minimum width and height. A rejection test is performed on the identified contour to ensure it is actually the pupil that was detected. Non-pupil contours are usually jagged and / or noncircular in shape which can be detected by computing the isoperimetric quotient of the contour [46]. The isoperimetric quotient is a ratio of area to perimeter as shown:

$$IQ = \frac{(4 \cdot \pi \cdot A)}{P^2} \quad (2)$$

A : Contour area

P : Contour perimeter

IQ : Isoperimetric quotient

The isoperimetric quotient is used to identify how close a contour is to a circle, the closer the shape is to a circle the closer the quotient is to a value of 1. A jagged shape with a long perimeter but little internal area would have a quotient of closer to 0.

The computed contour quotient is compared against an experimentally determined minimum to determine if the contour should be rejected. If the pupil is not detected, the image processing algorithm exits and tries again on the next image. After a fixed number of such failures the systems gives up on the current ROI and reprocesses the entire image to relocate the pupil. Waiting for a number of failures before rejecting the ROI avoids re-

processing the entire image when the pupil is lost for only a short duration due to an eye blink.

Step 5 - Fit Ellipse

An ellipse fitting algorithm [47] is applied to the boundary points of the pupil contour. The algorithm uses a least mean squares approach to fitting the equation of an ellipse to the perimeter points of the contour. The center of the ellipse equation is used as the center of the rough pupil for further processing.

3.4. Pupil Glint Detection

The pupil glint formed on the bright pupil image by the on-axis lighting can interfere with the precise identification of the pupil contour by the fine pupil detection system. By identifying the location of this glint, it can be removed to prevent corrupting the perimeter of the pupil contour. A flowchart of the pupil glint detection operation is shown in Figure 3-11. Identification of the glint is made possible by the fact that the glint generates the highest intensity pixels in the image.

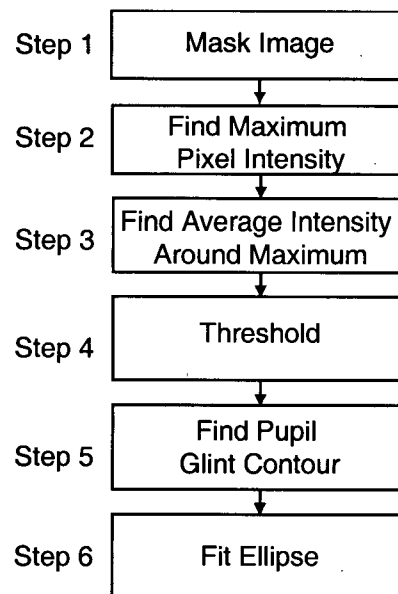


Figure 3-11 - Pupil Glint Detection Flowchart

Step 1 - Mask Image

The only pupil glint candidates of interest are located on the surface of the cornea. To reduce the chance of identifying a glint located on the sclera, a circular mask is

applied to the bright pupil image, centered on the rough pupil ellipse center. The radius of this circle is large enough to encompass the image of most individual's corneas. As seen in Figure 3-12 only the iris and pupil remain after the masking operation. The mask diameter is seen here to be slightly smaller than the iris diameter.

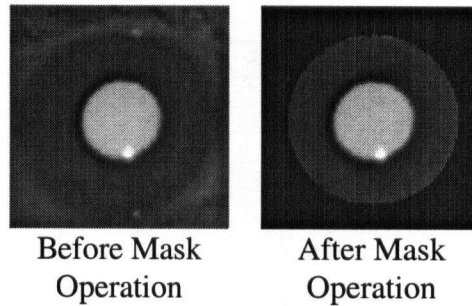


Figure 3-12 - Pupil Glint Mask

Step 2 - Find Maximum Pixel Intensity

To determine the threshold for the pupil glint, the location of the pixel with the maximum intensity level is found. This is done using a simple maximum function.

Step 3 - Find Average Intensity Around Maximum

The average intensity level of the maximum pixel and the 8 surrounding pixels is determined. This average level is reduced by an experimentally determined percentage resulting in the dual glint threshold level. The threshold is reduced to encompass more glint pixels than just the brightest 9. A magnified image (12x13 pixels) of one of the glints is shown in Figure 3-13 with the maximum pixel identified and the 8 surrounding pixels marked with 'X's.

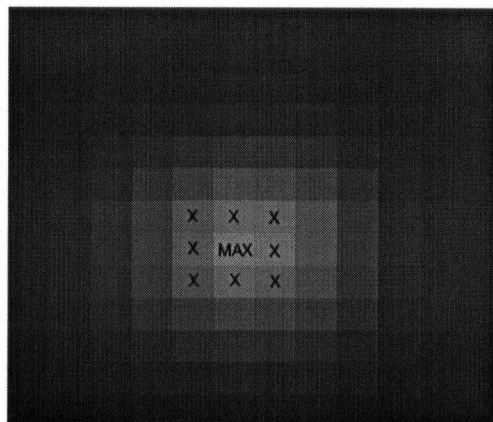


Figure 3-13 - Glint Threshold Extraction

Step 4 - Threshold

The masked bright pupil image is thresholded with the determined value, resulting in the binary image shown in Figure 3-14. In this image it is easy to determine the pupil glint contour, however, on occasion there are other spurious artifacts. These artifacts may be due to reflections at the cornea / sclera boundary or due to the threshold level being set too low and allowing non glint pixels through the thresholding operation.

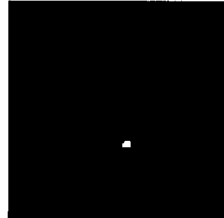


Figure 3-14 - Dual Glint Threshold

Step 5 - Find Pupil Glint Contour

To determine which contour is the pupil glint contour, a list of all contours in the image is created. The contour list is searched for a shape that fits the range of sizes expected for a pupil glint contour.

Step 6 - Fit Ellipse

The center of the detected contour is determined by fitting an ellipse to the perimeter of the glint contour. The center of the ellipse equation is used as the center of the pupil glint.

3.5. Fine Pupil Detection

The fine pupil detection system determines the exact perimeter of the pupil contour, which is required for the POG estimation algorithm. The fine pupil detection system operates on only the bright pupil image to avoid any motion artifacts that may be introduced by the image differencing operation. The rough pupil and pupil glint contours previously detected are used as masks to aid in detecting the fine pupil perimeter very accurately. Once the fine pupil contour has been detected, an ellipse equation is fit to the perimeter of the pupil for future use by the POG estimation system. The flowchart of the steps in the fine pupil detection is shown in Figure 3-15. Steps 1 through 7 are used to threshold the bright pupil image to segment the pupil contour. Steps 8 through 11 are

performed to refine the edges of the pupil contour to fit the best possible ellipse to the perimeter.

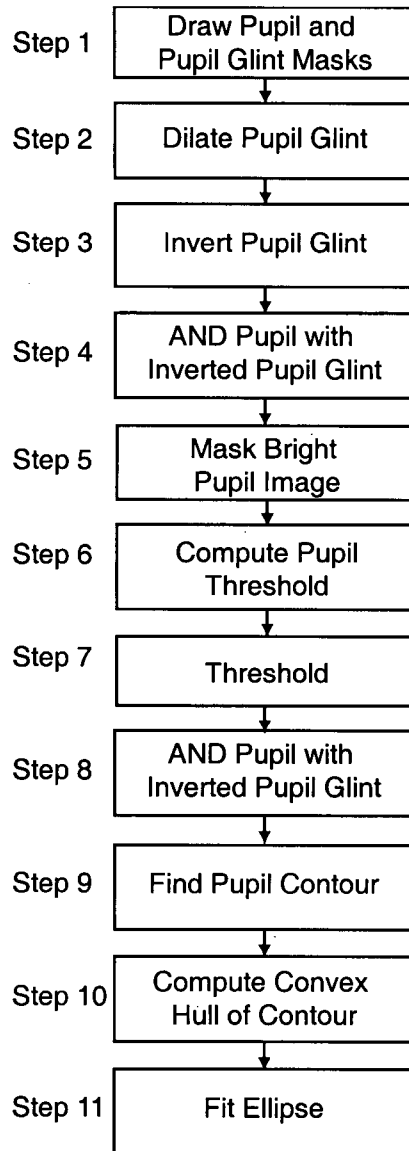


Figure 3-15 - Fine Pupil Ellipse Estimation Flowchart

Step 1 - Draw Pupil and Pupil Glint Masks

The previously identified rough pupil and pupil glint contours are used as masks in the following steps. Examples of the rough pupil and pupil glint masks are shown below in Figure 3-16.

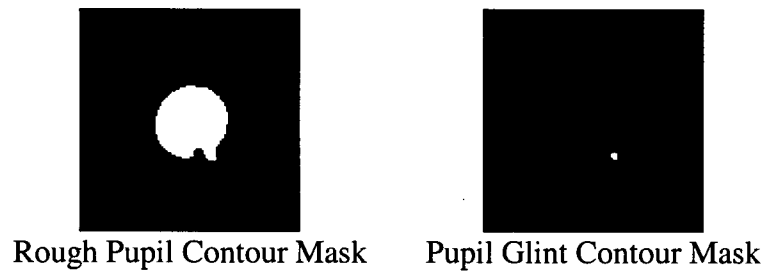


Figure 3-16 - Fine Pupil Extraction Masks

Step 2 - Dilate Pupil Glint Contour

The glint pixels have a higher intensity level than the pupil contour. To get an accurate average pupil intensity level, the glint pixels must be removed. The pupil glint identification procedure produces a contour that is slightly smaller than the actual pupil glint. Dilation of the pupil glint contour ensures that a sufficient amount of the pupil contour will be removed in Step 7. The dilated pupil glint contour shown in Figure 3-17 is larger than the actual pupil glint, which results in removing too much of the pupil in Step 7. The excess pupil removal is compensated for in Step 8.

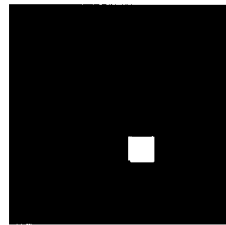


Figure 3-17 - Dilated Pupil Glint Contour

Step 3 - Invert Pupil Glint Contour

To perform the masking or cutting operation the dilated pupil glint contour is inverted as shown in Figure 3-18



Figure 3-18 - Inverted Pupil Glint Contour

Step 4 - AND Rough Pupil Contour with Inverted Pupil Glint

To remove the pupil glint from the pupil contour the inverted pupil glint is ANDed with the rough pupil contour, resulting in the image shown in Figure 3-19.

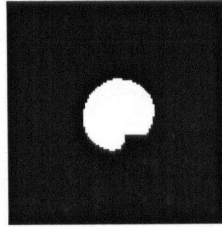
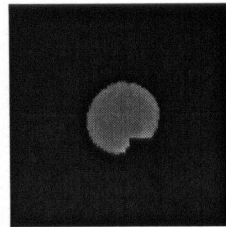
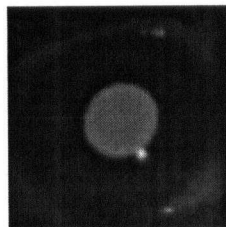


Figure 3-19 - Pupil Glint Mask AND Pupil Contour

Step 5 - Mask Bright Pupil Image

To determine the proper threshold value to segment the fine pupil contour the average intensity level of the pupil in the bright pupil image is required. To remove all but the pupil the mask from Step 4 is applied to the bright pupil image. The masking operation leaves only the grayscale pupil pixels as shown in Figure 3-20. The average value of the remaining pupil pixels is computed to provide an estimate for the average pupil intensity level.



Original Bright Pupil Image Masked Bright Pupil Image

Figure 3-20 - Masked Bright Pupil

Step 6 - Compute Pupil Threshold

The fine pupil threshold is the average value reduced by an experimentally determined percentage. If only the average value was used as the threshold level some pupil pixels would be above and some would be below the threshold.

Step 7 - Threshold

The threshold value computed is applied to the bright pupil image resulting in the binary image shown in Figure 3-21. The pupil glint is still attached to the pupil contour in this image and is removed in the following step.

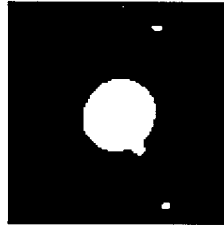


Figure 3-21 - Fine Pupil Contour Thresholded

Step 8 - AND Fine Pupil Contour with Inverted Pupil Glint

If the pupil glint is on the border of the pupil it can cause a slight distortion in the pupil perimeter. If the glint is entirely internal to the pupil contour or completely separated then no perimeter distortion occurs. To remove the pupil glint the fine pupil contour is ANDed with the inverted pupil glint as shown in Figure 3-22. The excess pupil contour that is removed is compensated for with the convex hull operation in Step 10.



Figure 3-22 - Fine Pupil Contour ANDed with Inverted Pupil Glint

Step 9 - Find Pupil Contour

The fine pupil operations above may have introduced multiple contours in the resulting image. To identify the pupil contour, a bounding box is fit to each contour in the image and the distance from the center of each bounding box to the center of the rough pupil ellipse is determined. The smallest distance corresponds with the fine pupil contour.

Step 10 - Compute Convex Hull of Contour

The pupil contour should be an ellipse with no concave sections along the perimeter. Concave sections may arise from eye lashes overlapping the pupil or the masking operation with the pupil glint. To restore the contour to a convex structure the convex hull operation is performed. The convex hull operation fills in any concave sections of the contour as illustrated in Figure 3-23.

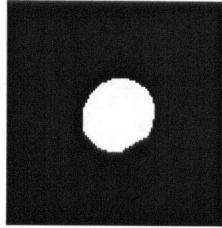


Figure 3-23 - Fine Pupil Convex Hull

Step 11 - Fit Ellipse

The perimeter pixels of the detected fine pupil contour are finally fit to an ellipse equation using the ellipse fitting algorithm [47]. A magnified image of the identified pupil contour is shown below in Figure 3-24.

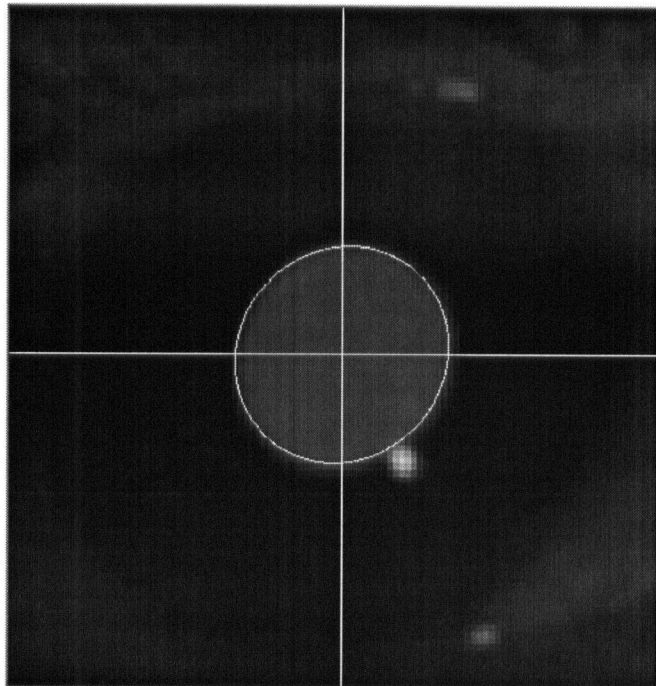


Figure 3-24 - Identified Fine Pupil Contour

3.6. Dual Glint Detection

The dual glint detection system determines the image locations of the dual glints formed by the off-axis lighting in the dark pupil images. The centers of the dual glint contours are required to estimate the POG. A flowchart of the dual glint detection operation is shown in Figure 3-25. This procedure is the same as the pupil glint detection system in Section 3.4, with the exception of Step 6.

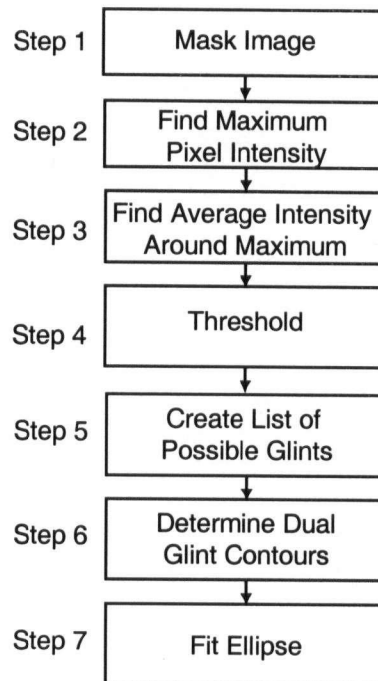


Figure 3-25 - Dual Glint Detection Flowchart

Step 6 - Determine Dual Glint Contours

The glint contour list is searched to find a pair of contours which match a pre-determined relative X and Y pixel displacement with one another. The displacements to search for depend on which of the 6 glint light sources are activated. The maximum and minimum X and Y displacements for each possible combination of glints was measured from the image shown in Figure 3-26, which shows the glints formed by all 6 of the glint light sources.

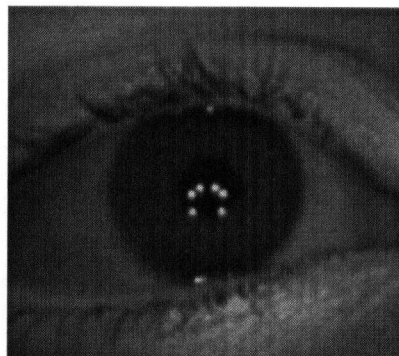


Figure 3-26 - All Off-Axis Glint Light Sources Enabled

An example of the extracted dual glints is shown in Figure 3-27.

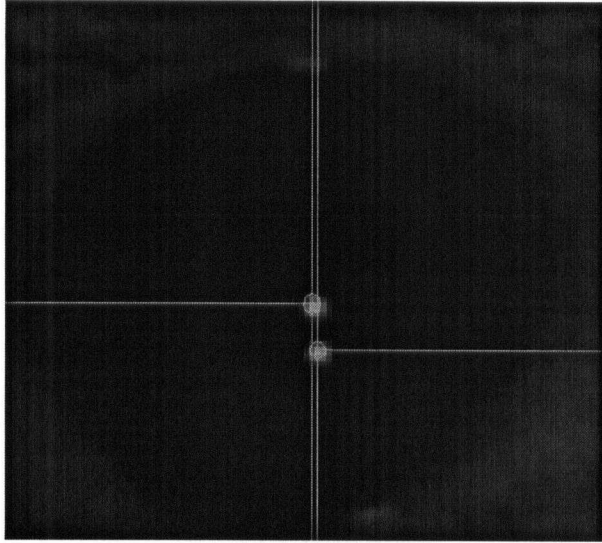


Figure 3-27 - Identified Dual Glint Contours

4. Optical Geometry for Point of Gaze Estimation

An image based approach has been developed for estimating the point of gaze of a subject. Structured lighting is used to illuminate the users face, generating images with information sufficient to compute the POG. A flowchart outlining the POG estimation algorithm is shown in Figure 4-1 with a detailed explain of each step in the following sections. A schematic of the physical system is shown in Figure 4-2. The notation used in this chapter is that non-bold variables are scalars, lower case variables are points and upper case variables are vectors.

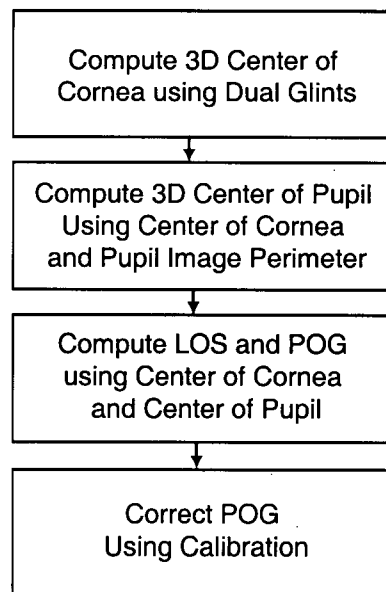


Figure 4-1 - POG Estimation Algorithm Flowchart

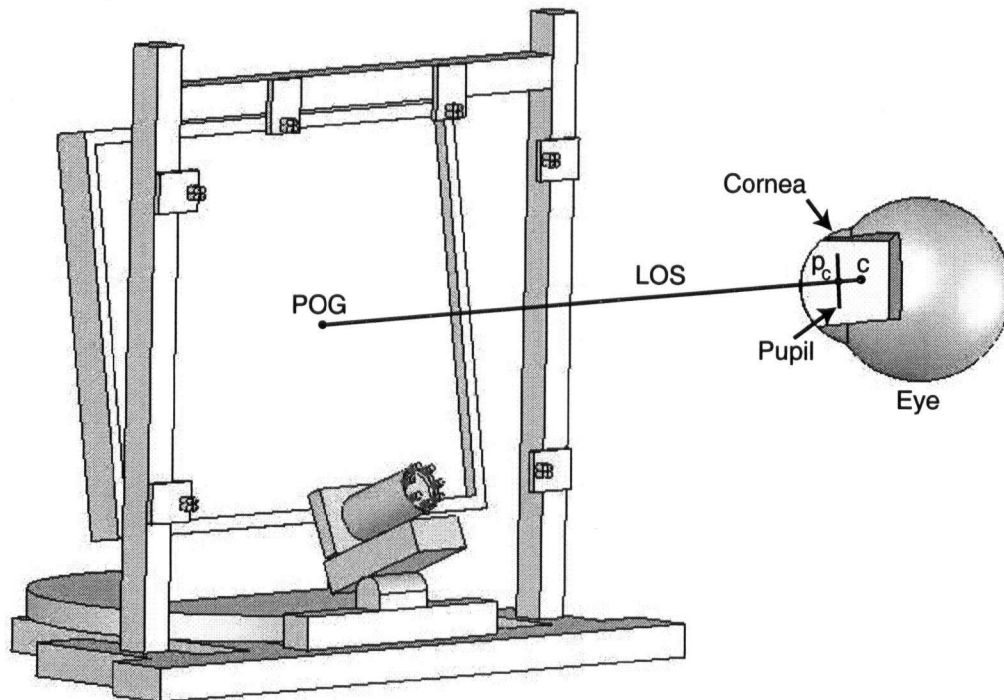


Figure 4-2 – POG Estimation System Overview

c : Center of corneal sphere model

p_c : Center of the pupil

LOS : Line-of-sight the vector formed from **p_c** to **c**

POG : Point-of-gaze found by intersecting the **LOS** with the monitor plane

In addition to the pupil contour perimeter and location of the dual glints in the captured images, models of the eye, camera and system are required to compute the POG. The camera and system models were described in Section 2.2.3.2 and Section 2.2.6 respectively. The schematic eye developed by the ophthalmologic researcher Alvar Gullstrand [13] is used as the eye model. A simplified diagram of the schematic eye showing the parameters of interest is shown in Figure 4-3. This model assumes that the surface of the cornea is spherical.

The parameters of interest in the schematic eye are the radius of the cornea, the distance from the center of the cornea to the center of the pupil, and the index of refraction of the aqueous humor. Gullstrand compiled population norms for the values for the parameters of the schematic eye which are listed in Table 4-1. These population averages are used in the POG algorithm. The final stage of POG estimation involves

correcting the estimated POG for possible sources of error, some of which would be the difference between the population norms and the actual values of the user operating the system.

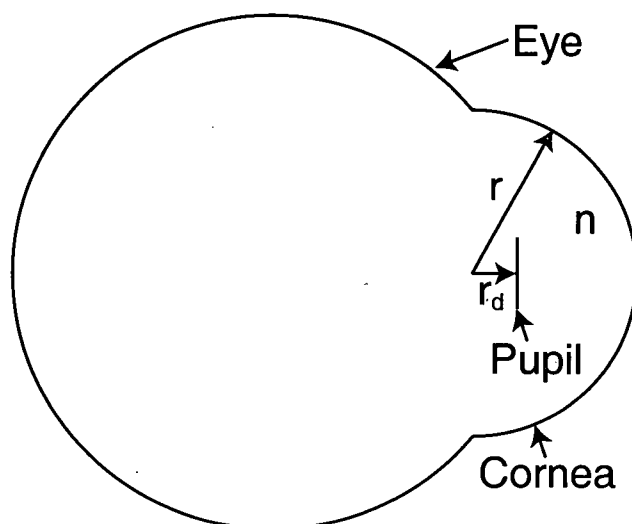


Figure 4-3 - Eye Model

- r : Radius of the corneal sphere model
- r_d : Distance from the center of corneal sphere to the center of pupil
- n : Index of refraction of the aqueous humor fluid

Table 4-1- Schematic Values for the Eye

Parameter	Value
r	0.7 cm
r_d	0.42 cm
n	1.376

4.1. Computing the Center of the Corneal Sphere Model

The first step in computing the POG is to compute the center of the cornea sphere. The geometry first outlined by Shih [31] is used to triangulate the center of the cornea sphere model using the image locations of two glints off the surface of the cornea. The diagram in Figure 4-4 illustrates the flow of light from the glint light sources q_1 and q_2 to the surface of the cornea at g_1 and g_2 and reflected through the pin-hole camera model focal point o onto the camera sensor plane at i_1 and i_2 . A flowchart of the steps used in computing the cornea center are shown in Figure 4-5.

Note that any two of the six off-axis light sources located around the monitor can be used as the glint generators. If any of the active light sources cause reflections off eyeglasses that interfere with the images, the offending light can be switched off and another switched on.

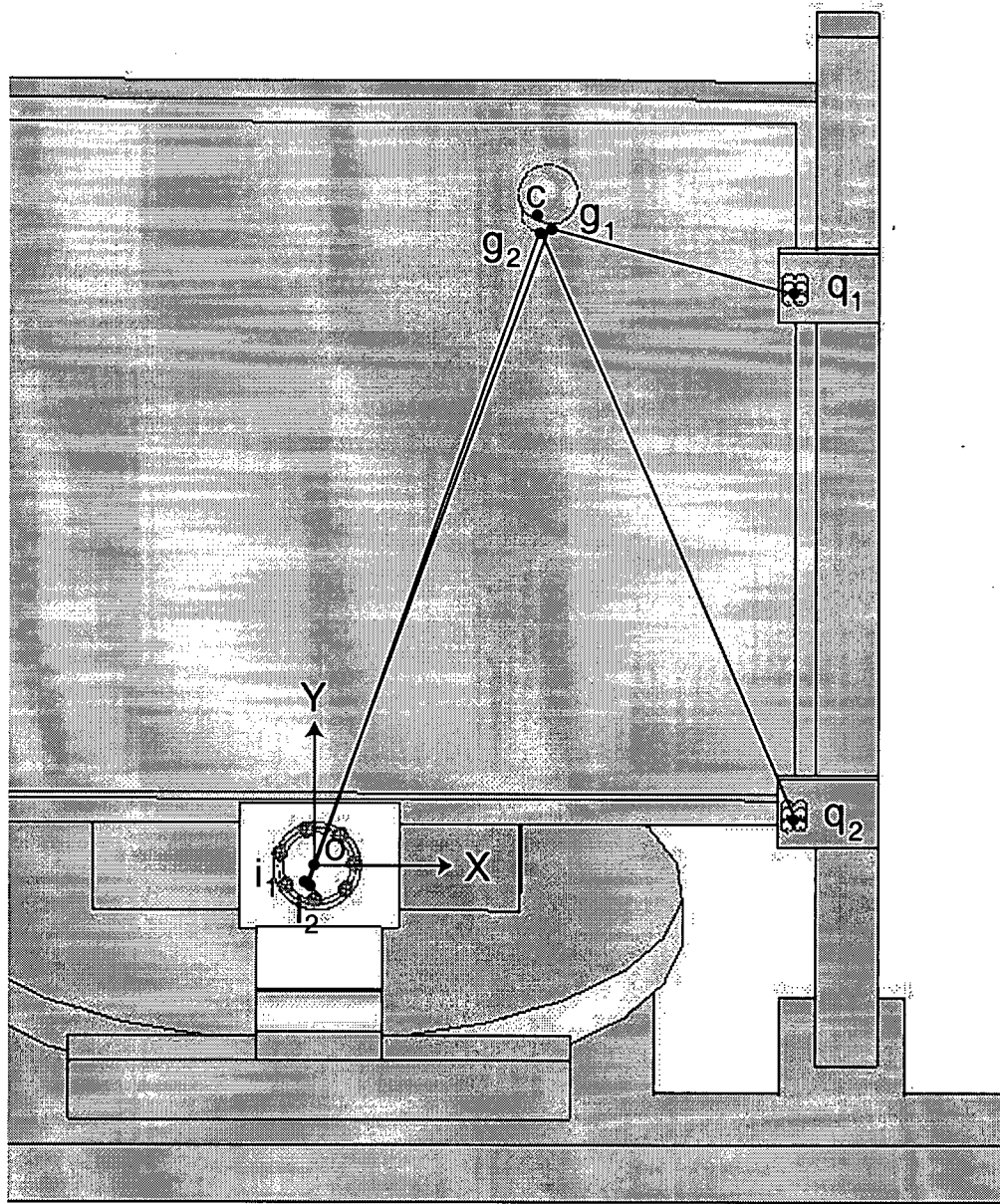


Figure 4-4 - Cornea Center Estimation World Coordinate System

- o** : World coordinate system origin (focal point of camera)
- c** : Center of the corneal sphere model
- q₁** : Location of the glint 1 light source

- q_2** : Location of the glint 2 light source
- g_1** : Location of glint 1 on the surface of the cornea
- g_2** : Location of glint 2 on the surface of the cornea
- i_1** : Location of the image of glint 1
- i_2** : Location of the image of glint 2
- X** : The X axis of the world coordinate system, horizontal to CCD sensor
- Y** : The Y axis of the world coordinate system, vertical to CCD sensor
- Z** : The Z axis of the world coordinate system, normal to the CCD sensor
(out of the page)

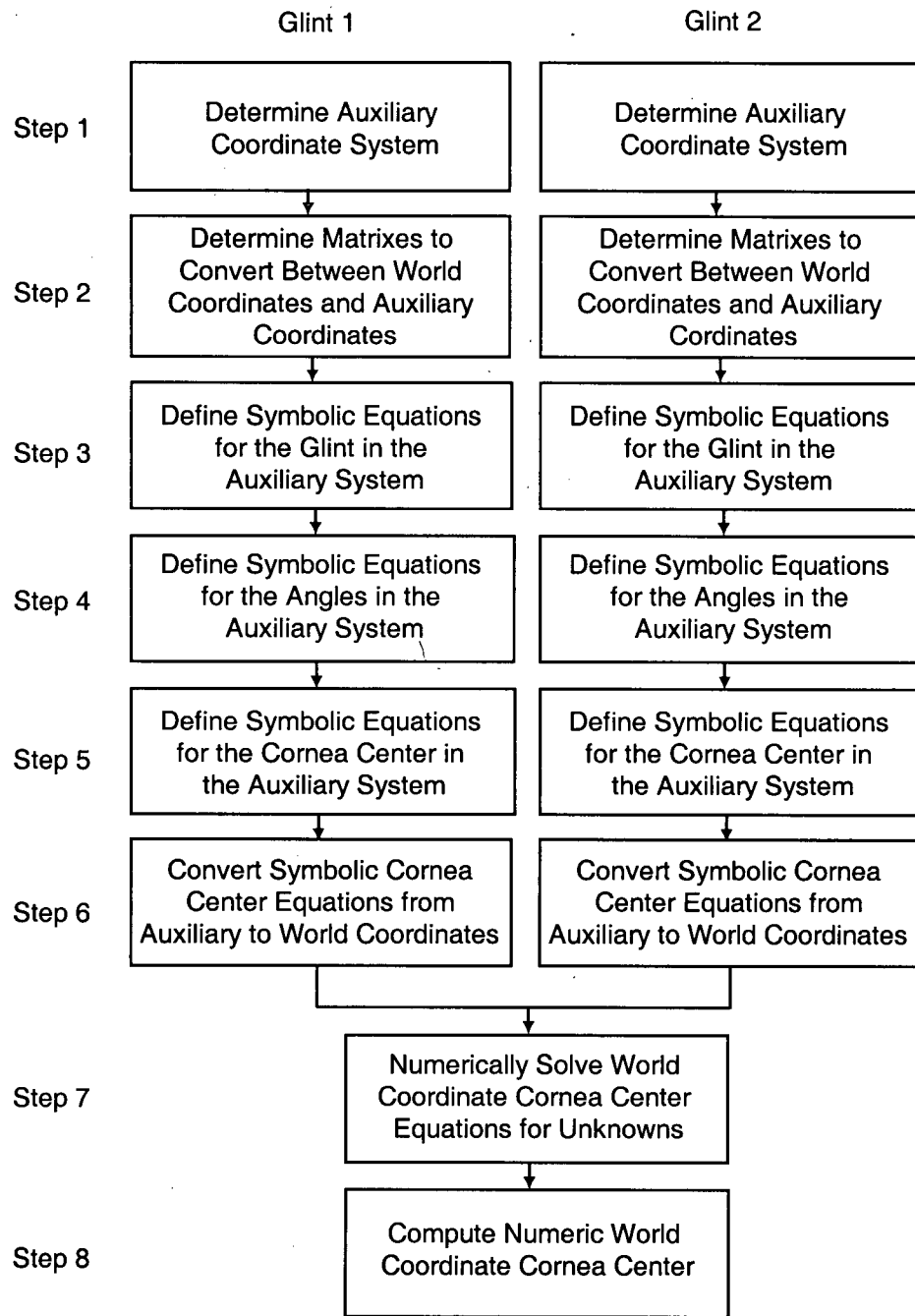


Figure 4-5 - Cornea Center Calculation Flowchart

Step 1 - Determine Auxiliary Coordinate System

Solving for the center of the cornea \mathbf{c} is simplified by noting that because the reflections of the glints are off of the surface of a spherical object, the set of points \mathbf{q}_1 , \mathbf{g}_1 , \mathbf{c} , \mathbf{o} and \mathbf{i}_1 are all co-planar, as are the set of points \mathbf{q}_2 , \mathbf{g}_2 , \mathbf{c} , \mathbf{o} and \mathbf{i}_2 . An auxiliary coordinate system can be defined for each set of glint points such that they then lie in an

axis plane, reducing the problem complexity from 3D to 2D. The auxiliary coordinate system for each set of points is defined with the \hat{X}_i axis along the vector Q_i . The \hat{Y}_i axis is defined to be normal to \hat{X}_i and the vector $-I_i$. The \hat{Z}_i axis is defined such that the \hat{X}_i - \hat{Z}_i plane contains the vector $-I_i$. The diagram of Figure 4-5 has been modified to illustrate the auxiliary coordinate system for the first glint and is shown in Figure 4-6.

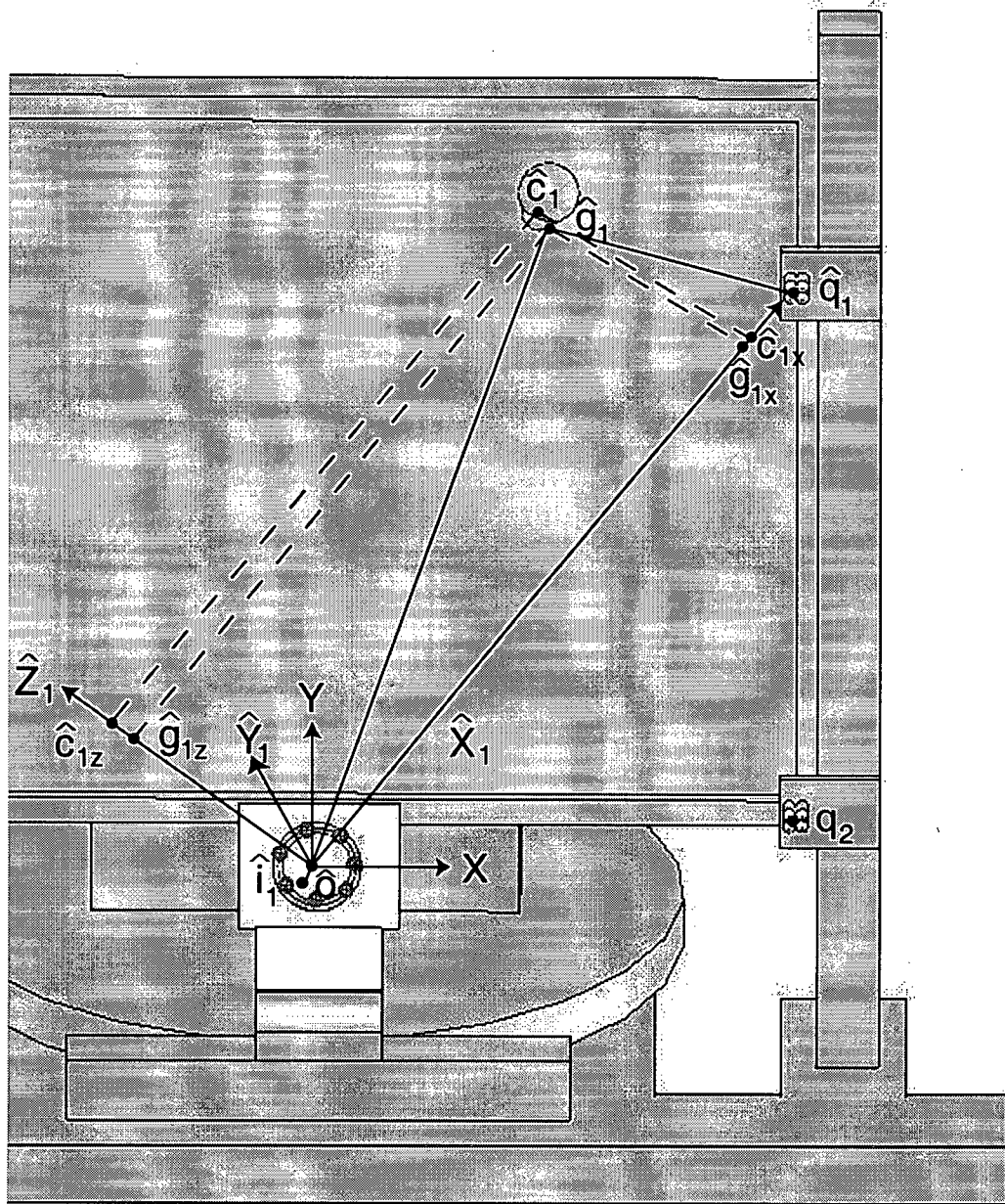


Figure 4-6 – Auxiliary Coordinate System for Cornea Center Estimation (as viewed along the world Z axis by a user)

- \hat{o} : Auxiliary coordinate system origin (focal point of camera)
- \hat{q}_1 : Location of the glint light source
- \hat{g}_1 : Location of glint on the surface of the cornea
- \hat{g}_{1x} : X coordinate of glint location
- \hat{g}_{1z} : Z coordinate of glint location
- \hat{c}_1 : Location of cornea center on the surface of the cornea
- \hat{c}_{1x} : X coordinate of cornea center location
- \hat{c}_{1z} : Z coordinate of cornea center location
- \hat{i}_1 : Location of the image of the glint
- \hat{X}_1 : The X axis of the auxiliary coordinate system
- \hat{Y}_1 : The Y axis of the auxiliary coordinate system
- \hat{Z}_1 : The Z axis of the auxiliary coordinate system
- X : The X axis of the world coordinate system
- Y : The Y axis of the world coordinate system
- Z : The Z axis of the world coordinate system

For each of the auxiliary coordinate systems the axes are computed using the following equations (where \times is the cross product and $\| \|$ is the norm operator):

$$\hat{X}_i = \frac{\mathbf{Q}_i}{\|\mathbf{Q}_i\|} \quad (3)$$

$$\hat{Y}_i = \frac{\mathbf{Q}_i \times \mathbf{I}_i}{\|\mathbf{Q}_i \times \mathbf{I}_i\|} \quad (4)$$

$$\hat{Z}_i = \hat{X}_i \times \hat{Y}_i \quad (5)$$

Step 2 - Determine Conversion Matrixes

To convert between the world and auxiliary coordinate systems a transformation matrix and its inverse are required. Only a rotation matrix is required as the origin is the same for both coordinate systems. There are three rotations required to rotate from the world to an auxiliary system. The rotation angles are easily calculated as both the world and auxiliary coordinates are known. The C++ implementation used to compute the

rotations is listed in Appendix F. A 3x3 rotation matrix **R** is defined for the auxiliary coordinate system for glint 1 and a 3x3 rotation matrix **S** is defined for the auxiliary coordinate system for glint 2 as shown in equations 6 and 7. The matrixes will remain in symbolic form for the subsequent equation derivations.

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \quad (6)$$

$$\mathbf{S} = \begin{bmatrix} s_{11} & s_{12} & s_{13} \\ s_{21} & s_{22} & s_{23} \\ s_{31} & s_{32} & s_{33} \end{bmatrix} \quad (7)$$

Step 3 - Determine Symbolic Equations for Glint Location

The auxiliary coordinate system has been defined such that all points lie within the $\hat{\mathbf{X}}_i - \hat{\mathbf{Z}}_i$ plane. The three dimensional system shown in Figure 4-6 is reduced to just this plane and shown in Figure 4-7.

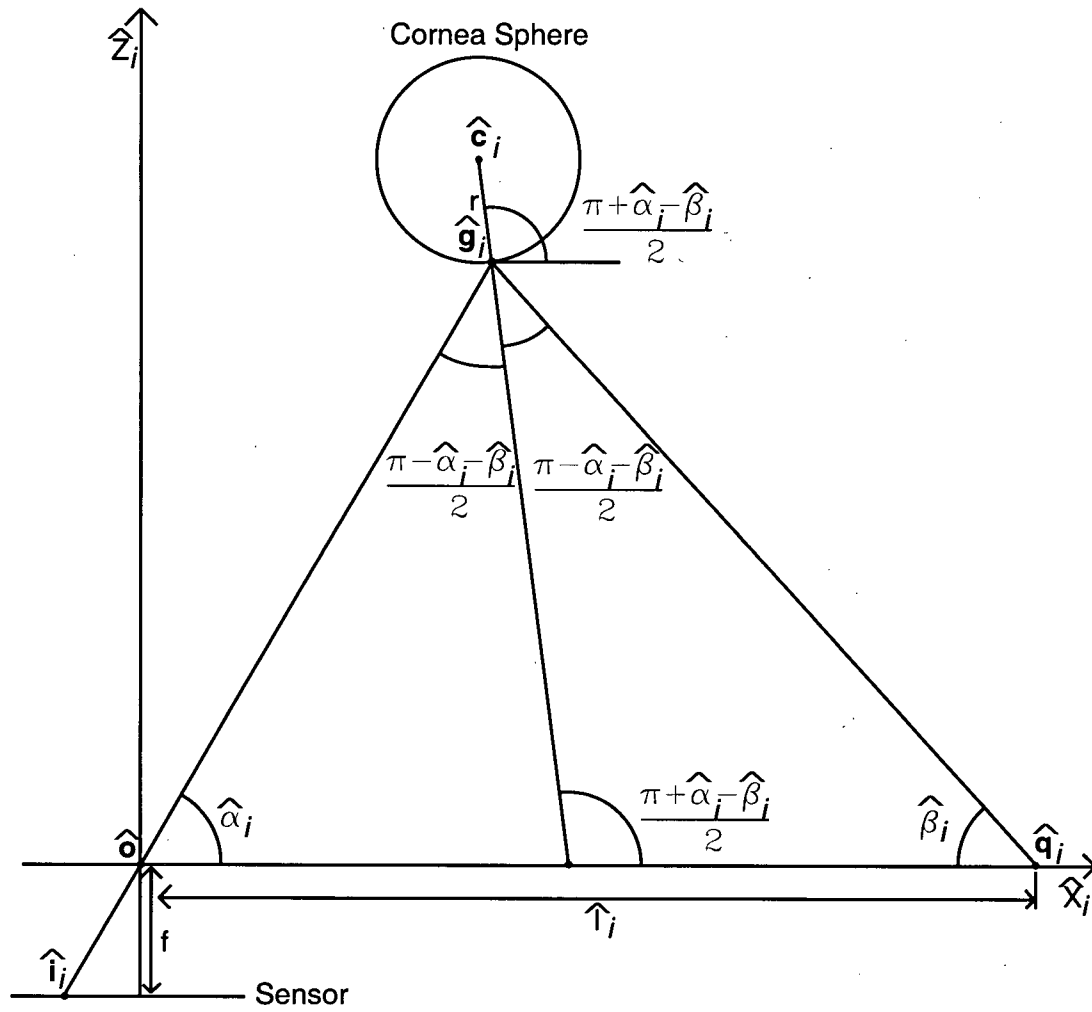


Figure 4-7 - 2D Auxiliary Coordinate System

- \hat{o} : Auxiliary coordinate system origin
- \hat{q}_i : Location of the glint point light source
- \hat{c}_i : Center of the cornea
- \hat{g}_i : Location of the glint
- \hat{i}_i : Image location of the glint on the CCD sensor
- f : Focal length distance
- r : Radius of cornea
- \hat{l}_i : Distance from \hat{o} to \hat{q}_i
- $\hat{\alpha}_i$: $\angle \hat{g}_i\hat{o}\hat{q}_i$
- $\hat{\beta}_i$: $\angle \hat{g}_i\hat{q}_i\hat{o}$

The values for $\hat{\mathbf{i}}_i$, $\hat{\mathbf{o}}$, $\hat{\mathbf{q}}_i$ and r are known for each of the auxiliary coordinate systems defined by the dual glints. The variable \hat{g}_{ix} remains an unknown; the remaining equations are defined with respect to this variable. The glint location \hat{g}_i is defined as:

$$\hat{g}_{ix} = \hat{g}_{ix} \quad (8)$$

$$\hat{g}_{iz} = \hat{g}_{ix} \tan(\hat{\alpha}_i) \quad (9)$$

Step 4 - Determine Symbolic Equations for Angles

The angles $\hat{\alpha}_i$ and $\hat{\beta}_i$ from Figure 4-6 are defined as (where \cdot is the dot product operator):

$$\hat{\alpha}_i = \cos^{-1} \left(\frac{(-\hat{\mathbf{I}}_i) \cdot (\hat{\mathbf{Q}}_i)}{\|-\hat{\mathbf{I}}_i\| \|\hat{\mathbf{Q}}_i\|} \right) \quad (10)$$

$$\hat{\beta}_i = \tan^{-1} \left(\frac{(\hat{g}_{ix}) (\tan(\hat{\alpha}_i))}{\hat{l}_i - \hat{g}_{ix}} \right) \quad (11)$$

The value for $\hat{\alpha}_i$ can be determined numerically while the equation for $\hat{\beta}_i$ is defined in terms of the unknown \hat{g}_{ix} .

Step 5 - Determine Symbolic Equations for Cornea Center in Auxiliary Coordinate System

The location of $\hat{\mathbf{c}}_i$ is defined in terms of the one unknown \hat{g}_{ix} as follows:

$$\hat{c}_{ix} = \hat{g}_{ix} - r \sin \left(\frac{\hat{\alpha}_i - \hat{\beta}_i}{2} \right) \quad (12)$$

$$\hat{c}_{iz} = \hat{c}_{ix} + r \cos \left(\frac{\hat{\alpha}_i - \hat{\beta}_i}{2} \right) \quad (13)$$

Step 6 - Convert Equations from Auxiliary to World Coordinate System

The rotation matrixes **R** and **S** defined in equations 6 and 7 are used to convert the equations for the cornea center in each of the two glint auxiliary coordinate systems as follows:

$$\mathbf{c}_1 = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} \hat{c}_{1x} \\ 0 \\ \hat{c}_{1z} \end{bmatrix} \quad (14)$$

$$\mathbf{c}_2 = \begin{bmatrix} s_{11} & s_{12} & s_{13} \\ s_{21} & s_{22} & s_{23} \\ s_{31} & s_{32} & s_{33} \end{bmatrix} \begin{bmatrix} \hat{c}_{2x} \\ 0 \\ \hat{c}_{2z} \end{bmatrix} \quad (15)$$

Step 7 - Numerically Solve for Cornea Center

The cornea center determined from each of the auxiliary coordinate systems should be located at exactly the same position in the world coordinate system resulting in the constraint:

$$\mathbf{c}_1 = \mathbf{c}_2 \quad (16)$$

Equations 14 and 15 can be multiplied out and collected to one side in equation 16 to obtain the following set of three equations, each of which contain two unknowns \hat{g}_{1x} and \hat{g}_{2x} :

$$(r_{11} \cdot \hat{c}_{1x} + r_{13} \cdot \hat{c}_{1z}) - (s_{11} \cdot \hat{c}_{2x} + s_{13} \cdot \hat{c}_{2z}) = 0 \quad (17)$$

$$(r_{21} \cdot \hat{c}_{1x} + r_{23} \cdot \hat{c}_{1z}) - (s_{21} \cdot \hat{c}_{2x} + s_{23} \cdot \hat{c}_{2z}) = 0 \quad (18)$$

$$(r_{31} \cdot \hat{c}_{1x} + r_{33} \cdot \hat{c}_{1z}) - (s_{31} \cdot \hat{c}_{2x} + s_{33} \cdot \hat{c}_{2z}) = 0 \quad (19)$$

Newton's method for solving an over defined set of non-linear equations [48] can be used to solve the three equations for the two unknowns \hat{g}_{1x} and \hat{g}_{2x} . To use the numerical method equations 17, 18 and 19 must be converted into a matrix representation as follows:

$$\begin{bmatrix} F_1 \\ F_2 \\ F_3 \end{bmatrix} = \begin{bmatrix} (r_{11} \cdot \hat{c}_{1x} + r_{13} \cdot \hat{c}_{1z}) - (s_{11} \cdot \hat{c}_{2x} + s_{13} \cdot \hat{c}_{2z}) \\ (r_{21} \cdot \hat{c}_{1x} + r_{23} \cdot \hat{c}_{1z}) - (s_{21} \cdot \hat{c}_{2x} + s_{23} \cdot \hat{c}_{2z}) \\ (r_{31} \cdot \hat{c}_{1x} + r_{33} \cdot \hat{c}_{1z}) - (s_{31} \cdot \hat{c}_{2x} + s_{33} \cdot \hat{c}_{2z}) \end{bmatrix} \quad (20)$$

The Jacobian matrix, or matrix of first partial differentials (F_1 , F_2 and F_3 differentiated with respect to \hat{g}_{1x} and \hat{g}_{2x}) is then defined as:

$$J = \begin{bmatrix} \frac{\partial F_1}{\partial \hat{g}_{1x}} & \frac{\partial F_1}{\partial \hat{g}_{2x}} \\ \frac{\partial F_2}{\partial \hat{g}_{1x}} & \frac{\partial F_2}{\partial \hat{g}_{2x}} \\ \frac{\partial F_3}{\partial \hat{g}_{1x}} & \frac{\partial F_3}{\partial \hat{g}_{2x}} \end{bmatrix} \quad (21)$$

Newton's method for computing the values of \hat{g}_{1x} and \hat{g}_{2x} is then as follows:

while (norm (K) > 0.01 & cnt < maxcnt)

cnt = cnt + 1;

F = compute_F (\hat{g}_{1x} , \hat{g}_{2x}); % as in equation 20

J = compute_J (F); % as in equation 21

K = [- F(1), - F(2), - F(3)]';

ΔG = inverse (J' * J) * (J' * K); % where ' is the transpose operator

\hat{g}_{1x} = \hat{g}_{1x} + ΔG (1);

\hat{g}_{2x} = \hat{g}_{2x} + ΔG (2);

end

cnt : Iteration counter

maxcnt : Maximum number of iterations to converge

F : Matrix of F_i equations

J : Matrix of F_i partial differential equations

K : Solution to F_i equations

ΔG : Change in \hat{g}_{1x} and \hat{g}_{2x}

\hat{g}_{1x} : Unknown variable from glint 1

\hat{g}_{2x} : Unknown variable from glint 2

Step 8 - Compute Cornea Center in World Coordinate System

Provided the algorithm in Step 7 above converges, either of equations 14 or 15 can be used along with the estimated value of \hat{g}_{1x} or \hat{g}_{2x} to compute the final value of c .

4.2. *Computing the Center of the Pupil*

To compute the **LOS** the center of the pupil along with the center of the cornea is required. The center of the real pupil is computed by ray tracing multiple points from the perimeter of the pupil image out from the camera sensor and into the eye. Computing the average of the real pupil perimeter points provides an estimate for the pupil center. Only two opposing pupil perimeter points are required to estimate the pupil center when the average is computed; however increasing the number of perimeter points reduces the susceptibility to noise. A total of 5 pupil perimeter points was chosen experimentally to provide robust pupil center estimates in a minimum of time. A flowchart of the algorithm for identifying the pupil center is shown in Figure 4-8. All calculations are performed in three dimensions in this algorithm.

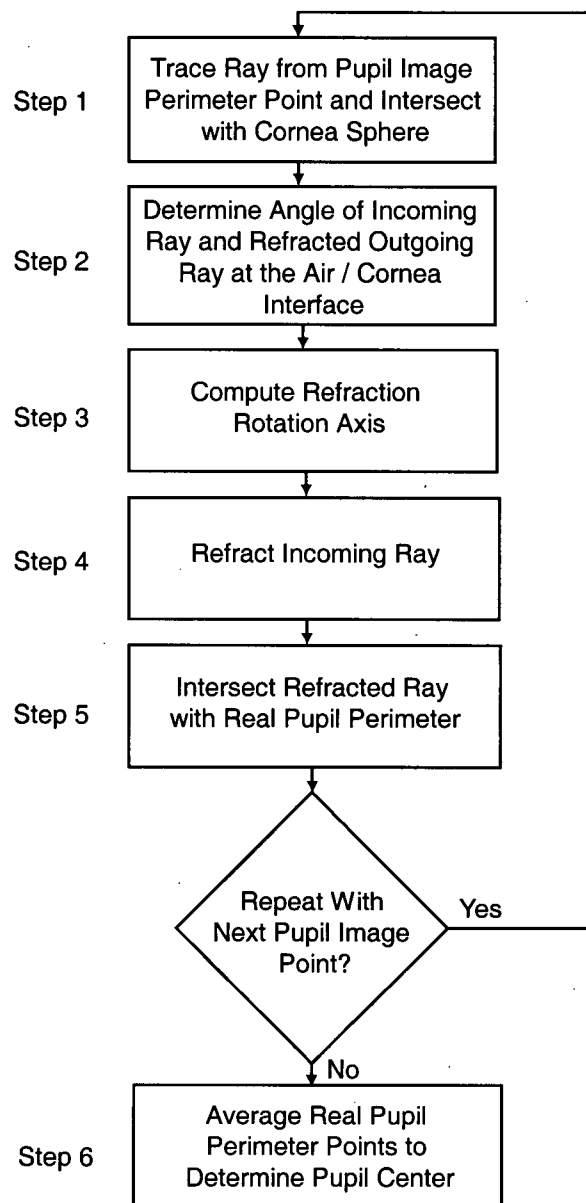


Figure 4-8 - Pupil Center Calculation Flowchart

Step 1 - Trace Pupil Perimeter Ray to Surface of Cornea Sphere

Using the equation of the ellipse that was fit to the pupil contour perimeter in the image acquisition section, a series of perimeter points i_i are chosen to be traced out from the image and into the eye. The diagram shown in Figure 4-9 illustrates two rays being traced from the sensor to the real pupil perimeter.

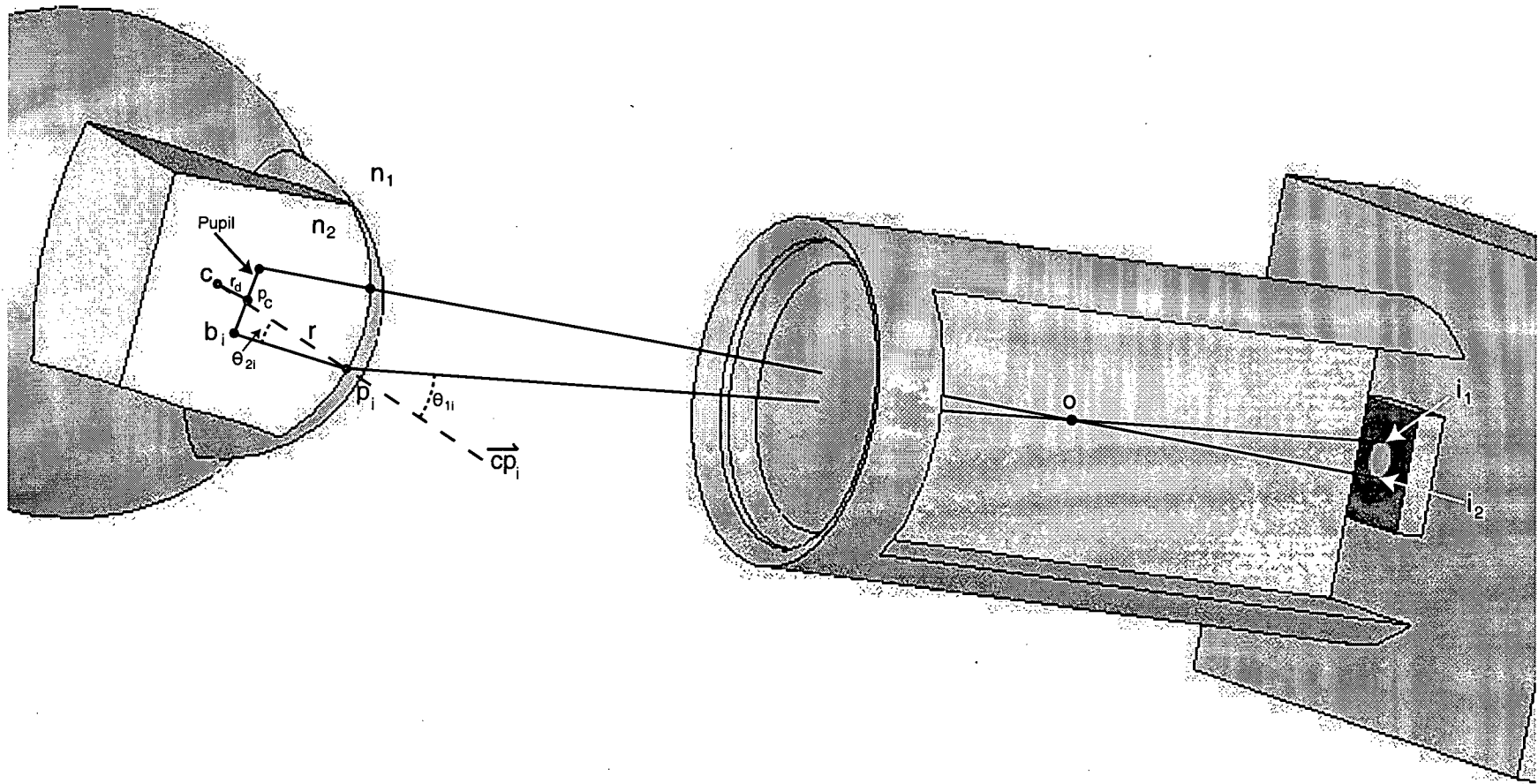


Figure 4-9 - Computing the Center of the Pupil

- o** : World coordinate system origin
- i_i** : Pupil perimeter image point on camera sensor
- p_i** : Pupil perimeter point on surface of cornea
- b_i** : Real pupil perimeter point inside the eye
- c** : Center of the cornea
- p_c** : Pupil Center
- r** : Radius of the cornea
- r_d** : Distance from center of cornea to center of pupil
- $\overrightarrow{cp_i}$: Vector normal to sphere **c** at point **p_i**
- θ_{1i}** : Incoming ray angle with respect to normal
- θ_{2i}** : Refracted ray angle with respect to normal
- n₁** : Index of refraction of the air
- n₂** : Index of refraction of the aqueous humor

The point **p_i** on the surface of the cornea is found by intersecting vector **-I_i** with sphere **c** of radius **r**. The point **p_i** is represented using a parametric equation of a line with one unknown parameter **t**:

$$\begin{bmatrix} p_{ix} \\ p_{iy} \\ p_{iz} \end{bmatrix} = \begin{bmatrix} i_{ix} \\ i_{iy} \\ i_{iz} \end{bmatrix} + t \begin{bmatrix} -I_{ix} \\ -I_{iy} \\ -I_{iz} \end{bmatrix} \quad (22)$$

The cornea is modeled as a sphere using the previously computed cornea center and population norm for the cornea radius:

$$(p_{ix} - c_x)^2 + (p_{iy} - c_y)^2 + (p_{iz} - c_z)^2 = r^2 \quad (23)$$

Equation 22 can be substituted into equation 23 and explicitly solved for the parameter **t**. The symbolic mathematical program Maple was used to solve for the parameter explicitly. The C++ implementation for determining the intersection point(s) of a line with a sphere is listed in Appendix G. After solving for **t** numerically, substitution back into equation 22 results in the pupil perimeter point **p_i** on the surface of cornea.

Step 2 - Compute Angles for Incoming and Refracted Rays

The angle the incoming ray makes with the normal vector to the sphere at point \mathbf{p}_i is calculated as follows (where \cdot is the dot product):

$$\theta_{1i} = \cos^{-1} \left(\frac{(\mathbf{p}_i - \mathbf{c}) \cdot (-\mathbf{P}_i)}{\|(\mathbf{p}_i - \mathbf{c})\| \|(-\mathbf{P}_i)\|} \right) \quad (24)$$

Refraction at the surface of the cornea is performed by rotating the normal vector $\overrightarrow{\mathbf{cp}_i}$ according to Snell's law of refraction. Snell's law states that when a ray transits from one medium to another with a different index of refraction, the following equation holds:

$$n_1 \sin(\theta_{1i}) = n_2 \sin(\theta_{2i}) \quad (25)$$

Rearranging equation 25 to solve for θ_{2i} and substituting in θ_{1i} results in:

$$\theta_{2i} = \frac{\sin^{-1} \left(n_1 \sin \left(\cos^{-1} \left(\frac{(\mathbf{p}_i - \mathbf{c}) \cdot (-\mathbf{P}_i)}{\|(\mathbf{p}_i - \mathbf{c})\| \|(-\mathbf{P}_i)\|} \right) \right) \right)}{n_2} \quad (26)$$

Step 3 - Compute Refraction Rotation Axis

To refract the incoming ray at the air / cornea interface a rotation axis is required. The rotation axis is denoted as the vector \mathbf{K} , which is the vector orthogonal to the normal $\overrightarrow{\mathbf{cp}_i}$ and vector $-\mathbf{P}_i$:

$$\mathbf{K} = \frac{\overrightarrow{\mathbf{cp}_i} \times -\mathbf{P}_i}{\|\overrightarrow{\mathbf{cp}_i} \times -\mathbf{P}_i\|} \quad (27)$$

Step 4 - Compute Refracted Ray

To compute the refracted ray an equivalent angle rotation [49] is performed in 3D on the vector $-\overrightarrow{\mathbf{cp}_i}$ about the vector \mathbf{K} by the angle θ_{2i} . The rotation matrix required is computed as follows:

$$\mathbf{R}_\mathbf{K} = \begin{bmatrix} K_x K_x (1 - \cos(\theta_{2i})) + \cos(\theta_{2i}) & K_x K_y (1 - \cos(\theta_{2i})) - K_z \sin(\theta_{2i}) & K_x K_z (1 - \cos(\theta_{2i})) + K_y \sin(\theta_{2i}) \\ K_x K_y (1 - \cos(\theta_{2i})) + K_z \sin(\theta_{2i}) & K_y K_y (1 - \cos(\theta_{2i})) + \cos(\theta_{2i}) & K_y K_z (1 - \cos(\theta_{2i})) - K_x \sin(\theta_{2i}) \\ K_x K_z (1 - \cos(\theta_{2i})) - K_y \sin(\theta_{2i}) & K_y K_z (1 - \cos(\theta_{2i})) + K_x \sin(\theta_{2i}) & K_z K_z (1 - \cos(\theta_{2i})) + \cos(\theta_{2i}) \end{bmatrix} \quad (28)$$

The equivalent angle rotation applied to the vector $-\overrightarrow{cp_i}$ results in the refracted vector $\overrightarrow{p_i b_i}$ as follows:

$$\overrightarrow{p_i b_i} = R_k(c - p_i) + p_i \quad (29)$$

Step 5 - Intersect Refracted Ray with Real Pupil Perimeter

To determine the real pupil perimeter point b_i the refracted vector $\overrightarrow{p_i b_i}$ is traced further into the eye. Each point on the real pupil perimeter is the same distance r_{ps} from the center of the cornea as shown in Figure 4-10.

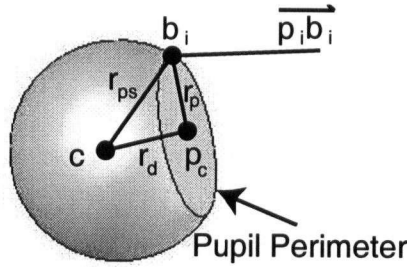


Figure 4-10 - Pupil Sphere Intersection

c : Center of the cornea

$\overrightarrow{p_i b_i}$: Refracted incoming ray

b_i : Real pupil perimeter point inside the eye

r_d : Distance from center of cornea to pupil

r_p : Radius of pupil

r_{ps} : Radius of truncated pupil sphere

A geometric solution for solving for the point b_i exists as the constraint

$$\|b_i - c\| = r_{ps} \quad (30)$$

results in a spherical shell of possible solutions for b_i . Finding the intersection point of the vector $\overrightarrow{p_i b_i}$ with the sphere centered at c with radius r_{ps} is easily performed using same line / sphere intersection method described in step 1 of the cornea center computation algorithm.

The distance r_{ps} is determined from the distance from the center of cornea to the center of the pupil r_d and the radius of the pupil r_p by the following equation:

$$r_{ps} = \sqrt{r_d^2 + r_p^2} \quad (31)$$

The distance r_d comes from the population norms compiled by Gullstrand while the pupil radius r_p is calculated using equation 1 from the camera model and the dimensions of the image of the pupil.

Step 6 - Average Pupil Perimeter Points to Obtain Pupil Center

The pupil center p_c is found by computing the average of all the perimeter points traced into the eye:

$$p_c = \frac{\sum_{i=1..5} b_i}{5} \quad (32)$$

4.3. Computing the POG

As mentioned in the introduction, the **POG** is determined by intersecting the vector formed from c to p_c (the **LOS**) with the surface of the monitor screen. The plane equation used to describe the surface of the screen are derived from location measurements of three of the screen corners. An outline of the system is shown in Figure 4-11.

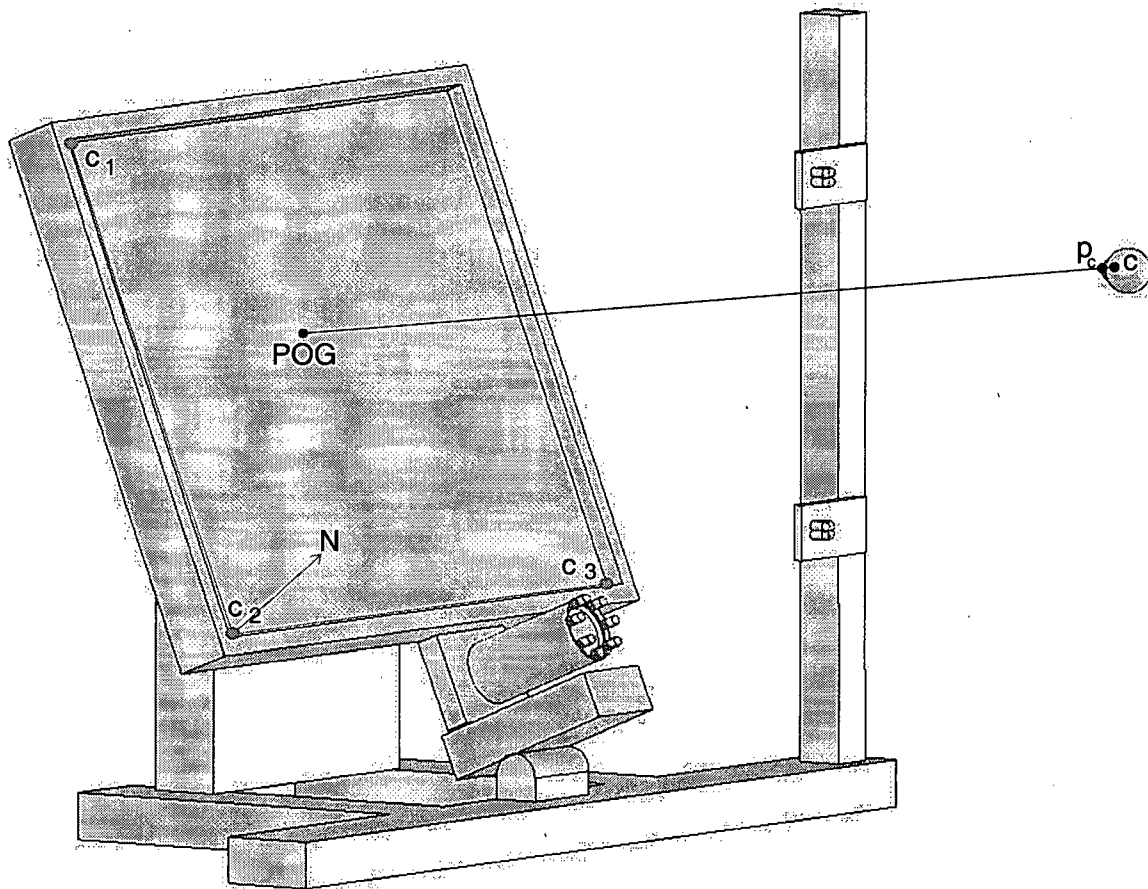


Figure 4-11 - Computing the POG

- c_1 : Monitor corner 1
- c_2 : Monitor corner 2
- c_3 : Monitor corner 3
- N : Normal vector to monitor plane
- c : Center of cornea
- p_c : Center of pupil
- POG** : Point of gaze on the computer screen

The vector normal to the monitor plane $N = [a \ b \ c]$ is computed by the following equation:

$$N = (c_1 - c_2) \times (c_3 - c_2) \quad (33)$$

The equation of a plane that defines the monitor plane is then as follows:

$$a(x - c_{1x}) + b(y - c_{1y}) + c(z - c_{1z}) = 0 \quad (34)$$

The **POG** can be described using a parametric equation of a line with one parameter t :

$$\begin{bmatrix} POG_x \\ POG_y \\ POG_z \end{bmatrix} = \begin{bmatrix} c_x \\ c_y \\ c_z \end{bmatrix} + t \begin{bmatrix} p_{cx} - c_x \\ p_{cy} - c_y \\ p_{cz} - c_z \end{bmatrix} \quad (35)$$

Substituting equation 35 into equation 34 results in one equation with one unknown, t :

$$a(POG_x - c_{1x}) + b(POG_y - c_{1y}) + c(POG_z - c_{1z}) = 0 \quad (36)$$

Maple was used to solve for the parameter t in equation 36 explicitly. The C++ implementation for determining the intersection point of a line with a plane is given in Appendix H. Once the value for t is known it can be substituted back into equation 35 to determine the **POG**.

4.4. **POG Correction**

A number of simplifications and possible sources of error in the algorithms above may result in inaccuracies in the computed POG. Some possible sources of error include the measurements of the physical glint light source locations and screen corners, the camera calibration parameters and the variations in eye model parameters for different subjects.

To correct for these sources of error a simple per user calibration is performed. A calibration is required for each pair of the dual glint light sources that are to be used. For users who only require one active pair of glint light sources such as those without eye-glasses, a single calibration is sufficient. For those who require the eye-glasses reflection compensation method enabled, in which the active dual glint light sources may change, a calibration is required for each set of alternative lights.

The calibration procedure involves recording the computed POG while the user stares at known reference points on the screen. Through experimentation it was found that using the four corners of the monitor as the reference points provides a sufficiently

accurate correction with minimal user interaction. An example of the 4 calibration points m_1 to m_4 , and one test point m are shown in Figure 4-12.

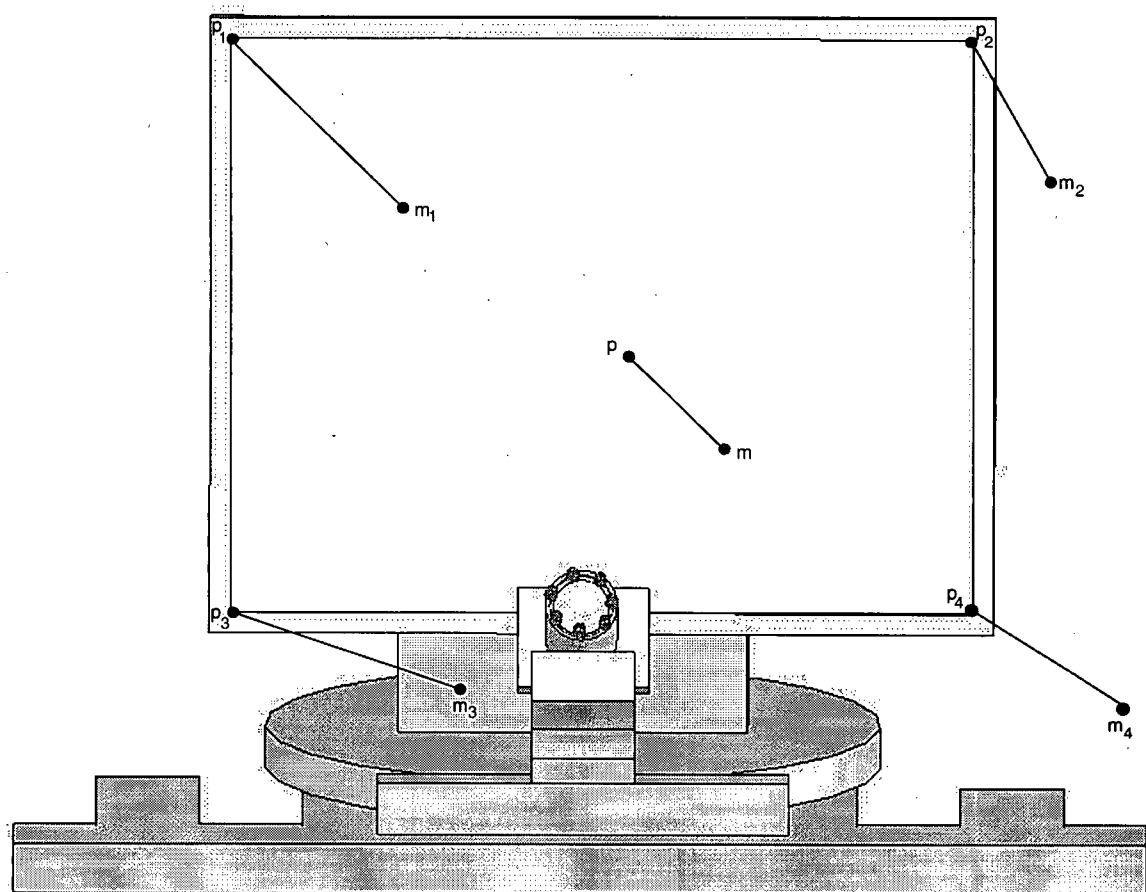


Figure 4-12 - Calibration Example

$m_1..m_4$: Compute POG for the four corners

$p_1..p_4$: Reference points

m : Example computed POG

p : Example corrected POG

The algorithm developed for correcting the computed POG M uses weighted correction factors from each of the calibration points. The correction algorithm flowchart is shown in Figure 4-13.

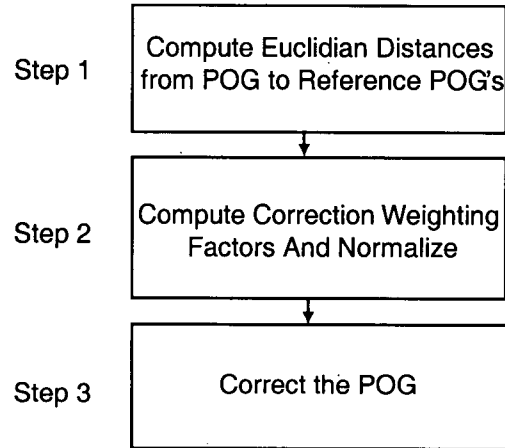


Figure 4-13 - POG Correction Algorithm Flowchart

Step 1 - Compute Euclidian Distances from POG to Reference POG's

For each calibration point a correction factor, which is the difference between the reference point \mathbf{p}_i and the calibration point \mathbf{m}_i , is weighted and then applied to the POG to correct. The distance from each calibration point \mathbf{m}_i to the point to be corrected \mathbf{m} is used as the weighting factor. For example, if the point to be corrected lies directly atop calibration point \mathbf{m}_i then 100% of the correction used to convert \mathbf{m}_i to \mathbf{p}_i is used. If the point \mathbf{m} lies exactly between each of the calibration points then 25% of each correction factor is used. To compute the distance from \mathbf{m} to each \mathbf{m}_i the following operation is performed, with $i = 1$ to 4 reference points:

$$dist_i = \|\mathbf{m} - \mathbf{m}_i\| + 1 \quad (37)$$

A value of 1 is added to each $dist_i$ to prevent a possible divide by zero in the following steps.

Step 2 - Compute Weighting Factors And Normalize

The weighting factors are computed depending on the relative distances computed in the previous step:

$$x_i = \frac{\sum_{i=1..n} dist_i}{dist_i} \quad (38)$$

The weighting factors are then normalized to percentages:

$$w_i = \frac{x_i}{\sum_{i=1..n} x_i} \quad (39)$$

Step 3 - Correct the POG

Each weighting factor is multiplied by the correction factor for each of the reference points and added to the POG to correct.

$$p = \left(\sum_{i=1..n} w_i (m_i - p_i) \right) + m \quad (40)$$

Finally the POG is averaged over time using a moving window average to reduce random jitter.

5. System Testing and Results

Evaluation of the eye-gaze tracking system we developed was performed to determine the effectiveness of our design decisions in overcoming the limitations found in many of the leading eye-gaze tracking systems. This chapter outlines the details of our accuracy metric, the tests performed and a discussion of the results of each test. The POG plots for each of the tests are shown in Appendix I. A comparison of how well our system performs with respect to the other leading systems is provided in the following chapter.

5.1. Accuracy Metrics

Eye-gaze tracking system accuracy is typically reported in terms of degrees of visual angle, while the actual error is usually measured in terms of pixel error from the reference POG to the estimated POG. Degrees of visual angle are preferred to pixels as the accuracy metric as visual angle is independent of screen size and resolution. In this thesis, accuracy will be reported in pixel error in the results section and degrees of visual angle when comparing our system with other eye-gaze tracking systems.

To convert from screen pixel error to degrees of visual angle the distance from the eye to the screen is required as shown in Figure 5-1. This distance is not known for our system as we do not fix the location of the user's head. To convert from pixel error to visual angle the most conservative distance of 60 cm is used, which is the closest distance the users face can come to the screen before going out of focus.

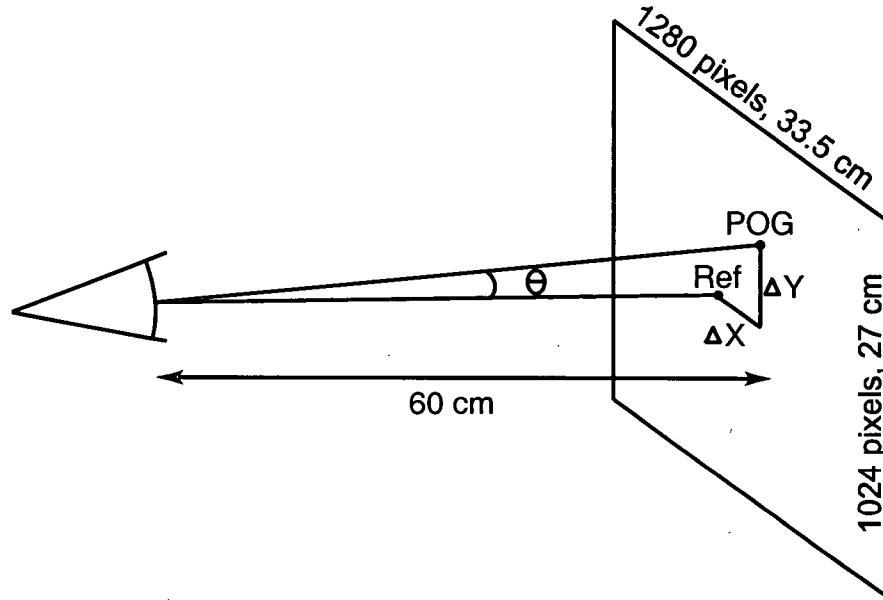


Figure 5-1 - Pixel Error to Visual Angle Error

POG : Estimated point of gaze on screen

Ref : Reference point on screen

ΔX : X error

ΔY : Y error

θ : Visual angle error

To convert from screen pixels to centimeters measurements of the screen were made, resulting in the following conversions:

$$\Delta X_{cm} = \Delta X_{pixel} \frac{33.5cm}{1280pixels} \quad (41)$$

$$\Delta Y_{cm} = \Delta Y_{pixel} \frac{27.0cm}{1024pixels} \quad (42)$$

By trigonometry we can then convert to degrees of visual angle:

$$\theta = 2 \tan^{-1} \left(\frac{\sqrt{\Delta X_{cm}^2 + \Delta Y_{cm}^2}}{60 \text{ cm}} \right) \quad (43)$$

5.2. Eye Tracking

Real-time performance is critical for an eye-gaze tracking based device such as ours. To reduce the computational load of the image processing algorithms a series of ROIs are used to reduce the required size of the images to process. The processing time required to perform a full system loop on a full sized image is 110 ms. When the ROI has locked onto the eye the processing time reduces to 28 ms.

To avoid losing the ROI lock on the eye in the event of an eye blink, 3 frames are processed before finally giving up on the ROI location and reprocessing the entire image. If the eye was lost due to an eye blink, it will return to the current ROI location. If the eye was lost due to rapid translation, it will take 3 frames at 15 Hz before the system gives up on the current ROI location, and 1 full sized image frame to reacquire the eye, for a total of $3 * (1 / 15 \text{ Hz}) + 110 \text{ ms}$, or 308 ms.

The size of the ROI limits the allowable head motion speed. A larger ROI permits larger eye motion without losing lock but increases the processing requirement for the larger number of pixels. The eye tracking ROI is 200 pixels wide by 175 pixels high and is re-centered on the pupil each system loop. An ROI of this size allows the image of the pupil to move up to 100 pixels left or right and 87.5 pixels up or down between each frame. To maintain the eye within the ROI it may move up to a distance of 0.87 cm horizontally and 0.76 cm vertically when the eye is 60 cm from the camera (using the pinhole camera model and equation 1 to convert from pixels to centimeters). With a frame rate of 15 Hz the allowable head motion speed while still maintaining the ROI lock is 13.2 cm/s horizontally and 11.5 cm/s vertically. The natural head movements observed during system testing were much less than this when the subject is looking at objects on the screen.

5.3. Feature Extraction

The pupil and glint feature extraction algorithms developed in Chapter 3 were tested on real images as part of the whole system during development. To evaluate the performance of just the feature extraction algorithms a quantitative test is required. A set of synthetic images were created by digitally manipulating copies of real images in which the real pupil and glints were removed and then redrawn with ellipse drawing functions

using known ellipse centers. The difference between the known pupil and glint ellipse centers and the extracted ellipse centers gives a measure of the algorithm's performance.

Two sets of bright and dark pupil images were created to test the algorithms. One set of images had the pupil glint drawn close to the pupil perimeter and the other set of images had the pupil glint drawn external to the pupil contour. The original and synthetic images are for each set of images are shown in Figure 5-2 and Figure 5-3. The locations of the ellipses drawn on the synthetic images and the values of the extracted ellipses for each of the two data sets are shown in Table 5-1.

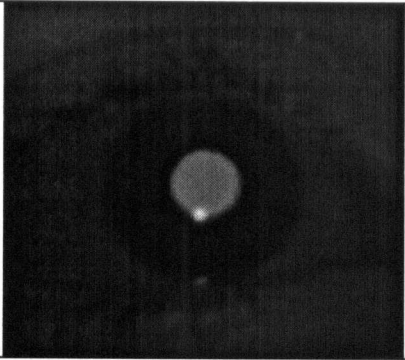
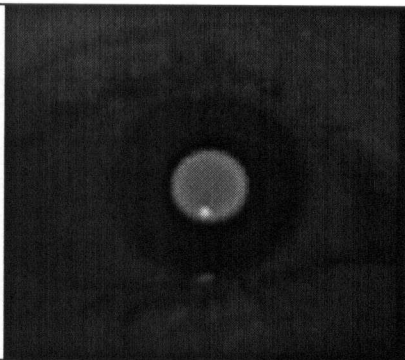
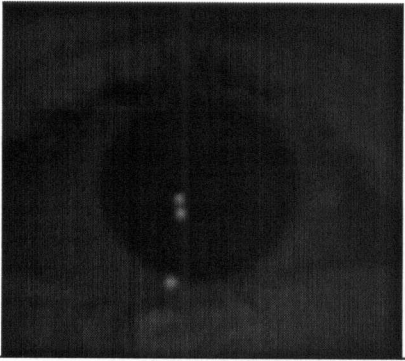
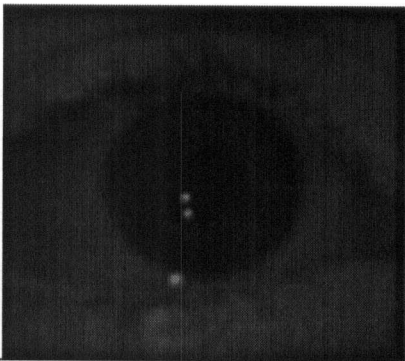
Bright Pupil Image		
Dark Pupil Image		
	Original Image	Synthetic Image

Figure 5-2 – Image Test Set 1 for Feature Extraction Testing

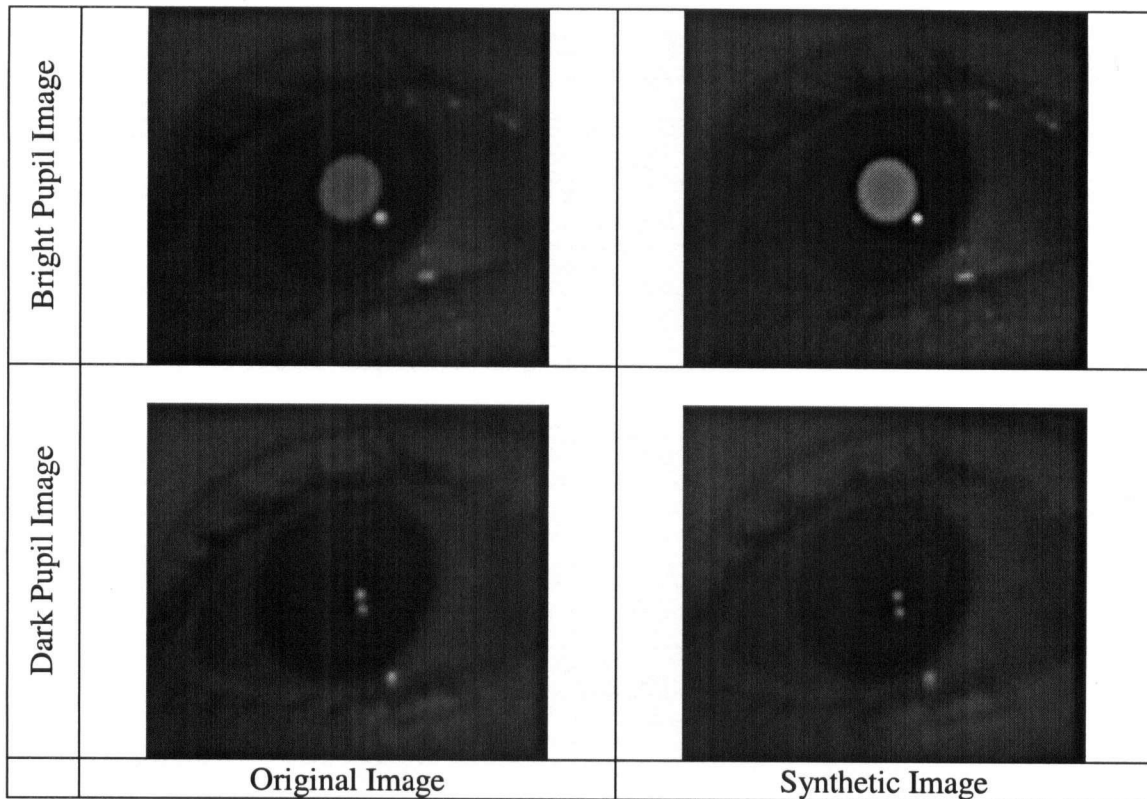


Figure 5-3 - Image Test Set 2 for Feature Extraction Testing

Table 5-1- Results of Feature Extraction Test on Synthetic Images

		Identified Pupil Center (pixels)		Identified Pupil Glint Center (pixels)		Identified Glint 1 Center (pixels)		Identified Glint 2 Center (pixels)	
Coordinate		X	Y	X	Y	X	Y	X	Y
Set 1	Actual Locations	649.00	450.00	647.00	463.00	638.00	456.00	639.00	464.00
	Extracted Locations	649.03	450.01	647.00	463.00	638.00	456.00	639.00	464.00
	Difference	0.03	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Set 2	Actual Locations	598.00	441.00	613.00	455.00	604.00	447.00	605.00	455.00
	Extracted Locations	598.05	441.05	613.00	455.00	604.00	447.00	605.00	455.16
	Difference	0.05	0.05	0.00	0.00	0.00	0.00	0.00	0.16

The feature extraction system identified the centers of the generated pupil and glint ellipses in the static images very well, to within 0.16 pixels in the worst case. Dynamic effects such as motion blur would likely reduce the methods accuracy.

5.4. Feature Extraction Sensitivity

The sensitivity of the estimated POG to errors in the extracted image feature locations was tested by first determining the POG for a given set of inputs. The inputs to the POG estimation system are the X and Y coordinates of the dual glint image centers and the ellipse equation for the pupil perimeter (ellipse center X and Y, major and minor axis length, and rotation angle). Each of the inputs to the POG estimation algorithm were individually varied by ± 1 and ± 2 pixels, the POG estimated again, and the change in POG recorded. The difference between the correct POG (900, 417) and the POG calculated with the incorrect image feature locations are tabulated in Table 5-2. Note that in the case of the angle of rotation of the pupil ellipse the value was varied in units of radians.

Table 5-2- Feature Extraction Sensitivity Differences

Input	Input - 2		Input - 1		Input + 1		Input + 2	
	POG X error (pixels)	POG Y error (pixels)	POG X error (pixels)	POG Y error (pixels)	POG X error (pixels)	POG Y error (pixels)	POG X error (pixels)	POG Y error (pixels)
g1x	117	-1	60	4	-61	-12	Fail	Fail
g1y	64	44	28	19	-21	-14	-39	-26
g2x	Fail	Fail	-4	-13	2	5	0	0
g2y	-40	-125	-22	-70	28	88	63	202
px	-117	3	-58	1	58	-1	116	-1
py	1	122	1	61	0	-61	0	-122
major	2	-5	1	-2	0	3	-1	6
minor	2	-4	1	-2	-1	3	-2	5
phi	0	0	0	0	0	0	0	0

g1x : Glint 1 center X coordinate (pixels)

g1y : Glint 1 center Y coordinate (pixels)

g2x : Glint 2 center X coordinate (pixels)

g2y : Glint 2 center Y coordinate (pixels)

px : Pupil center X coordinate (pixels)

py : Pupil center Y coordinate (pixels)

major : Length of pupil ellipse major axis (pixels)

minor : Length of pupil ellipse minor axis (pixels)

ϕ : Rotation angle of pupil ellipse (radians)

Varying the dual glint X co-ordinates by +2 pixels for glint 1 and -2 pixels for glint 2 resulted in the POG estimation method failing to converge. A correlation appears to exist between the image X coordinate error and the POG X coordinate error and the image Y coordinate error and the POG Y coordinate error. The most sensitive parameters to variation appear to be the X coordinate of glint 1, the Y coordinate of glint 2, and both the X and Y coordinates of the pupil ellipse center. Variation of the major and minor axis lengths, as well as the pupil ellipse rotation angle ϕ , did not significantly change the estimated POG. To get an idea of these errors in terms of visual angle the conversion factor is 40 pixels / ° of visual angle.

5.5. Single User Accuracy Testing

Single user testing was performed to test the system on a number of different issues with a consistent user. A chin rest was used for the single user tests to position the head at the same location for each test and to prevent the possibility of fatigued muscles from affecting the results. The use of a headrest helped to isolate the particular issue under test but does not affect the accuracy of the testing as indicated by the results of Section 5.5.1. The level of ambient IR lighting was low for all of the tests except the ambient lighting tests.

Single user testing involves recording the POG for a number of estimations at each point on a 4x4 grid drawn on the computer screen. Once the subject has completed looking at the grid they are asked to stare at a single point in the center of the screen while the system records a large number of samples collected at the full system frame rate. The grid provides a measure of accuracy over the entire screen, while the single point provides an indication of the standard deviation of the computed POG.

The analysis of variance (ANOVA) statistical method was used to evaluate differences in average error between tests. The ANOVA was performed on the X and Y pixel average errors with a critical alpha (statistical significance cut off value) of 0.05 for all tests. The result of an ANOVA test is reported as $F(df_between, df_within) = F_result, p = p_result$. The $df_between$ variable is the number of groups tested minus one (e.g. for a 2 group test $df_between$ would equal $2 - 1 = 1$). The value of df_within can be

used to determine the total number of samples by adding $df_{\text{between}} + df_{\text{within}} + 1$ (e.g. $F(1,80)$ means there were $1 + 80 + 1 = 82$ samples in the study). F_{result} is the resulting value of the F statistic and p_{result} is the final value that is compared with alpha. If $p_{\text{result}} > \alpha$ then no statistically significant difference was found.

For tests comparing 3 or more means a significant result using the ANOVA test does not specify which of the means is significantly different. The Bonferroni post-hoc test can be used to determine which of the means are significantly different.

5.5.1. With and Without Headrest

The first of the single user tests is to determine the effect of the use of the headrest. The headrest is used for the remaining single user tests to provide a consistent head position, and to ensure that possible muscle fatigue will not affect the results.

Table 5-3 – With and Without Headrest

		Without Headrest (pixels)		With Headrest (pixels)	
		X	Y	X	Y
4x4 Grid	Average Error	12.41	12.52	12.52	17.34
	Standard Deviation	10.33	9.95	9.95	14.76
	Maximum Error	9.5	11.59	11.59	44.39
	Maximum Error Loc	421	1263	1263	969
Single Point	Average Error	12.86	19.77	19.77	21.91
	Standard Deviation	9.81	13.66	13.66	7.78

No statistically significant difference was found for the X coordinate ($F(1, 151) = 0.004$, $p = .949$) and for the Y coordinate ($F(1, 151) = 0.019$, $p = 0.890$).

5.5.2. Shutter speed

A small lens aperture is required to maintain a large depth of focus (DOF) but a longer shutter time is then required to capture enough incoming light to fully expose the images. A long shutter time may increase the effect of motion blur on the recorded images when the head is in motion. The following testing was performed at four different speeds (from a maximum shutter duration of 66.66 ms to as fast as 8.33 ms) to determine

the effect of shutter speed on system accuracy. Shutter durations faster than 8.33 ms were not tested as the aperture required to expose these images resulted in a DOF that was not sufficiently large.

Table 5-4 – Varying Shutter Speeds

		8.33 ms (pixels)		25.00 ms (pixels)		58.33 ms (pixels)		66.66 ms (pixels)	
		X	Y	X	Y	X	Y	X	Y
4x4 Grid	Average Error	13.86	10.58	14.16	11.12	19.46	12.88	14.68	12.64
	Standard Deviation	13.28	8.72	12.33	10.69	16.91	10.31	13.4	11.27
	Maximum Error	37.75	8.5	35.14	6.42	42.14	35	33.33	35.33
	Maximum Error Loc	421	323	421	969	842	0	842	646
Single Point	Average Error	24.91	5.77	11.45	7.02	25.35	10.67	12.55	8.24
	Standard Deviation	9.68	4.85	8.67	5.3	15.3	7.61	10.17	5.71

Statistical analysis of the differences in the average errors for the 4 different shutter speeds resulted in a statistically significant difference for X ($F(3, 344)=3.056$, $p = 0.028$) and no significant difference in Y ($F(3, 344) = 0.968$, $p = 0.408$). A Bonferroni post-hoc analysis of the average X error data indicated that the X pixel error for the 58.33 ms shutter time was slightly larger than the rest of the average X errors. While the difference in X error is statistically significant, we did not feel it was part of a trend given the 66.66 ms average X error was not significantly different than for the 8.33 and 24 ms shutter speeds. As a result of this investigation we determined that we could use the longest possible shutter time of 66.66 ms (or 1 / 15 fps) without loss of accuracy.

5.5.3. Sunlight Compensation

The operation of the system is robust to most ambient lighting conditions except for when direct sunlight is present. Direct sunlight contains far more IR light than our lighting system can produce. When sunlight shines through a nearby window and strikes the subject's face, the resulting images are almost completely washed out

In an attempt to solve this problem, we tried to increase the proportion of system-generated light to sunlight reaching the camera. To reduce the amount of sunlight we replaced the long pass SFX filter with a narrow band interference filter from Edmund Optics, part number NT43-098. The narrow band filter was matched as closely as

possible to our LED system lighting frequency (875 nm). The filter is centered at 880 nm and has a full width-half maximum (FWHM) of 10 nm as shown in Figure 5-4.

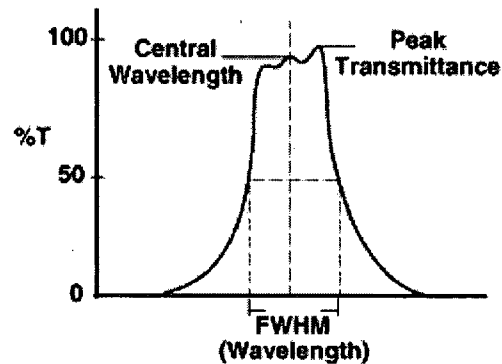


Figure 5-4 - Narrowband Filter¹

A set of four test images were performed under the conditions listed in Table 5-5 with the corresponding captured images shown in Figure 5-5.

Table 5-5 - Sunlight Compensation Test Conditions

Test Case	Test conditions
A)	<ul style="list-style-type: none"> • No direct sunlight • Interference filter • Moderate aperture size
B)	<ul style="list-style-type: none"> • Direct sunlight • Interference filter • Moderate aperture size
C)	<ul style="list-style-type: none"> • Direct sunlight • Interference filter • Smaller aperture size than A and B
D)	<ul style="list-style-type: none"> • Direct sunlight • Long pass filter • Smaller aperture size than C

¹ Figure adapted from:

<http://www.edmundoptics.com/onlinecatalog/DisplayProduct.cfm?productid=1903>



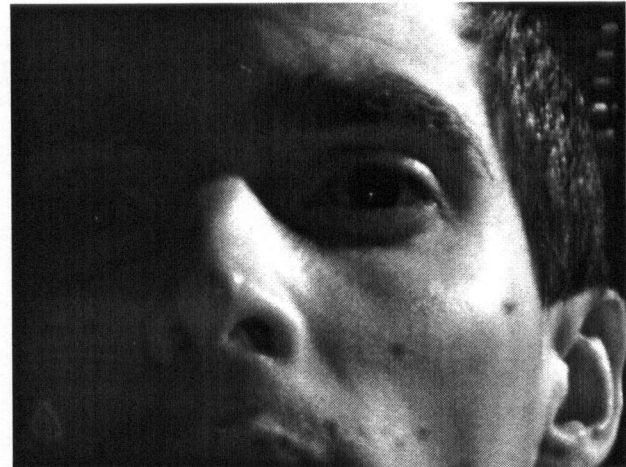
A



B



C



D

Figure 5-5 – Sunlight Compensation Test Images

The bright pupil is clearly seen in *A* when direct sunlight is not striking the face. In *B* the bright pupil is still visible though only slightly as is the pupil in *C* with a reduced aperture to prevent the saturation seen in *B*. When the long pass filter is used in image *D* and the aperture reduced significantly to prevent saturation, the bright pupil cannot be seen at all and the glint is quite faint.

The narrowband filter appears to block greater amounts of sunlight while passing a larger percentage of system generated light. However, the amount of sunlight blocked is still insufficient to allow the eye-gaze tracking system to operate in the presence of direct sunlight.

5.5.4. IR LED Wavelength

The sensitivity of the camera sensor is lower at longer wavelengths, as shown in Figure 2-5. In addition to light with a wavelength of 875 nm, the most commonly used in eye tracking systems, Epitex L760-30-AU diodes with a wavelength of 760 nm were tried. Using lower wavelength diodes should increase the relative response of the camera allowing a smaller lens aperture for an equivalent intensity image and consequently a greater depth of field. Unfortunately we discovered that light generated at 760 nm is still slightly visible to the human eye. The IR LED wavelength test is performed to determine if any difference in accuracy exist due to the different wavelengths.

Table 5-6 – Different LED Wavelengths

		760 nm (pixels)		875 nm (pixels)	
		X	Y	X	Y
4x4 Grid	Average Error	14.34	13.35	13.94	14.05
	Standard Deviation	16.66	10.17	13.26	13.1
	Maximum Error	50.57	11.57	29.16	49.5
	Maximum Error Loc	421	0	421	0
Single Point	Average Error	10.46	17.05	22.27	20.87
	Standard Deviation	6.62	10.33	11.5	7.54

A statistical comparison of the average errors for the two cases indicated that there was no statistically significant difference between the average errors for X ($F(1, 181) = 0.033$, $p = 0.857$) and Y ($F(1, 181) = 0.164$, $p = 0.686$). These results indicate that a lower wavelength could be used without loss in system accuracy, provided a wavelength could be found that was still invisible to humans.

5.5.5. Eye-Glasses

A test was performed to determine if there was any difference in accuracy with and without eye-glasses. The test user does not require corrective lenses but for this test used a pair of non-prescription reading glasses without an anti-glare coating. It was found that the anti-glare coating on some eye-glasses reduces the transmission of light and increases the difficulty of image feature extraction. The algorithm to prevent eye-glass

reflections from corrupting the data was enabled for the test in which eye-glasses were used.

Table 5-7 – Eye Glass Reflection Compensation Routine

		<i>No Eye-glasses No Correction (pixels)</i>		<i>Eye-glasses Correction (pixels)</i>	
		X	Y	X	Y
4x4 Grid	Average Error	17.03	9.95	12.57	8.66
	Standard Deviation	12.98	9.97	10.65	9.37
	Maximum Error	36.57	0	26.75	4.75
	Maximum Error Loc	842	0	421	323
Single Point	Average Error	11.6	7.94	17.45	6.36
	Standard Deviation	8.9	4.95	11.96	4.03

Statistical analysis of the average errors indicated that there was a statistically significant difference in X ($F(1, 269) = 9.446$, $p = 0.002$) and no significant difference in Y ($F(1, 269) = 1.185$, $p = 0.277$). The situation in which eye-glasses were worn actually resulted in better accuracy which is an unexpected result, though the difference in average error was relatively small (4.46 pixels).

5.5.6. Ambient IR Levels

System accuracy was measured with three levels of ambient IR light to roughly determine the susceptibility of the system to ambient lighting conditions. To more precisely define the ability to handle ambient IR light a measure of the actual ambient IR level should be recorded for each test. Unfortunately no such measurement device was available for our testing. Testing took place next to large 6' tall windows that were covered with a perforated mesh curtain. The test with no sunlight was performed in the evening, cloudy day testing was performed on a day with thick cloud cover and the indirect sunlight test was performed on a sunny day with sunlight in the laboratory, but no direct sunlight striking the subject's face.

Table 5-8 – Ambient IR Levels

		No Sunlight (pixels)		Cloudy Day (pixels)		Indirect Sunlight (pixels)	
		X	Y	X	Y	X	Y
4x4 Grid	Average Error	12.41	17.01	15.22	16.52	13.96	15.71
	Standard Deviation	10.33	14.98	11.87	11.51	9.15	14.53
	Maximum Error	9.5	67.5	29	20	10.79	51.39
	Maximum Error Loc	421	0	842	0	1263	646
Single Point	Average Error	12.86	24.31	30.68	19.29	19.73	13.51
	Standard Deviation	9.81	4.35	13.45	8.18	14.32	13.66

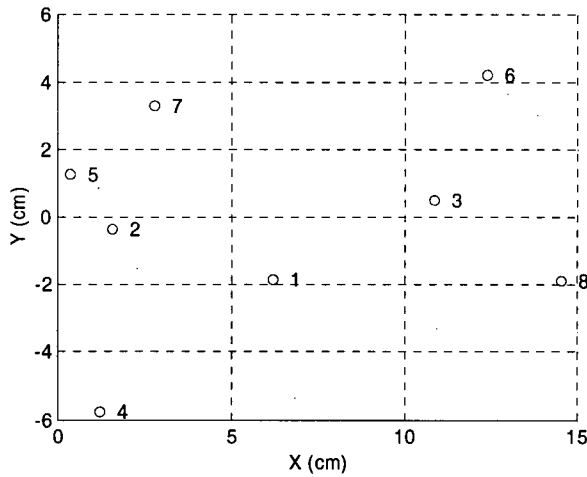
A statistical analysis of the average errors indicates there is no statistically significant difference in X ($F(2, 301) = 1.791$, $p = 0.169$) or in Y ($F(2, 301) = 0.198$), $p = 0.829$). The system appears to be robust to varying levels of ambient sunlight although as mentioned previously, cannot operate when direct sunlight strikes the subjects face.

5.5.7. Head Locations

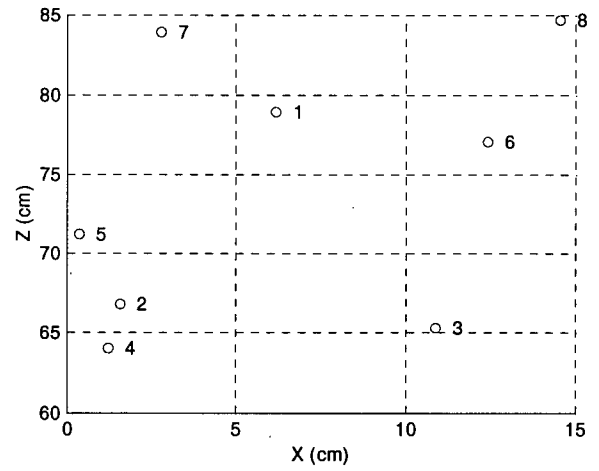
To test the ability of the system to handle different head positions with a single calibration the system accuracy was measured with the head in 8 different locations. A single calibration was performed at position 1 which was then used for all subsequent positions. The locations of the different head positions were measured with the miniBird™ position tracker from Ascension Technology Corporation and are listed in Table 5-9 and plotted in Figure 5-6 . Note that the head positions were recorded by the miniBird sensor located on top of the users head, with respect to the miniBird coordinate system. A transformation was performed on the data to convert the miniBird positions into the eye-gaze tracker world coordinate system, however they should only be considered approximate values. The range of X, Y and Z positions was 14.2 cm, 9.95 cm, and 20.64 cm respectively, covering the full FOV. The results of the tests are shown in Table 5-10, Table 5-11, and Table 5-12.

Table 5-9 - miniBird Head Positions

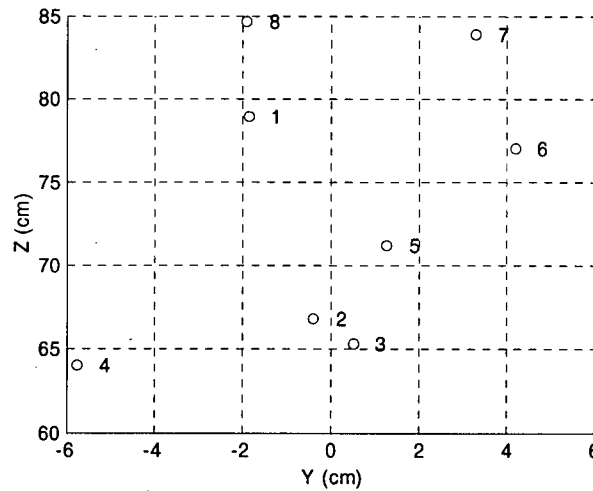
	Pos 1 (cm)	Pos 2 (cm)	Pos 3 (cm)	Pos 4 (cm)	Pos 5 (cm)	Pos 6 (cm)	Pos 7 (cm)	Pos 8 (cm)
X	6.21	1.58	10.87	1.24	0.37	12.40	2.80	14.57
Y	-1.85	-0.38	0.52	-5.75	1.28	4.20	3.31	-1.90
Z	78.89	66.81	65.35	64.03	71.22	77.04	83.89	84.67



Front View



Top View



Side View

Figure 5-6 – 3 Planar Views of the Head Position Locations

Table 5-10 - Head Position Results 1 – 3

		Position 1 (pixels)		Position 2 (pixels)		Position 3 (pixels)	
		X	Y	X	Y	Y	Y
4x4 Grid	Average Error	16.52	18.65	31.65	21.25	64.09	18.41
	Standard Deviation	14.4	13.22	27.68	13.5	39.6	15.21
	Maximum Error	38.12	26.5	87.16	26.5	110.71	41.85
	Maximum Error Loc	421	646	421	646	421	646
Single Point	Average Error	14.8	7.1	55.35	19.78	92.56	21.96
	Standard Deviation	10.97	5.16	14.29	5.66	12.95	4.73

Table 5-11 - Head Position Results 4 – 6

		Position 4 (pixels)		Position 5 (pixels)		Position 6 (pixels)	
		X	Y	X	Y	Y	Y
4x4 Grid	Average Error	59.32	26.74	20.2	27.08	15.22	15
	Standard Deviation	35.08	15.88	18.82	16.46	13.48	9.98
	Maximum Error	115.5	43.83	57.16	29.16	37	35.71
	Maximum Error Loc	421	646	421	646	842	969
Single Point	Average Error	55.94	52.52	15.76	22.55	9.36	5.69
	Standard Deviation	14.34	9.76	10.67	6.15	6.68	3.6

Table 5-12 - Head Position Results 7 – 8

		Position 7 (pixels)		Position 8 (pixels)	
		X	Y	X	Y
4x4 Grid	Average Error	39.76	20.77	34.86	23.6
	Standard Deviation	29.95	15.38	20.56	15.06
	Maximum Error	101.18	33.9	76.25	36.12
	Maximum Error Loc	842	969	842	969
Single Point	Average Error	67.53	6.05	47.92	27.16
	Standard Deviation	17.75	4.83	22.4	9.06

Statistical analysis of the average error indicates a statistically significant differences in X ($F(7, 937) = 54.388, p = 0.000$) and Y ($F(7, 937) = 9.304, p = 0.000$). The average error for X and Y for the 8 difference head positions are shown graphically in Figure 5-7.

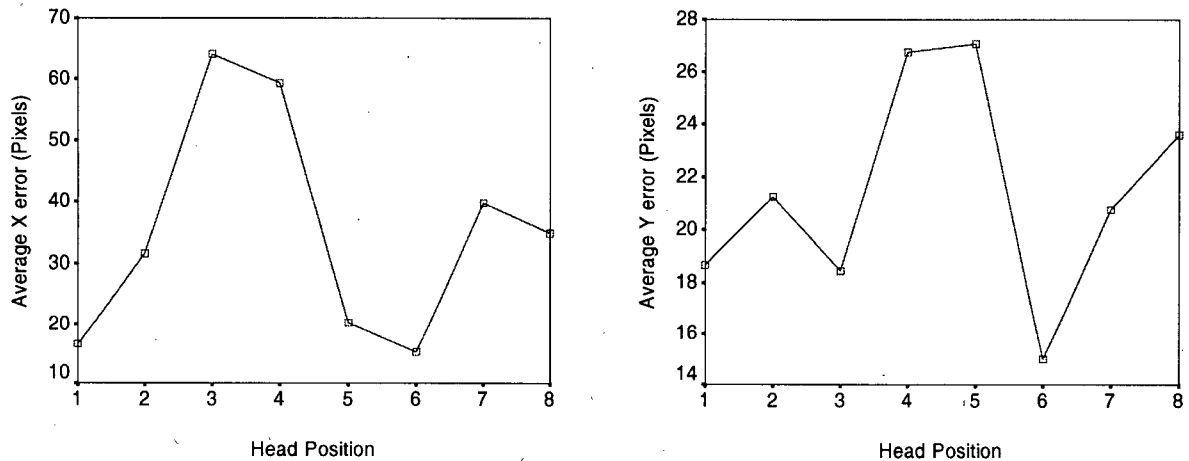


Figure 5-7 - Average X and Y Errors for 8 Head Positions

Over the full FOV average system accuracy ranged from 15 to 64 pixels of error. Analysis of the images recorded at positions 3 and 4 showed that they were too close to the camera and slightly out of focus, resulting in poor system accuracy. If positions 3 and 4 are not considered the system accuracy ranges from 15 to 39 pixels of error. The best pixel error converts to 0.53° of visual angle, using equation 43 to perform the conversion. The worst case pixel error results in 1.1° of visual angle.

5.6. Multi User Testing

The system was tested on a population sample to evaluate how well it performs with respect to different eye sizes and shapes, genders and ethnicities and the use of eye-glasses and contacts. Authorization for human testing was granted by the Behavioural Research Ethics Board in the Office of Research Services at the University of British Columbia. Test subject recruitment was informal, individuals who knew of the system and were curious to try it out were asked if they would also like to participate in the study. All test subjects were graduate or undergraduate students at the University of British Columbia. The test subjects were not paid for participating in the study.

In total there were 12 subjects comprised of 5 Caucasian subjects, 4 Middle Eastern subjects, 2 Indian subjects, and 1 Asian subject. There were 10 male subjects and 2 female, all with varying eye colours and shapes. Eye-glasses were worn by 2 of the subjects and contact lenses were worn by 1 subject, the remaining 9 did not require

corrective lenses. The system failed to operate on one Asian volunteer whose eye lids and eye lashes covered much of the pupil. This subject was not included in the 12 subjects reported here.

Multi user testing was performed without a chin rest and in a number of different ambient lighting conditions ranging from indirect sunlight during the day to no ambient sunlight in the evening. Multi user testing involved asking the user to sit comfortably in front of the computer screen while the system was calibrated and a data set was captured, similar to that of the single user testing. The user was then asked to reposition themselves in front of the monitor. A second data set was captured using the calibration data from the first trial, no re-calibration was performed for the second trial. The summary of the results for both trials for all subjects are shown in Table 5-13 through Table 5-18.

Table 5-13 - Multi User Trials 1 - 2

		Subject 1 Trial 1 (pixels)		Subject 1 Trial 2 (pixels)		Subject 2 Trial 1 (pixels)		Subject 2 Trial 2 (pixels)	
		X	Y	X	Y	X	Y	X	Y
4x4 Grid	Average Error	21.25	19	16.11	18.39	21.74	20.83	22.41	24.73
	Standard Deviation	19.31	14.97	18.79	15.64	17.51	16	18.7	17.95
	Maximum Error	42.5	36.39	65.11	47.11	58	23.19	55.4	48
	Maximum Error Loc	842	969	421	969	0	323	0	323
Single Point	Average Error	21.01	22.05	30.56	38.52	9.17	19.79	32.15	10.29
	Standard Deviation	14.32	9.69	12.95	15.09	6.85	5.87	16.06	8.49

Table 5-14 - Multi User Trials 3 - 4

		Subject 3 Trial 1 (pixels)		Subject 3 Trial 2 (pixels)		Subject 4 Trial 1 (pixels)		Subject 4 Trial 2 (pixels)	
		X	Y	X	Y	X	Y	X	Y
4x4 Grid	Average Error	15.22	16.52	22.72	25.37	21.14	33.76	20.13	12.53
	Standard Deviation	11.87	11.51	19.08	13.92	23.94	27.81	18.91	9.54
	Maximum Error	29	20	45.5	46.12	2.7	112.79	46.38	3.15
	Maximum Error Loc	842	0	842	0	1263	646	421	646
Single Point	Average Error	30.68	19.29	31.82	9.21	63.05	21.63	39.25	11.3
	Standard Deviation	13.45	8.18	18.19	7.59	27.67	19.1	17.44	7.18

Table 5-15 - Multi User Trials 5 - 6

		Subject 5 Trial 1 (pixels)		Subject 5 Trial 2 (pixels)		Subject 6 Trial 1 (pixels)		Subject 6 Trial 2 (pixels)	
		X	Y	X	Y	X	Y	X	Y
4x4 Grid	Average Error	22.39	40.6	32.28	35.59	26.77	16.82	26.93	15.42
	Standard Deviation	16.24	35.51	24.02	32.38	18.1	11.82	20.21	13.43
	Maximum Error	15.6	118	19.75	96.25	61.57	25.28	74.14	25.71
	Maximum Error Loc	421	0	842	969	0	969	0	0
Single Point	Average Error	18.08	21.12	31.8	56.53	28.63	14	19.02	6.26
	Standard Deviation	11.44	11.38	14.14	22.35	14.56	6.85	17.78	4.53

Table 5-16 - Multi User Trials 7 - 8

		Subject 7 Trial 1 (pixels)		Subject 7 Trial 2 (pixels)		Subject 8 Trial 1 (pixels)		Subject 8 Trial 2 (pixels)	
		X	Y	X	Y	X	Y	X	Y
4x4 Grid	Average Error	23.58	36.59	30.15	38	19.43	21.31	17.24	22.9
	Standard Deviation	22.25	26.68	20.87	22.24	16.36	14.72	15.72	21.82
	Maximum Error	40.88	86.11	56	65.33	54.6	33.33	10.79	76.4
	Maximum Error Loc	842	969	0	323	0	646	842	0
Single Point	Average Error	22.68	49.01	18.9	31.45	14.77	15.39	41.33	17.8
	Standard Deviation	21.18	7.45	8.95	9.35	9.75	11.14	28.85	10.56

Table 5-17 - Multi User Trials 9 - 10

		Subject 9 Trial 1 (pixels)		Subject 9 Trial 2 (pixels)		Subject 10 Trial 1 (pixels)		Subject 10 Trial 2 (pixels)	
		X	Y	X	Y	X	Y	X	Y
4x4 Grid	Average Error	25.27	14.79	35.12	16	19.09	15.24	23.14	14.63
	Standard Deviation	21.54	11.46	27.89	14.84	18.63	14.41	23.01	16.38
	Maximum Error	56.83	23.5	74.66	30.66	55.2	5	81.79	20.5
	Maximum Error Loc	421	969	421	969	842	323	842	969
Single Point	Average Error	45.63	24.5	42.64	36.11	8.68	13.1	22.89	10.1
	Standard Deviation	12.9	6.75	10.4	8.17	5.47	4.62	14.66	6.62

Table 5-18 - Multi User Trials 11 - 12

		Subject 11 Trial 1 (pixels)		Subject 11 Trial 2 (pixels)		Subject 12 Trial 1 (pixels)		Subject 12 Trial 2 (pixels)	
		X	Y	X	Y	X	Y	X	Y
4x4 Grid	Average Error	19.18	25.26	20.2	32.97	40.98	18.65	48.98	15.33
	Standard Deviation	15.65	21	18.71	22.12	30.42	13.76	36.62	14.97
	Maximum Error	14.5	69.66	55.83	48.83	112.11	32.33	117.61	36.84
	Maximum Error Loc	1263	646	842	323	842	969	842	969
Single Point	Average Error	6.5	25.56	34.78	9.1	53.98	22.65	50.11	8.53
	Standard Deviation	3.8	9.64	17.52	7.67	22.8	6.75	25.14	5.88

A comparison of the average error between each of the 12 subjects was performed for both trial 1 and trial 2. In trial 1, statistically significant differences were found between subjects for both the average errors in X ($F(11, 1427) = 14.129, p = 0.000$) and Y ($F(11, 1427) = 23.763, p = 0.000$). In trial 2, statistically significant differences were found between subjects for both the average errors in X ($F(11, 1590) = 24.198, p = 0.000$) and Y ($F(11, 1590) = 30.492, p = 0.000$). The average errors for each subject for trial 1 and trial 2 are plotted in Figure 5-8. The variation in accuracy may be due to the anatomical differences between subjects, as well as the different physical abilities of the subjects to use the system. The POG correction algorithm does not appear to be able to compensate for all of this variability to an equal extent.

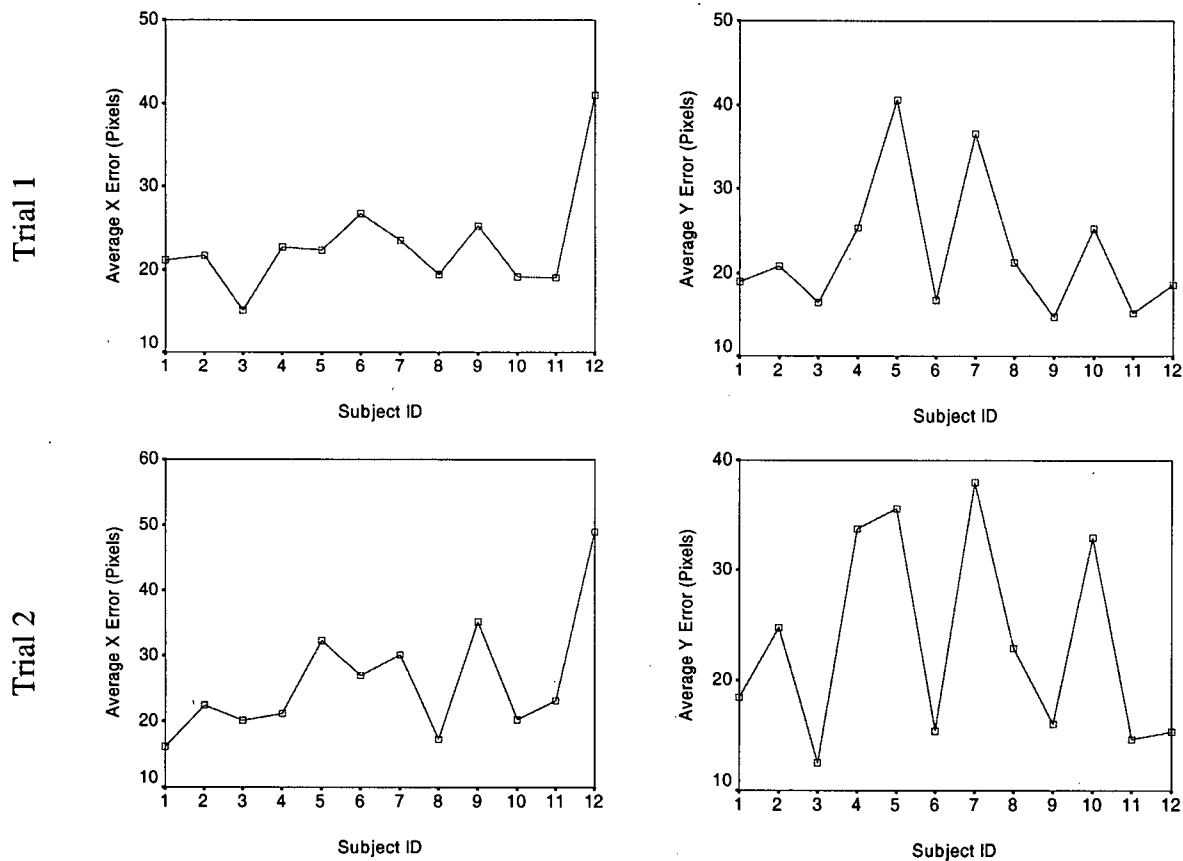


Figure 5-8 - Average Error Comparison Between Subjects for Trial 1 and Trial 2

Statistical comparisons between the average errors of Trial 1 and the average errors of Trial 2 were also performed for each of the subjects. The results of the tests indicate a statistically significant difference in the average X error for subjects 1, 3, 5, 7, 9, and 12 and a statistically significant difference in the average Y error for subjects 3 and 4. A summary of the ANOVA results are listed in Table 5-19.

Table 5-19 – Multi User Trial 1 and 2 ANOVA Comparison Results

Subject	X Coordinate ANOVA Result	Y Coordinate ANOVA Result
1	F (1, 244) = 4.095, p = 0.044	NS
2	NS	NS
3	F (1, 244) = 7.206, p = 0.008	F (1, 319) = 11.558, p = 0.001
4	NS	F (1, 264) = 9.250, p = 0.003
5	F (1, 210) = 12.042, p = 0.001	NS
6	NS	NS
7	F (1, 255) = 5.926, p = 0.016	NS
8	NS	NS
9	F (1, 176) = 6.931, p = 0.009	NS
10	NS	NS
11	NS	NS
12	F (1, 295) = 4.051, p = 0.045	NS

NS = No statistically significant difference

For some of the subjects the average error improved from trial 1 to trial 2 and for some the average error decreased from trial 1 to trial 2. Overall the average of the average X and Y errors for all 12 subjects was 23.00 pixels and 23.28 pixels respectively for trial 1 and 26.28 pixels and 22.66 pixels respectively for trial 2. The average of the maximum error for X and Y was 45.29 and 48.80 pixels respectively for trial 1 and 58.58 and 45.41 pixels respectively for trial 2. Using equation 43 to convert from pixels to degrees of visual angle results in average errors of 0.82° and 0.87° for trial 1 and trial 2 respectively and average maximum errors of 1.67° and 1.85° for trial 1 and trial 2 respectively.

6. Conclusions

6.1. Discussion and System Comparison

The objective of this thesis was to develop and evaluate a single camera, 3D model based, eye-gaze tracking system which utilizes the dual glint method. We were able to meet our objectives using in part, a novel combination of ideas existing in the eye-gaze tracking literature including:

- The image differencing technique by Ebisawa [39]
- The schematic eye by Gullstrand [13]
- The dual glint method for computing the center of the cornea by Shih [31]
- The approach to detecting the rough pupil by Nouredin [15]
- The ellipse fitting algorithm by Fitzgibbon [47]

Original Contributions of this Thesis

The original contributions of this thesis are as follows:

1. A novel combination of existing and new algorithms to provide acceptable performance in estimating the POG while maintaining real-time performance on a low-cost hardware platform. The new algorithms employed in this system are:

- The fine pupil detection algorithm
- Dual glint pattern detection
- The cornea center computation approach
- The pupil center identification approach
- The per-user POG calibration method

The approach we have taken has enabled the system to respond quickly to head movements and to reacquire tracking within 110 ms of losing lock on the eye, after a 3 frame delay of 198 ms which was added to prevent reprocessing the full image in the event of an eye blink. To the best of our knowledge, none of the pan/tilt/zoom systems can match this performance, and only the commercial Tobii system claims to be faster.

2. A method for dealing with the reflections from eye-glasses that may interfere with the image feature extraction system. Reflections off eye-glasses can occur when the lenses

are in a particular orientation with respect to the camera. We developed a method for the off-axis lighting system that activated an alternate set of lights when reflections were detected. The alternative set of lights were located such that no reflections were generated for that particular orientation of the glasses. For on-axis lighting we developed a method for blanking out the reflections to prevent false pupil detection. This allows the user to re-orientate their head at will without concern for reflections corrupting the images.

A comparison between the system we designed and the systems by Ohno and Tobii Technologies is shown below in Table 6-1. We chose to compare our system against the Ohno and Tobii systems as we feel that at this time they are the leading research and commercial systems.

Table 6-1- Comparison of Leading Systems with Our Design

	Ohno	Tobii	Our System
Accuracy (average error)	0.68° to 1° for 9 subject test	0.5° for 10 subject test, 0.5° to 1° for free head test	0.87° for 12 subject test, 0.54° to 1.1° for free head test
Cost	Possibly high, large amounts of hardware	\$25,000 USD (commercial price)	~1500 CAD (parts cost)
Camera Type	Analog with Frame Grabber	Digital Firewire	Digital Firewire
Update rate	30 Hz	50 Hz	15 Hz
Tracking recovery time	Not mentioned	< 100 ms	< 308 ms
Personal Calibration	2 point calibration	2 or more points	4 point or more
Non contact	Yes	Yes	Yes
Non intrusive	Yes	Yes	Yes
Non restrictive (XxYxZ)	10x10 with a large DOF at 60 cm	20x15x20 at 60 cm or 30x15x20 cm with both eyes	13x10x20 at 60 cm
Ability to handle ambient light	Not mentioned	Works with all but direct sun	Works with all but direct sun
Processor Overhead	High	Low	Low

POG estimation method	Stereo cameras, depth from focus, ray tracing	PG vector, multiple glints and ray tracing	Dual glint and ray tracing
Number of subjects tested	9 subjects tested	10 subjects tested	12 subjects tested
Tested with different head locations	2 positions	Yes, but no details given	6 locations tested
Works in the presence of eye-glasses and contact lenses	Yes, but no details given	Yes, but no details given	Yes with eye-glasses compensation Algorithm

When tested on multiple different users (12 subjects) the average system accuracy was 0.87° of visual angle. The average maximum error was 1.85° of visual angle, a statistic not reported for the other two systems. The accuracy of our system over the range of head locations tested ranged from 0.54° to 1.1° of visual angle. This is similar to the results reported by the Tobii system though they do not specify how their test was performed.

Lighting conditions from dark to indirect sunlight were tested and found not to affect the accuracy of the system. We were unable to design the system to operate with the subject's face in direct sunlight, though we proposed a method utilizing a narrow band filter which holds promise. The lighting requirements are similar between the Tobii system and our own.

The digital camera approach we used was very similar to the method used by Tobii with the fixed FOV and high resolution imaging. There are no moving parts to the system which leads to a fast re-acquisition time. Our FOV is slightly smaller than the Tobii system and the frame rate is lower, due to the limitation of frame rate of our camera. The speed of allowable head motion is up to 13 cm/s horizontally and 11 cm/s vertically compared with 10 cm/s reported for the Tobii system.

The system cost was approximately \$1500 CAD in materials including the camera and lens, electronics and mounting structure. The majority of the cost was for the digital camera and as digital imaging technology progresses this cost may be reduced.

Our work has resulted in a fully functional, free head motion, eye-gaze tracking system for which the full design specification and software source code is available. This system may be used as a testbed for future refinements, alternative design implementations, and as a tool for developing eye-gaze tracking applications.

6.2. Future Refinements

Currently the system is capable of tracking only one eye at a time. The addition of a face tracking algorithm would allow the system to know which eye is being tracked and make identification of the location of the eye in the full FOV much simpler. Tracking both eyes would allow for estimating the POG in free space by intersecting the LOS from both the right and left eyes, and would no longer require a geometrically simple gaze object (in our case the monitor plane).

The active lighting system for the off-axis lights is capable of completely preventing reflections from eye-glasses from corrupting the image of the dual glints. The software developed for compensating for the glasses reflections caused by the on-axis lights only works provide the reflection does not overlap the pupil contour. A method for avoiding the reflections caused by the on-axis lighting altogether would be desirable, similar to the off-axis lighting system.

A number of improvements to the camera would greatly improve the system. A higher resolution camera would help to increase the size of the FOV which would increase the allowable range of head positions, as well as improve the resolution of the images captured. A higher frame rate would increase the speed of the POG estimation and allow faster head motion without motion blur corrupting the POG estimation. Higher sensitivity to IR light would allow a smaller aperture for the equivalent image intensity which would increase the DOF and reduce the output intensity required from the system lighting. A search for an improved camera was performed but unfortunately no camera was found at the time that provided significant improvements that outweighed their costs. A summary of our camera search is reported in Appendix B.

Infrared diodes with a shorter wavelength for increased camera sensitivity while still being invisible to the human visual system would be desirable for improving the camera sensitivity to system lighting. Changes to the filter and system lighting could also

be made to allow for operation of the system in direct sunlight. A more precise testing method is also desirable to quantify the levels of ambient light in which the system is capable of operating.

A more accurate method for measuring the physical locations of the glints and monitor corners with respect to the world origin would reduce the reliance on per user calibrations.

The feature extraction algorithms should be analyzed to determine if fewer experimentally determined parameters could be used, possibly increasing detection robustness. A faster computer would decrease the CPU usage time allowing other CPU intensive application to run in conjunction with the eye-gaze tracking application.

Finally, the overall system accuracy may be improved by using multiple dual glint locations for computing the POG and averaging the results. Alternatively three or more glints could be used to estimate the location of the center of the cornea with higher accuracy. As well, system accuracy may be improved through the development of a more advanced POG correction algorithm which more directly compensates for each possible source of error found in the system. The free head POG algorithm is based on the assumption that the cornea can be modeled as a spherical surface. Further investigation as to the accuracy of this assumption is desirable as a refined cornea surface model may lead to improved POG estimation accuracy as well.

7. References

- [1] Jacob, R. J. K., and K. S. Karn, "Eye tracking in human-computer interaction and usability research: Ready to deliver the promises", In J. Hyona, R. Radach, and H. Deubel (Eds.), *The Mind's Eyes: Cognitive and Applied Aspects of Eye Movements*. Oxford: Elsevier, 2003.
- [2] Rayner, K., "Eye Movements in Reading and Information processing: 20 Years of Research", *Psychological Bulletin* 124 (3), 372-422, 1998.
- [3] Petersson, L, L. Fletcher, N. Barnes, A. Zelinsky, "An interactive driver assistance system monitoring the scene in and out of the vehicle", *Proc. IEEE International Conference on Robotics and Automation (ICRA'04)*, Vol 4, pp 3475 – 3481, 2004.
- [4] Anders, G., "Pilot's Attention Allocation during Approach and Landing - Eye and Head-Tracking Research in an A330 Full Flight Simulator", *Proc. of the 11th International Symposium on Aviation Psychology*, 2001.
- [5] Lohse, G.L., "Consumer Eye Movement Patterns on Yellow Pages Advertising", *Journal of Advertising*, 26 (1), pp. 61-73, 1997.
- [6] Wells, M.H. and M.J. Griffin, "A review and investigation of aiming and tracking performance with head-mounted sights", *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-17, pp. 210-221, 1987.
- [7] Bhuiyan, M.A. et al, "On tracking of eye for human-robot interface", *International Journal of Robotics and Automation*, v 19, n 1, pp. 42-54, 2004.
- [8] Jarvis, R., "A Go Where You Look Tele-Autonomous Rough Terrain Mobile Robot", 8th International Symposium on Experimental Robotics (ISER '02) in: *Advanced Robotics*, V. 5, pp. 624- 633, Springer-Verlag, 2003.
- [9] Loschky, L.C. and G.W. McConkie, "User performance with gaze contingent multiresolutional displays", *Proc. Eye Tracking Research and Applications Symposium (ETRA'00)*, pp 97-103, ACM, 2000.

- [10] Zhai, S., C. Morimoto, and S. Ihde, "Manual And Gaze Input Cascaded (MAGIC) Pointing", Proc. SIGCHI Conference on Computer Human Interaction (CHI'99), pp. 246-253, ACM, 1999.
- [11] Majaranta, P. and K.J. Raiha, "Twenty years of eye typing: systems and design issues", Proc. Eye Tracking Research and Applications Symposium (ETRA'02), pp. 15-22, ACM, 2002.
- [12] Morimoto, C. H., and M. R. M. Mimica, "Eye gaze tracking techniques for interactive applications", Computer Vision and Image Understanding 98, pp. 4-24, 2005.
- [13] Goss, D. A., and R. W. West, "Introduction to the Optics of the Eye", Butterworth-Heinemann, 2002.
- [14] Nguyen, K., C. Wagner, D. Koons, M. Flickner, "Differences in the Infrared Bright Pupil Response of Human Eyes", Proc. Eye Tracking Research and Applications Symposium (ETRA'02), pp 133-138, ACM, 2002.
- [15] Nouredin, B., "A Non-Contact Video-oculograph For Tracking Gaze in a Human Computer Interface", M.A.Sc Thesis, University of British Columbia, 2003.
- [16] Duchowski, A., "Eye Tracking Methodology – Theory and Practice", Springer-Verlag, 2003.
- [17] Merchant, S., "Eye Movement Research In Aviation and Commercially Available Eye Trackers Today", Technical Report, University of Iowa, 2001.
- [18] Larsen, J., L. Stark, "Difficulties in calibration or instrumentation for eye movements", Proceedings of IEEE International Conference on Systems, Man, and Cybernetics, Vol. 1, pp. 297-298, 1988.
- [19] Hutchinson, T., "Human-computer interaction using eye-gaze input", IEEE Transactions on Systems, Man and Cybernetics, Vol. 19, pp. 1527-1534, 1989.
- [20] Eye Response Technologies, <http://www.eyeresponse.com/ericasystem.html>
- [21] LC Technologies, <http://www.eyegaze.com/>
- [22] A-S-L Laboratories, <http://www.a-s-l.com/>

- [23] Mimica, M.R.M. and C.H. Morimoto, "A Computer Vision Framework for Eye Gaze Tracking", Proc. Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI'03), pp 406-412, ACM, 2003.
- [24] Schnipke, S. K., and M. W. Todd, "Trials and tribulations of using an eye-tracking system", Proc. SIGCHI Human Factors in Computing Systems Conference, pp. 273 -274, ACM, 2000.
- [25] Zhu, J., and J. Yang, "Subpixel Eye Gaze Tracking", Proc. of 5th International Conference on Automatic Face and Gesture Recognition, Vol. 4, pp 201-204, 2000.
- [26] Wang, J.-G., E. Sung, R. Venkateswarlu, "Estimating the eye gaze from one eye", Computer Vision and Image Understanding, Vol 98, Issue 1, pp 83-103, 2005.
- [27] C. Morimoto, A. Amir, M. Flickner, "Free Head Motion Eye Gaze Tracking Without Calibration", Proc. Conference on Computer Human Interaction (CHI'02), pp 586-587, 2002.
- [28] Yoo, D.H., and M.J. Chung, "A novel non-intrusive eye gaze estimation using cross-ratio under large head motion", Computer Vision and Image Understanding, Vol. 98, Issue 1, pp 25-51, 2005.
- [29] Nouredin, B., P.D. Lawrence, and C.F. Man, "A Non-Contact Device For Tracking Gaze in a Human Computer Interface", Computer Vision and Image Understanding, Vol 98, Issue 1, pp 52-82, 2005.
- [30] Beymer, D., and M. Flickner, "Eye Gaze Tracking Using an Active Stereo Head", Proc. IEEE Computer Soc. Conf. on Computer Vision and Pattern Recognition (CVPR'03), v. 2, pp 451-458, 2003.
- [31] Shih, S. and J. Liu, "A Novel Approach to 3-D Gaze Tracking Using Stereo Cameras", IEEE Trans. Syst. Man Cybernetics – Part B, Vol. 34, No 1, pp 234-245, 2004.
- [32] Ohno, T., and N. Mukawa, "A Free-head, Simple Calibration, Gaze Tracking System That Enables Gaze-Based Interaction", Proc. Eye Tracking Research and Applications Symposium (ETRA'04), pp115-122, ACM, 2004.

- [33] Ohno, T., N. Mukawa, and A. Yoshikawa, "FreeGaze: A Gaze Tracking System for Everyday Gaze Interaction", Proc. Eye Tracking Research and Applications Symposium (ETRA'02), pp125-132, ACM, 2002.
- [34] Tobii Technologies, <http://www.tobii.se/>
- [35] Seeing Machines, <http://www.seeingmachines.com/index.htm>
- [36] Matsumoto Y. and, A. Zelinsky, "An algorithm for real-time stereo vision implementation of head pose and gaze direction measurement", Proc. 4th IEEE International Conference on Automatic Face and Gesture Recognition, pp. 499-504, 2000.
- [37] Elvesjoe, J., M. Skogoe, and E. Gunnar, "Method and Installation for Detecting and Following an Eye and the Gaze Direction Thereof", World Patent WO20044045399, 2004.
- [38] Markerink, W. J., <http://www.a1.nl/phomepag/markerink/irfilter.htm>
- [39] Ebisawa, Y., "Improved Video-Based Eye-Gaze Detection Method", IEEE Transactions on Instrumentation and Measurement, 47:4, pp. 948-955, 1998.
- [40] "Compliance of Infrared Communication Products to IEC 825-1 and CENELEC EN 60825-1", Application Note 1118, Agilent Technologies, 1999.
- [41] Voke, J., "Radiation effects on the eye", Optometry Today, pp 27-32, 1999.
- [42] "Safety of laser products – Part 9: Complication of maximum permissible exposure to incoherent optical radiation", International Electrotechnical Commission (IEC) Technical Report 60825-9, 1999.
- [43] Open Computer Vision Library,
<http://www.intel.com/research/mrl/research/opencv/>
- [44] Zhu, D., S. T. Moore, T. Raphan, "Robust pupil center detection using a curvature algorithm", Computer Methods and Programs in Biomedicine, No. 59, pp. 145-157, 1999.
- [45] Haro, A., M. Flickner, I. Essa, "Detecting and Tracking Eyes By Using Their Physiological Properties, Dynamics, and Appearance", IEEE Conference on Computer Vision and Pattern Recognition, pp 163-168, 2000.
- [46] Isoperimetric quotient, <http://mathworld.wolfram.com/IsoperimetricQuotient.html>

- [47] Fitzgibbon, A.W., M. Pilu, R. B. Fisher, "Direct Least Squares Fitting of Ellipses", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, no. 5, pp. 476-480, 1999.
- [48] Fixed Point Iteration and Newtons Method,
<http://math.fullerton.edu/mathews/n2003/FixPointNewtonMod.html>
- [49] Craig, J., "Introduction to Robotics: Mechanics and Control", 2nd edition, Addison-Wesley Publishing Company, 1989.
- [50] Tsai, R.Y., "A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses", IEEE Journal of Robotics and Automation, Vol. RA-3, No. 4, pp. 323-344, 1987.
- [51] Zhang, Z., "A flexible new technique for camera calibration, IEEE Transactions on Pattern Analysis and Machine Intelligence", 22(11), pp. 1330-1334, 2000.
- [52] Heikkilä, J., O. Silvén, "A Four-step Camera Calibration Procedure with Implicit Image Correction", IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'97), pp. 1106-1112, 1997.
- [53] Matlab Camera Calibration Toolbox,
http://www.vision.caltech.edu/bouguetj/calib_doc/
- [54] Stewart, J., "Calculus – Early Transcendentals", 3rd edition, Brooks / Cole Publishing, Pacific Grove, CA, 1995.

8. Appendix A - Infrared LED Safety

The International Electrotechnical Commission (IEC) Technical Report 60825-9 specifies the maximum permissible exposure (MPE) for incoherent optical radiation. To ensure the REGT device is safe for human use it will only be operated below the MPE values specified. The MPEs listed are for an exposure duration of approximately eight hours.

8.1. System Diodes

Illumination of the face is achieved using Agilent HSDL-4220, high-performance AlGaAs infrared LEDs. The relevant component information for these diodes is summarized in Table 8-1.

Table 8-1 – HSDL-4220 LED Specification

HSDL-4220	Value	Units
Viewing Angle	30	deg
Operating Current	50	mA
Radiant Optical Power	19	mW
Radiant On-Axis Intensity	38	mW / sr
Peak Wavelength	875	nm

During on-axis illumination, six diodes are activated with approximately 50 mA of current each. During off-axis illumination, there are two sets of four diodes for a total of eight. To achieve similar illumination between the on-axis and off-axis lighting, the off-axis diodes current is set to approximately 100 mA each, or 800 mA total.

8.2. Maximum Permissible Exposure

There are three MPEs which relate to this particular system. The retinal thermal hazard MPE L_{RTH} ensures no thermal damage will occur to the retina. For infrared sources which do not activate the natural human aversion response to over stimulus, the L_{IR} MPE applies. Finally the MPE to prevent cornea and lens damage is specified by E_{IR} . These three MPE equations are shown below:

$$L_{RTH} = \frac{2.8 \cdot 10^4}{C_\alpha} \frac{W}{m^2 \cdot sr} \quad (44)$$

$$L_{IR} = \frac{6000}{C_\alpha} \frac{W}{m^2 \cdot sr} \quad (45)$$

$$E_{IR} = 100 \frac{W}{m^2} \quad (46)$$

To determine the MPE values for the system, the correction factor C_α for the limiting angular subtense of the eye must be determined. The angular subtense α is the visual angle subtended by the apparent source at the eye of an observer, see Figure 8-1.

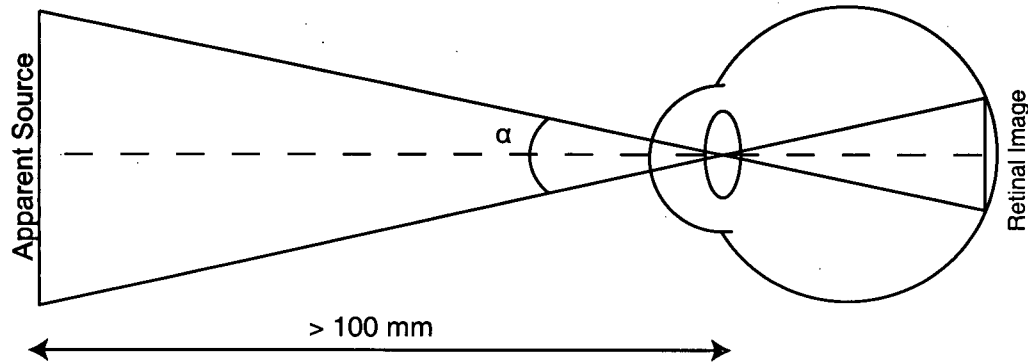


Figure 8-1 - Angular Subtense

The correction factor depends on the size of the apparent source. Measuring the actual apparent source size of the lights used would be difficult, as the LED elements are encased in an epoxy housing which both protects the element and acts as a lens. In addition arrays of diodes are being used, rather than single elements. The most restrictive apparent source size is used in the calculations to ensure the MPE used will match or exceed the actual MPE. The most restrictive correction factor for the MPE calculations is $C_\alpha = \alpha_{\max} = 0.1$ radians. Using this correction factor and converting units to mW / cm^2 , equations 44 through 46 become:

$$L_{RTH} = 2.8 \cdot 10^4 \cdot \frac{mW}{cm^2 \cdot sr} \quad (47)$$

$$L_{IR} = 6000 \frac{mW}{cm^2 \cdot sr} \quad (48)$$

$$E_{IR} = 10 \frac{mW}{cm^2} \quad (49)$$

By complying with the most restrictive MPE the safety of the system is ensured under all conditions. To compare E_{IR} with the other MPE's it must be converted from irradiance to luminance:

$$L = E \cdot \frac{4}{\alpha^2 \cdot \pi} \quad (50)$$

$$L = 10 \frac{mW}{cm^2} \cdot \frac{4}{(0.1)^2 \cdot \pi} = 1273 \frac{mW}{cm^2 \cdot sr} \quad (51)$$

The most restrictive MPE results in a maximum of $1273 \frac{mW}{cm^2 \cdot sr}$.

8.3. System Radiance

To determine the radiance of the system, the area over which the light falls must be determined. The MPE specification in the IEC technical report 60825-9 assumes an 8 hour setting with an average the pupil diameter of 7 mm and area of 0.3848 cm^2 .

$$P_{area} = \pi \cdot \left(\frac{0.7}{2} \right)^2 = 0.3848 \text{ cm}^2 \quad (52)$$

The radiance for the on-axis and off-axis LEDs may then be calculated as:

$$L_{on-axis} = 38 \frac{mW}{sr} \cdot \frac{1}{0.3848 \text{ cm}^2} \cdot 6 \text{ LEDs} = 588 \frac{mW}{cm^2 \cdot sr} \quad (53)$$

$$L_{8xLED} = 76 \frac{mW}{sr} \cdot \frac{1}{0.3848 \text{ cm}^2} \cdot 8 \text{ LEDs} = 1580 \frac{mW}{cm^2 \cdot sr} \quad (54)$$

As each set of lights is on for 50% of the time each the total system radiance is:

$$L_{total} = 0.5 \left(1580 \frac{mW}{cm^2 \cdot sr} \right) + 0.5 \left(588 \frac{mW}{cm^2 \cdot sr} \right) = 1084 \frac{mW}{cm^2 \cdot sr} \quad (55)$$

As has been shown, the radiance of the system is $1084 \frac{mW}{cm^2 \cdot sr}$ compared with a very conservative maximum permissible exposure of $1273 \frac{mW}{cm^2 \cdot sr}$.

9. Appendix B – Camera Search

While the Dragonfly camera used for this project was adequate, a number of possible improvements were identified. A search of available commercial cameras was performed to see if a better camera could be found. The following is a list of desirable attributes:

- Low cost
- > 1024x768 resolution
- > 15 fps update rate
- Good sensor sensitivity to 875 nm IR light
- External strobe to sync bright / dark pupil illumination
- General purpose IO to control lighting
- Global shutter for uniform image illumination and reduced blur
- Hardware ROI to reduce data bus bandwidth

Shown in Table 9-1 is a comparison between the Dragonfly and the cameras that came close to meeting the desired specifications. None of the cameras found appeared to offer a significant enough improvement to warrant purchasing, especially given their costs. Not all of the required information could be found for each of the cameras.

Table 9-1 - Camera Search

Company	Pt Grey	Pt Grey	Lumenera	Pixelink	C-Cam Technology	Imaging Source	Prosilica
Model	Dragonfly	SCOR-13FFM-CS	Lu 100	PL-A741	BCi5	DMK 41BF02	CV1280
Maximum Resolution	1024x768	1280x1024	1280x1024	1280x1024	1280x1024	1280x960	1280x960
Cost	Own	1095 USD	745 USD	1795 USD	EUR 1500	1290 USD	1895 USD
SDK Cost	Own	1195 USD	?	?	?	?	0
FPS at max	15 fps	15 fps	15 fps	27 fps	27.5 fps	15 fps	24 fps
HW ROI	Y	Y	Y	Y	Y	?	Y
External Strobe	Y	Y	Y	Y	Y	?	Y
Monochrome	Y	Y	Y	Y	Y	Y	Y
Sensitivity at 875 nm	~ 18%	? ~33%	80% eff	? ~33%	~33%	?	~33%
Global Shutter	Y	Y	Half Global	Y	Y	?	Y
CCD Size	1/3"	2/3"	1/2"	2/3"	2/3"	1/2"	2/3"
CCD Sensor	Sony ICX204	Fill Factory IBIS5A – 1300-M	?	? Probably Fill Factory	Fill Factory IBIS5A – 1300-M	?	? Probably Fill Factory

10. Appendix C – Camera Calibration

The intrinsic pin-hole camera model parameters depend on the lens, camera and mounting system used and as such cannot be specified by the camera manufacturer. The parameters must be determined through a method of camera calibration. Tsai [50], Heikkilä and Silvén [52] and Zhang [51] describe how the intrinsic parameters can be estimated automatically by imaging a large number of features with known relative displacements. Typically these features are the corners in a checkerboard pattern as shown in Figure 10-1 and Figure 10-2. An iterative approach to solving for the model parameters is performed by projecting the checkerboard points in the image space out to points in the object space. The error between the estimated points and the actual points can be determined as the size of the checkerboard pattern is known precisely. The error is minimized in a least mean squares sense using a gradient descent algorithm. A Matlab based implementation called The Camera Calibration Toolbox developed by Jean-Yves Bouguet [53] was used to estimate the intrinsic camera parameters for our camera.

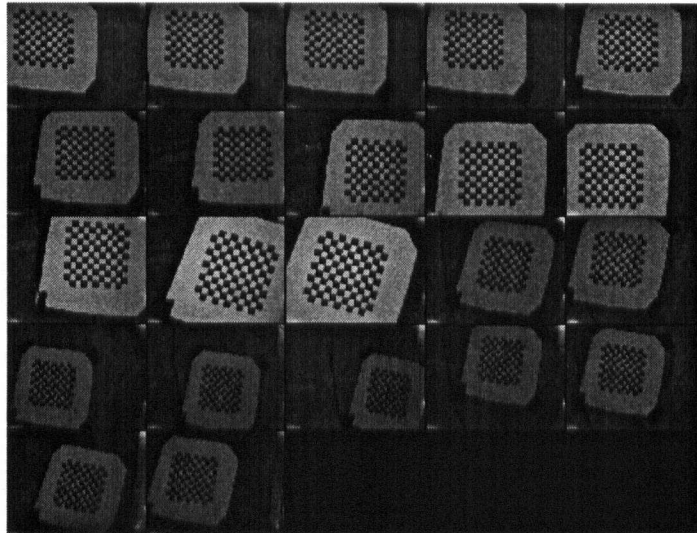


Figure 10-1 – Checkerboard Images Used to Calibrate the Camera

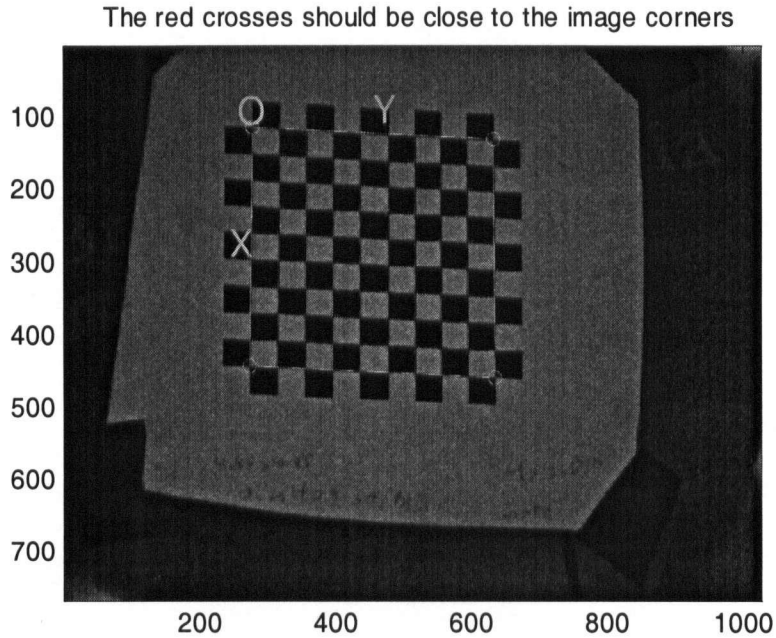


Figure 10-2 – Identified Corners in a Checkerboard Image

Table 10-1 - Camera Calibration Parameters

Variable	Description	Value (pixels)	Value (mm)
f	Effective focal length	6881	32
c_p	Critical point	(637.25, 46.84)	(3.0, 0.2)

The CCD sensor is a grid made up of 1024x768 pixels, with each square pixel 4.65 μm a side. The identified critical point is somewhat far from what might be expected for the center of the CCD ((1024, 768) / 2, or (512, 384)). This may be due to the physical mounting of the lens with respect to the CCD sensor or may in fact be incorrectly identified by the camera calibration. Both of the points (637.25, 46.84) and (512, 384) were tested as the critical point with little difference in the accuracy of the estimated POG. The per-user calibration compensates for offsets in POG estimation due to possible errors in the intrinsic camera parameters.

11. Appendix D - Electronics Schematic

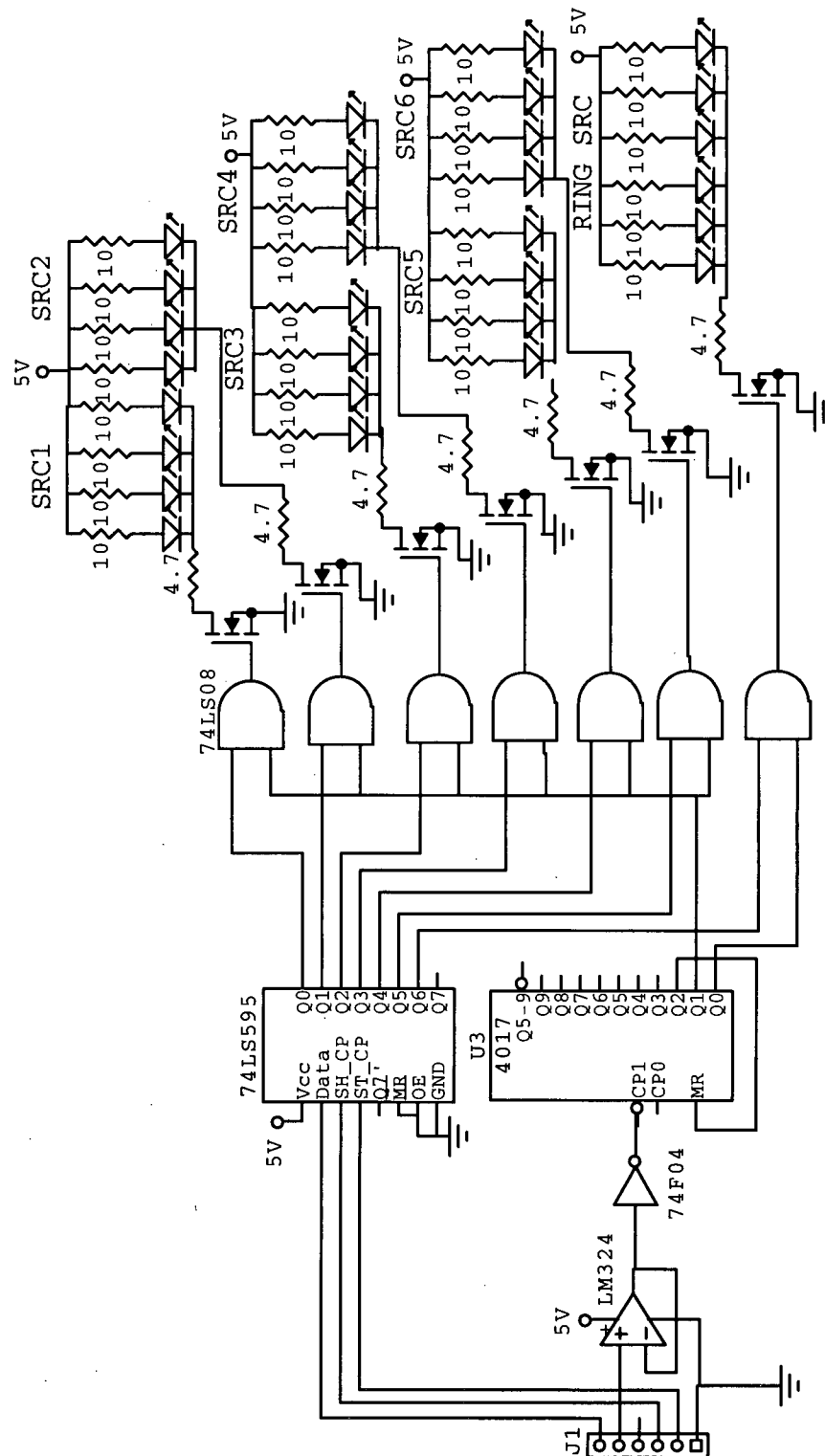


Figure 11-1 - Infrared Diode Schematic

12. Appendix E - Image Processing Parameters

<i>Parameter</i>	<i>Value</i>	<i>Units</i>
<i>Application</i>		
Size of eye ROI rectangle	200 x 175	Pixels
Size of pupil ROI rectangle	110 x 110	Pixels
<i>Rough Pupil</i>		
Gaussian smoothing kernel	5 x 5	Pixels
Pupil contour pixel count (for determining pupil threshold)	1000	Pixels
Minimum pupil contour size	50	Pixel Area
Maximum pupil contour size	200	Pixel Area
Minimum pupil width	15	Pixel Area
Maximum pupil width	65	Pixel Area
Minimum pupil height	15	Pixel Area
Maximum pupil height	65	Pixel Area
Pupil contour area / perimeter rejection limit	0.02	unitless
Pupil lost count	3	unitless
Glasses Reflection Identification Level	170	Intensity
Glasses Reflection Threshold Percentage	60	%
<i>Pupil Glint</i>		
Radius of Search Mask	42	Pixels
Pupil glint threshold scaling factor	90	%
Pupil glint contour maximum size	20	Pixel Area
<i>Fine Pupil</i>		
Pupil glint dilation kernel	5 x 5	Pixels
Bright pupil threshold scaling factor	65	%
Fine pupil contour minimum size	50	Pixel Area
Fine pupil contour maximum size	200	Pixel Area
<i>Dual Glint</i>		
Radius of Search Mask	42	Pixels
Dual Glint threshold scaling factor	80	%
Glasses Glare Average Intensity Threshold	2.0	Intensity
Dual glint minimum x difference	Varies	Pixels
Dual glint maximum x difference	Varies	Pixels
Dual glint maximum y difference	Varies	Pixels
Dual glint maximum y difference	Varies	Pixels
Glasses Maximum Dual Glint Failures	10	Tries

13. Appendix F – Auxiliary / World Rotation Matrixes

```
void CGeometryCalc::compute_rot (CMatrix xhat, CMatrix yhat, CMatrix zhat,
CMatrix &rot2new, CMatrix &rot2world)
{
    float thetazback;
    float thetayback;
    float thetaxback;

    CMatrix rz (3, 3);
    CMatrix ry (3, 3);
    CMatrix rx (3, 3);

    CMatrix xhatback1 (3, 1);
    CMatrix yhatback1 (3, 1);
    CMatrix zhatback1 (3, 1);
    CMatrix xhatback2 (3, 1);
    CMatrix yhatback2 (3, 1);
    CMatrix zhatback2 (3, 1);
    CMatrix xhatback3 (3, 1);
    CMatrix yhatback3 (3, 1);
    CMatrix zhatback3 (3, 1);

    // find z rotation to rotate xhat coordinate into XZ plane
    thetazback = -atan2 (xhat.get (1,0), xhat.get (0,0));

    rz.set (0, 0, cos(thetazback));
    rz.set (0, 1, -sin(thetazback));
    rz.set (0, 2, 0);
    rz.set (1, 0, sin(thetazback));
    rz.set (1, 1, cos(thetazback));
    rz.set (1, 2, 0);
    rz.set (2, 0, 0);
    rz.set (2, 1, 0);
    rz.set (2, 2, 1);

    xhatback1 = rz * xhat;
    yhatback1 = rz * yhat;
    zhatback1 = rz * zhat;

    // find y rotation to rotate xhat coordinate into x axis
    thetayback = atan2 (xhatback1.get (2, 0), xhatback1.get (0, 0));

    ry.set (0, 0, cos(thetayback));
    ry.set (0, 1, 0);
    ry.set (0, 2, sin(thetayback));
    ry.set (1, 0, 0);
    ry.set (1, 1, 1);
    ry.set (1, 2, 0);
    ry.set (2, 0, -sin(thetayback));
    ry.set (2, 1, 0);
    ry.set (2, 2, cos(thetayback));

    xhatback2 = ry * xhatback1;
    yhatback2 = ry * yhatback1;
    zhatback2 = ry * zhatback1;

    // find x rotation to rotate yhat and zhat coordinates into y and z axis
    thetaxback = -atan2 (yhatback2.get (2, 0), yhatback2.get (1, 0));

    rx.set (0, 0, 1);
    rx.set (0, 1, 0);
    rx.set (0, 2, 0);
    rx.set (1, 0, 0);
```

```

rx.set (1, 1, cos(thetaxback));
rx.set (1, 2, -sin(thetaxback));
rx.set (2, 0, 0);
rx.set (2, 1, sin(thetaxback));
rx.set (2, 2, cos(thetaxback));

xhatback3 = rx * xhatback2;
yhatback3 = rx * yhatback2;
zhatback3 = rx * zhatback2;

// rotatates PT TO new coordinate FROM world coordinate
rz.set (0, 0, cos(thetazback));
rz.set (0, 1, -sin(thetazback));
rz.set (0, 2, 0);
rz.set (1, 0, sin(thetazback));
rz.set (1, 1, cos(thetazback));
rz.set (1, 2, 0);
rz.set (2, 0, 0);
rz.set (2, 1, 0);
rz.set (2, 2, 1);

ry.set (0, 0, cos(thetayback));
ry.set (0, 1, 0);
ry.set (0, 2, sin(thetayback));
ry.set (1, 0, 0);
ry.set (1, 1, 1);
ry.set (1, 2, 0);
ry.set (2, 0, -sin(thetayback));
ry.set (2, 1, 0);
ry.set (2, 2, cos(thetayback));

rx.set (0, 0, 1);
rx.set (0, 1, 0);
rx.set (0, 2, 0);
rx.set (1, 0, 0);
rx.set (1, 1, cos(thetaxback));
rx.set (1, 2, -sin(thetaxback));
rx.set (2, 0, 0);
rx.set (2, 1, sin(thetaxback));
rx.set (2, 2, cos(thetaxback));

rot2new = rx * ry * rz;

// rotatates PT FROM new coordinate TO world coordinate
rz.set (0, 0, cos(-thetazback));
rz.set (0, 1, -sin(-thetazback));
rz.set (0, 2, 0);
rz.set (1, 0, sin(-thetazback));
rz.set (1, 1, cos(-thetazback));
rz.set (1, 2, 0);
rz.set (2, 0, 0);
rz.set (2, 1, 0);
rz.set (2, 2, 1);

ry.set (0, 0, cos(-thetayback));
ry.set (0, 1, 0);
ry.set (0, 2, sin(-thetayback));
ry.set (1, 0, 0);
ry.set (1, 1, 1);
ry.set (1, 2, 0);
ry.set (2, 0, -sin(-thetayback));
ry.set (2, 1, 0);
ry.set (2, 2, cos(-thetayback));

rx.set (0, 0, 1);
rx.set (0, 1, 0);
rx.set (0, 2, 0);

```

```
rx.set (1, 0, 0);
rx.set (1, 1, cos(-thetaxback));
rx.set (1, 2, -sin(-thetaxback));
rx.set (2, 0, 0);
rx.set (2, 1, sin(-thetaxback));
rx.set (2, 2, cos(-thetaxback));
    rot2world = rz * ry * rx;
}
```

14. Appendix G – Intersection of a Line with a Sphere

```
int linesphere (CMatrix P1, CMatrix P2, CMatrix C, float r, CMatrix &P1int,
CMatrix &P2int)
{
    int num;

    float d1, d2, d3;
    float disc;
    float t1, t2;
    float dist1, dist2;

    CMatrix v (3, 1);
    CMatrix Ptemp (3, 1);

    // These subtractions are performed many times, computing them once speeds
    execution
    d1 = P1.get (0, 0) - C.get (0, 0);
    d2 = P1.get (1, 0) - C.get (1, 0);
    d3 = P1.get (2, 0) - C.get (2, 0);

    v = P2 - P1;

    // Discriminant
    disc = 2 * d2 * v.get (1, 0) * d1 * v.get (0, 0) + \
        2 * d2 * v.get (1, 0) * d3 * v.get (2, 0) + \
        2 * d1 * v.get (0, 0) * d3 * v.get (2, 0) + \
        v.get (0, 0) * v.get (0, 0) * r * r - \
        v.get (0, 0) * v.get (0, 0) * d3 * d3 - \
        v.get (0, 0) * v.get (0, 0) * d2 * d2 - \
        v.get (2, 0) * v.get (2, 0) * d1 * d1 + \
        v.get (2, 0) * v.get (2, 0) * r * r - \
        v.get (2, 0) * v.get (2, 0) * d2 * d2 - \
        v.get (1, 0) * v.get (1, 0) * d1 * d1 + \
        v.get (1, 0) * v.get (1, 0) * r * r - \
        v.get (1, 0) * v.get (1, 0) * d3 * d3;

    // No intersection
    if (disc < 0)
    {
        num = 0;
    }
    // One intersection (tangent)
    else if (disc == 0)
    {
        num = 1;
    }
    // Two intersections
    else
    {
        num = 2;
    }

    // Zero intersection points
    P1int.set (0, 0, 0);
    P1int.set (1, 0, 0);
    P1int.set (2, 0, 0);

    P2int.set (0, 0, 0);
    P2int.set (1, 0, 0);
    P2int.set (2, 0, 0);

    if (disc >= 0)
    {
```

```

        // Compute parameteric 't' for both intersection points
        t1 = -1.0*(d2*v.get (1,0)+d1*v.get (0,0)+d3*v.get (2,0)-sqrt(disc))/(v.get
(0,0)*v.get (0,0)+v.get (2,0)*v.get (2,0)+v.get (1,0)*v.get (1,0));
        t2 = -1.0*(d2*v.get (1,0)+d1*v.get (0,0)+d3*v.get (2,0)+sqrt(disc))/(v.get
(0,0)*v.get (0,0)+v.get (2,0)*v.get (2,0)+v.get (1,0)*v.get (1,0));

        // Compute intersection point
        P1int = v * t1 + P1;
        P2int = v * t2 + P1;
    }

    // Find closer intersection point by finding the point closer to P1
    dist1 = norm (P1int - P1);
    dist2 = norm (P2int - P1);

    // P1int is the closer one
    if (dist1 > dist2)
    {
        Ptemp=P1int;
        P1int=P2int;
        P2int=Ptemp;
    }

    return num;
}

```

15. Appendix H – Intersection of a Line with a Plane

```
void CGeometryCalc::compute_POG (CMatrix C, CMatrix LOS, CMatrix &POG)
{
    float t;
    CMatrix n (3, 1);

    // find normal to monitor plane
    n = cross (fM2-fM1, fM3-fM1);

    // From maple, parametric value for t
    t = -(n.get (0, 0) * C.get (0, 0) - n.get (1, 0) * fM1.get (1, 0) - n.get (0,
0) * fM1.get (0, 0) + n.get (1, 0) * C.get (1, 0) - n.get (2, 0) * fM1.get (2,
0) + n.get (2, 0) * C.get (2, 0)) / (n.get (0, 0) * LOS.get (0, 0) + n.get (1,
0) * LOS.get (1, 0) + n.get (2, 0) * LOS.get (2, 0));

    POG = C + t * LOS;
}
```


16. Appendix I – System Testing Figures

16.1. Head Rest

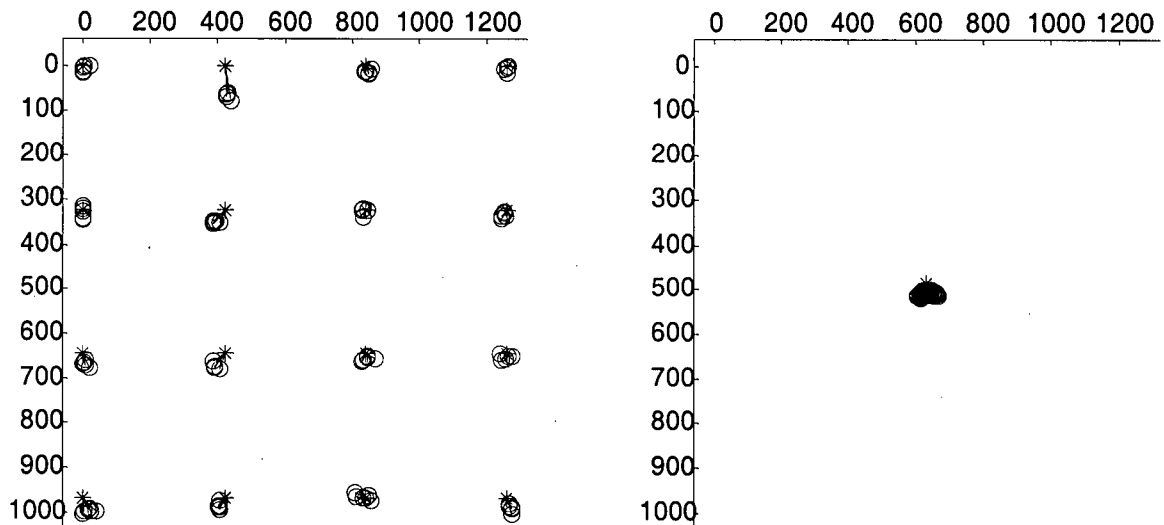


Figure 16-1 - Without Headrest

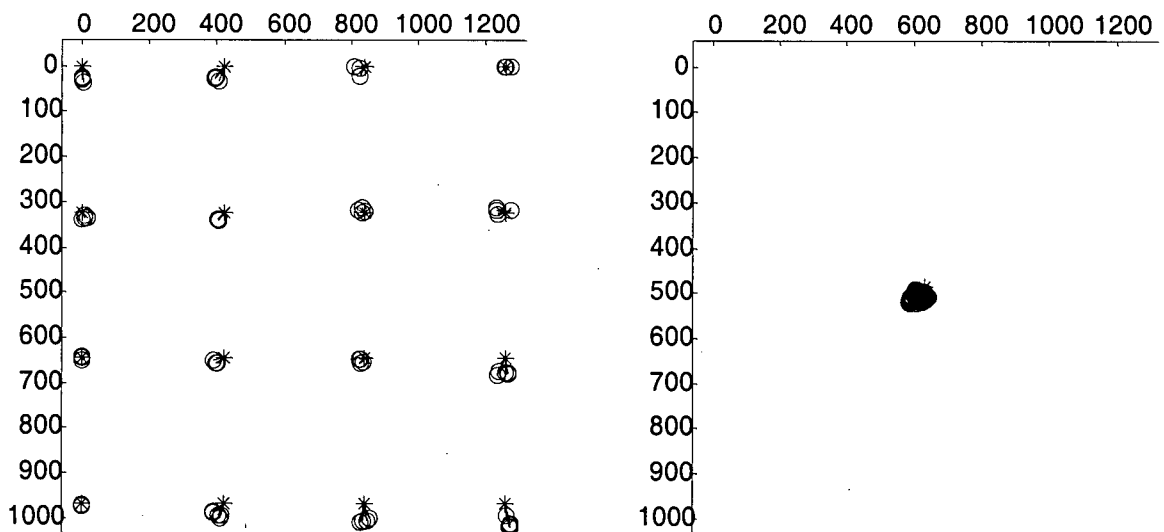


Figure 16-2 - With Headrest

16.2. Shutter Speed

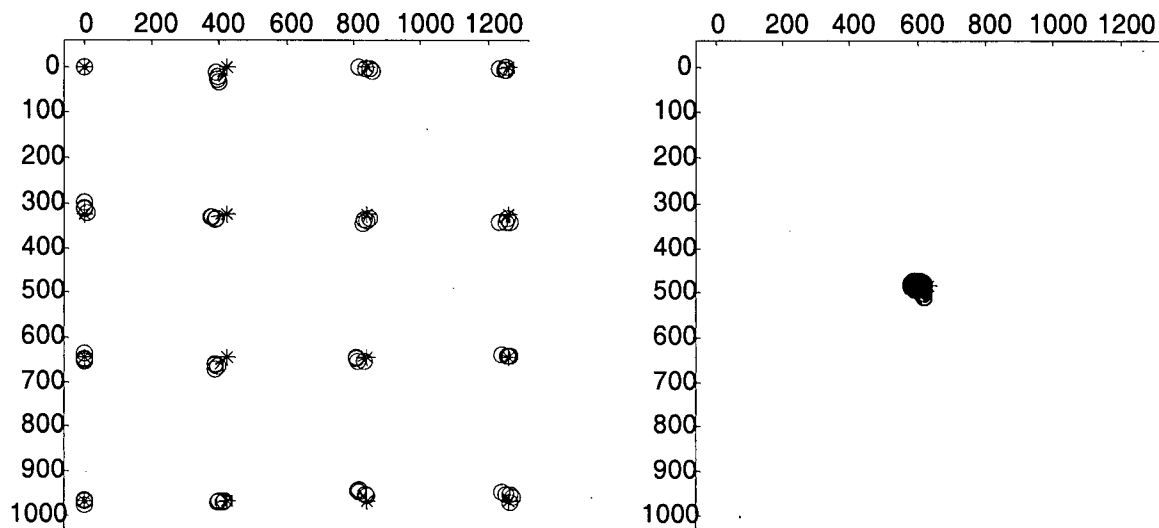


Figure 16-3 – 8.33 ms Shutter

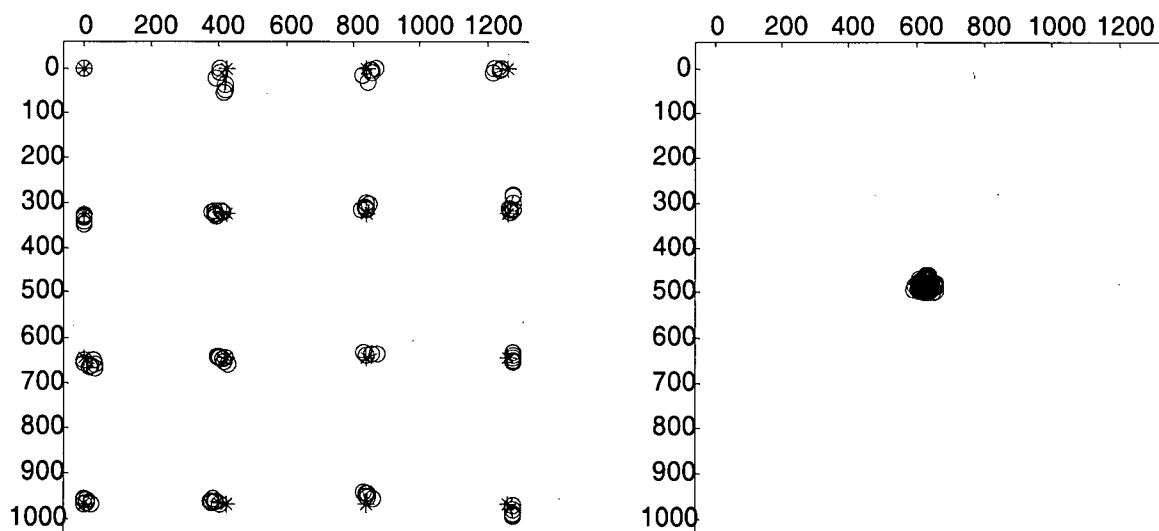


Figure 16-4 – 25.00 ms Shutter

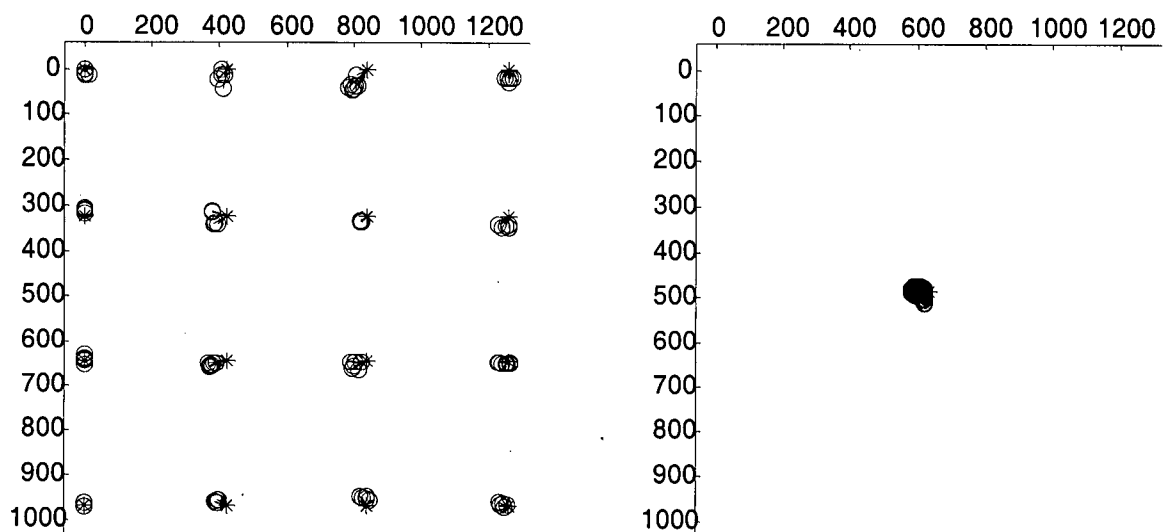


Figure 16-5 – 58.33 ms Shutter

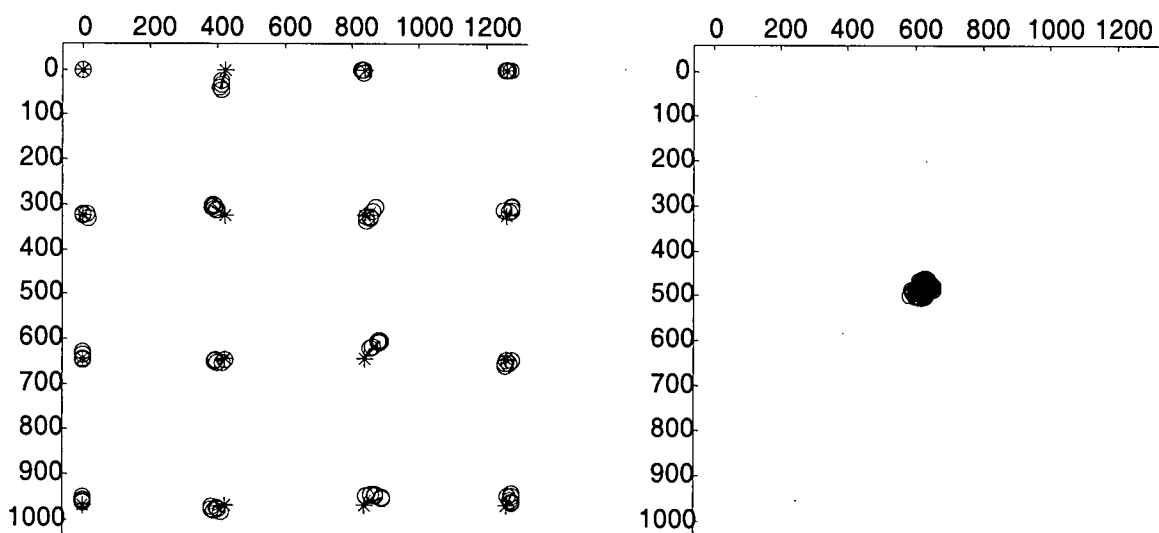


Figure 16-6 – 66.66 ms Shutter

16.3. LED Wavelength

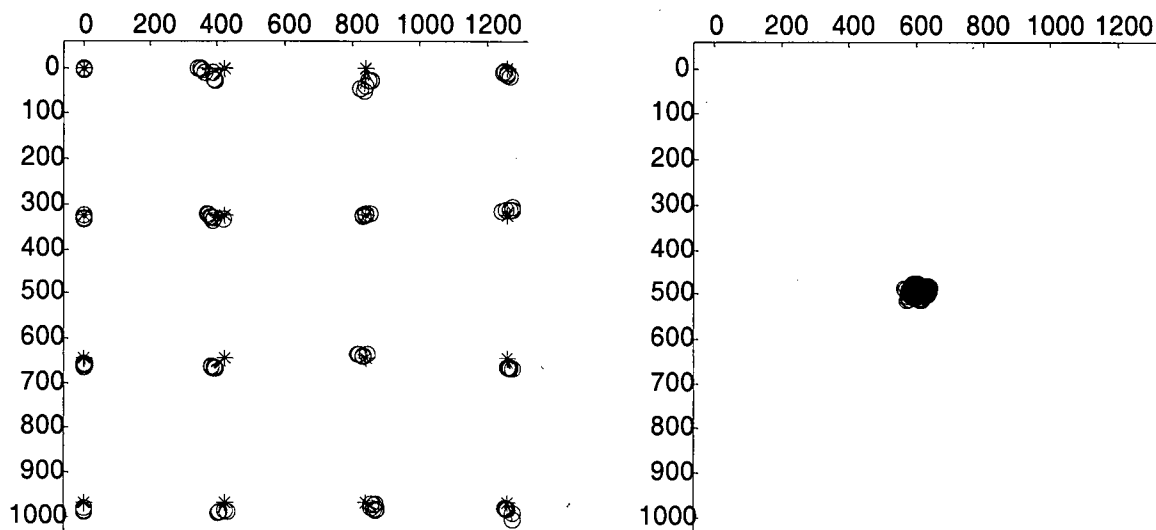


Figure 16-7 – 760 nm LEDS

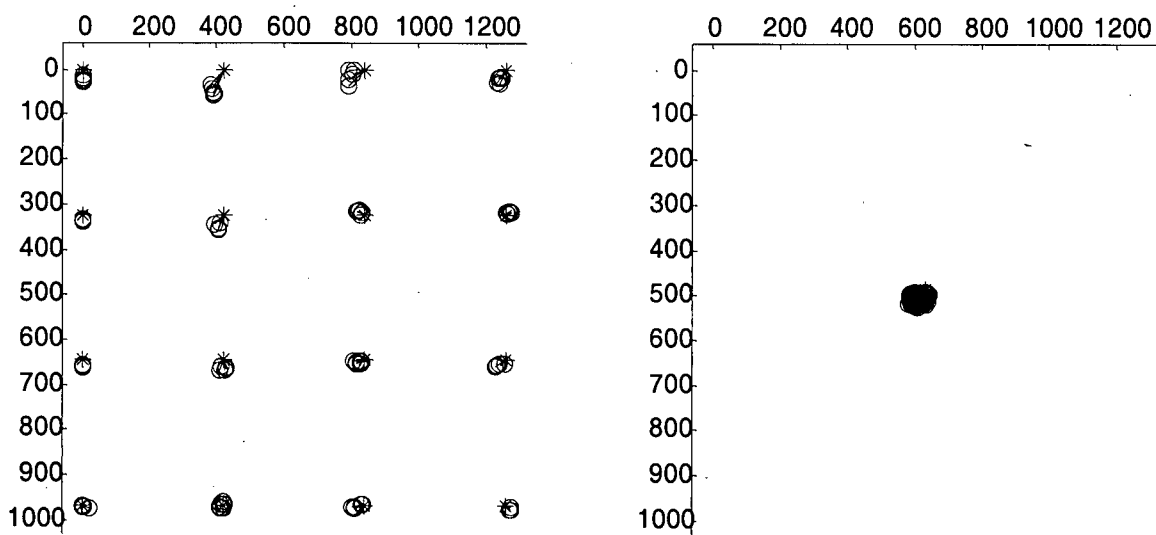


Figure 16-8 – 875 nm LEDs

16.4. Eye-Glasses

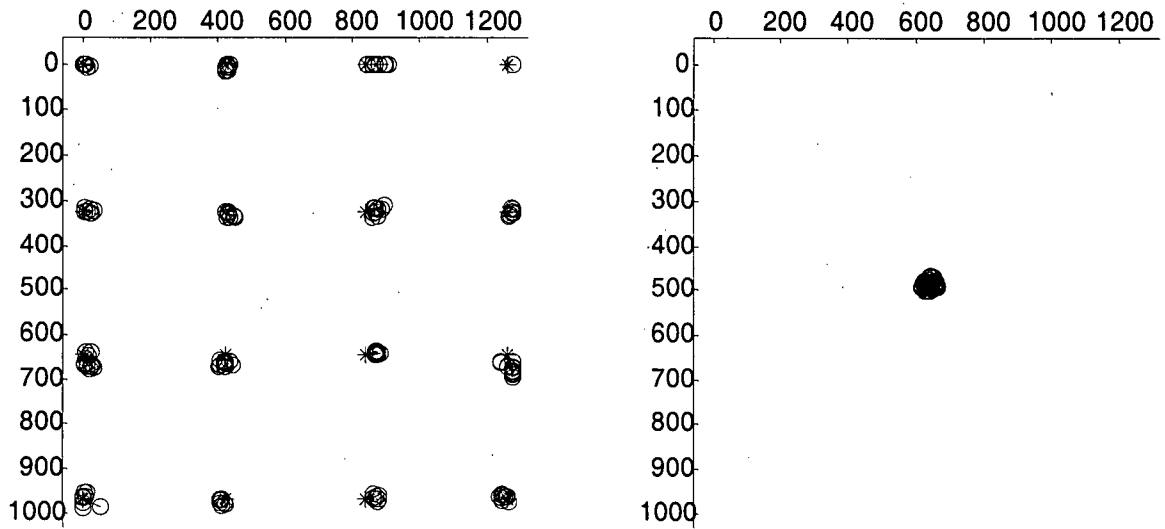


Figure 16-9 – No Glasses and No Correction

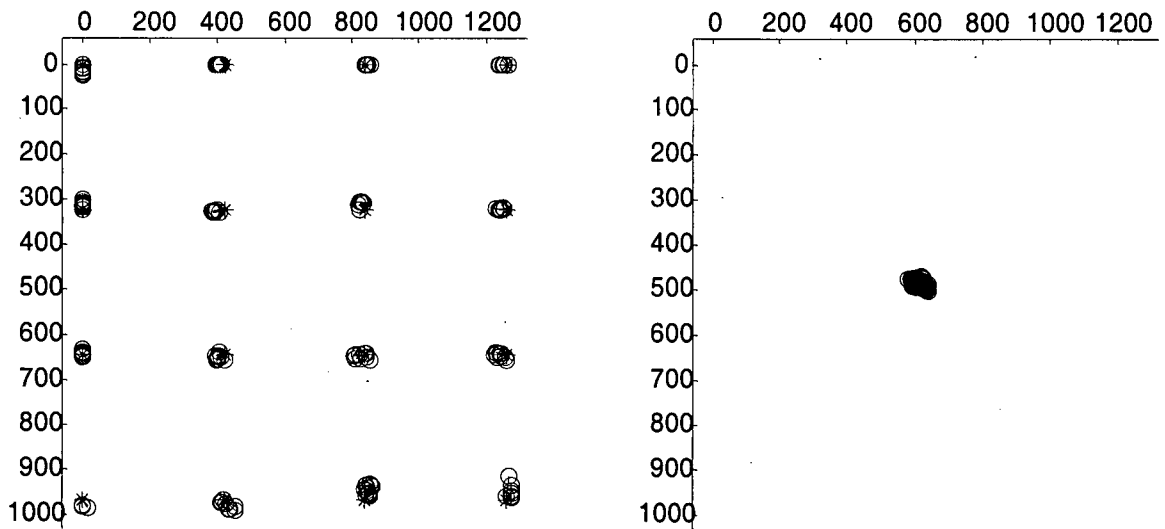


Figure 16-10 – Glasses and Correction

16.5. Ambient Lighting

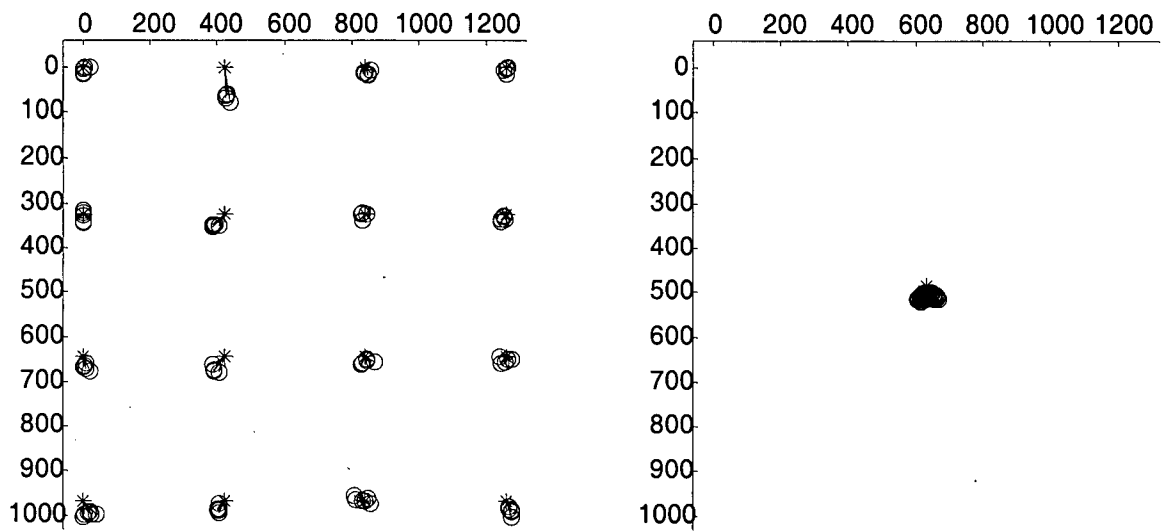


Figure 16-11 – No Sunlight

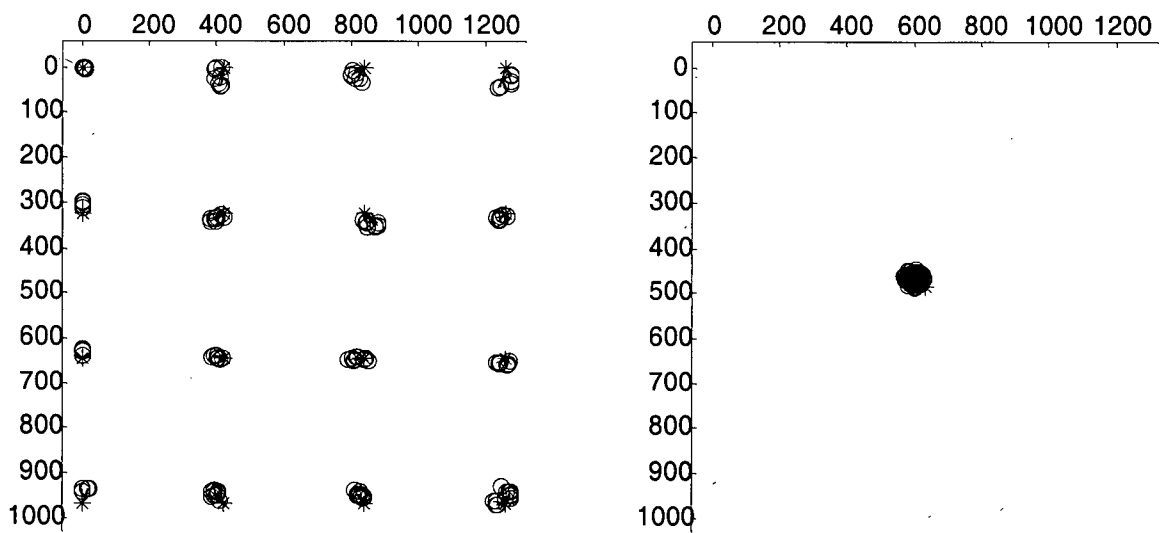


Figure 16-12 – Cloudy Day

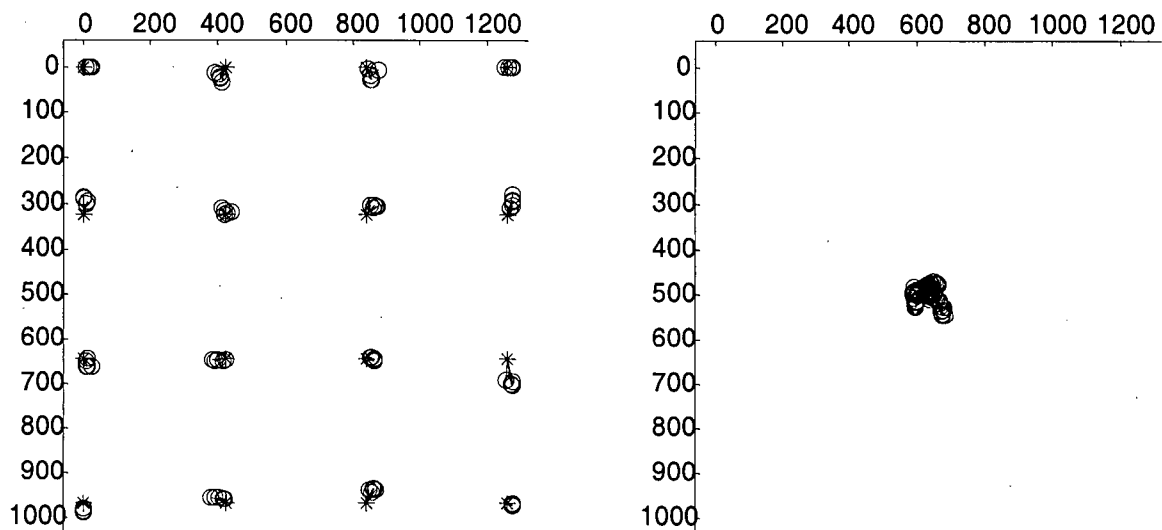


Figure 16-13 – Indirect Sunlight

16.6. Free Head

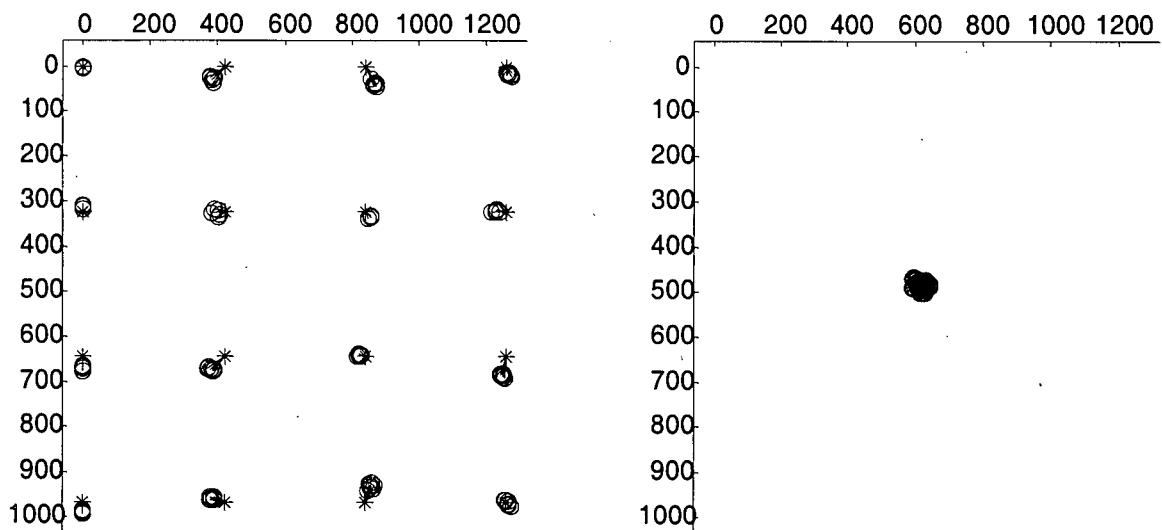


Figure 16-14 – Position 1

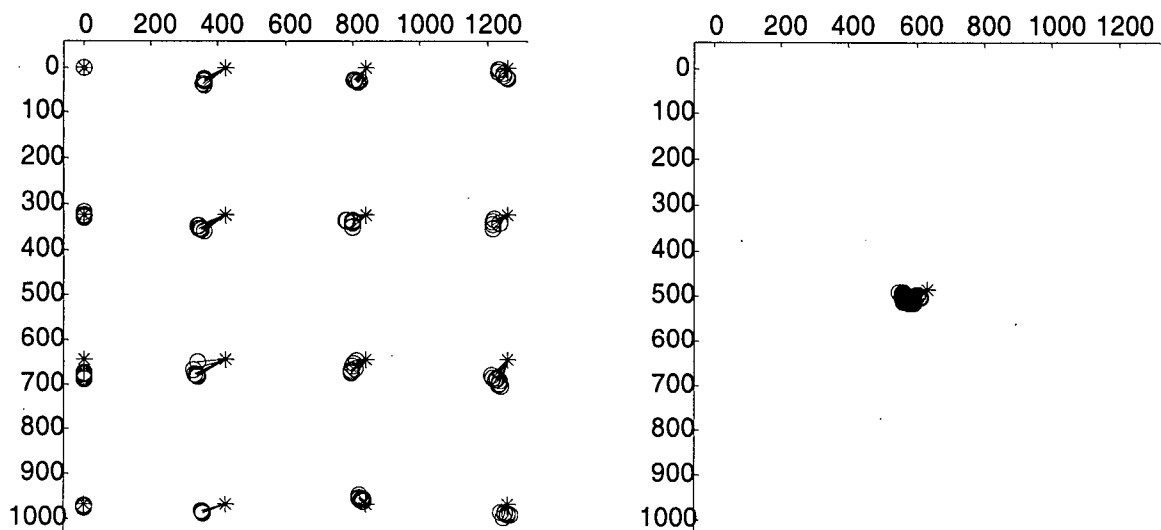


Figure 16-15 – Position 2

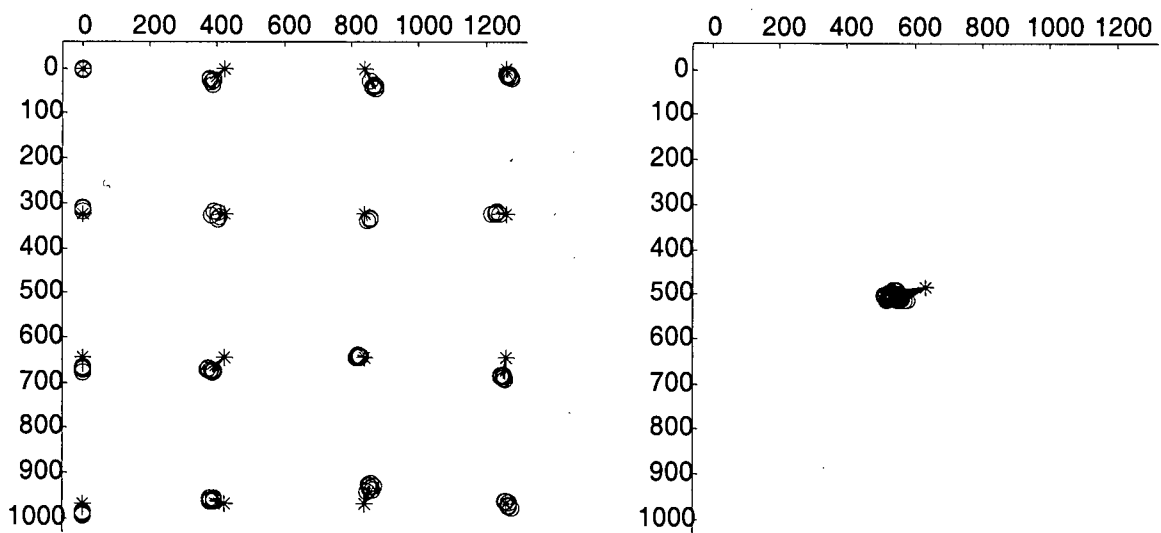


Figure 16-16 – Position 3

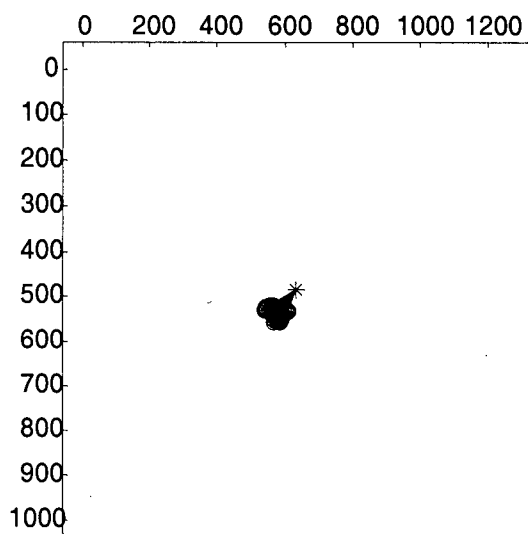
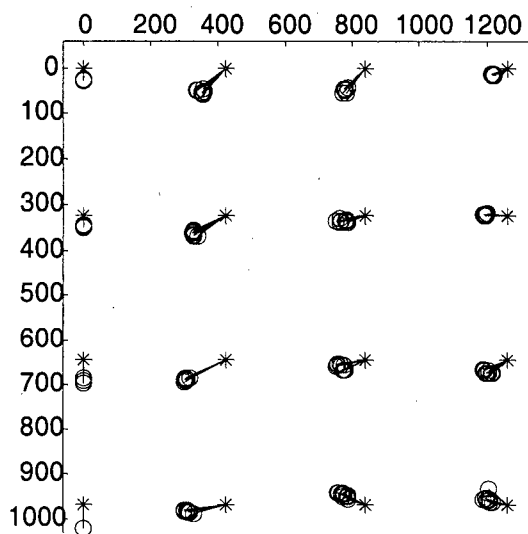


Figure 16-17 – Position 4

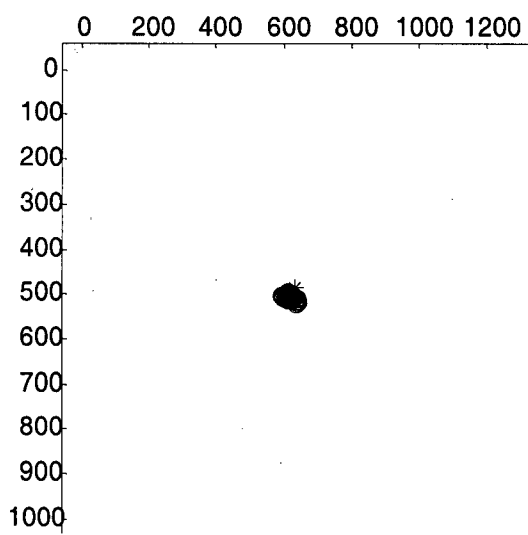
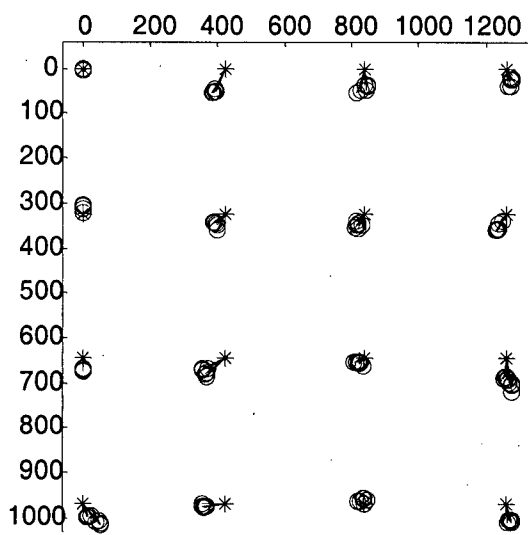


Figure 16-18 – Position 5

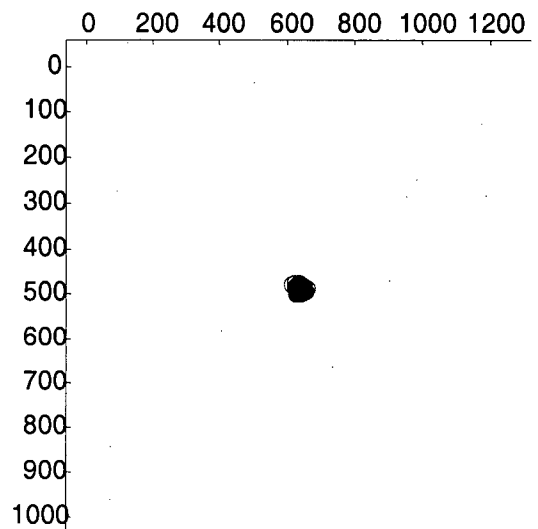
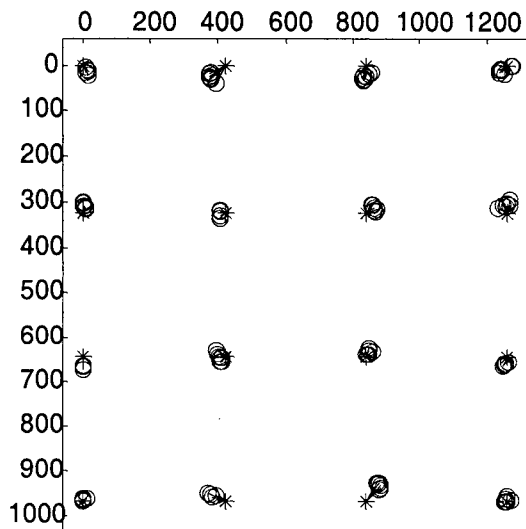


Figure 16-19 – Position 6

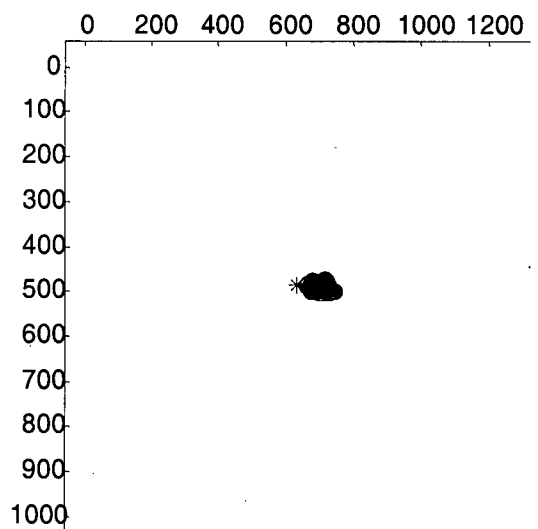
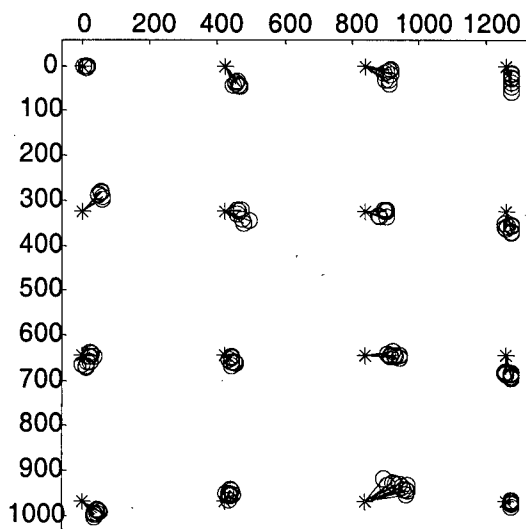


Figure 16-20 – Position 7

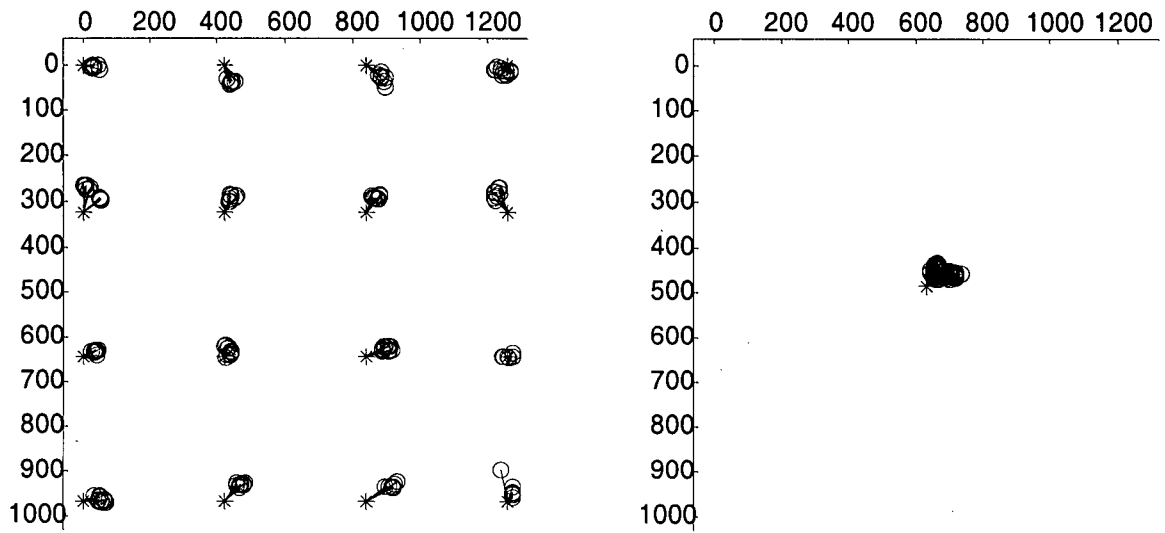


Figure 16-21 – Position 8

16.7. Multi User Trials

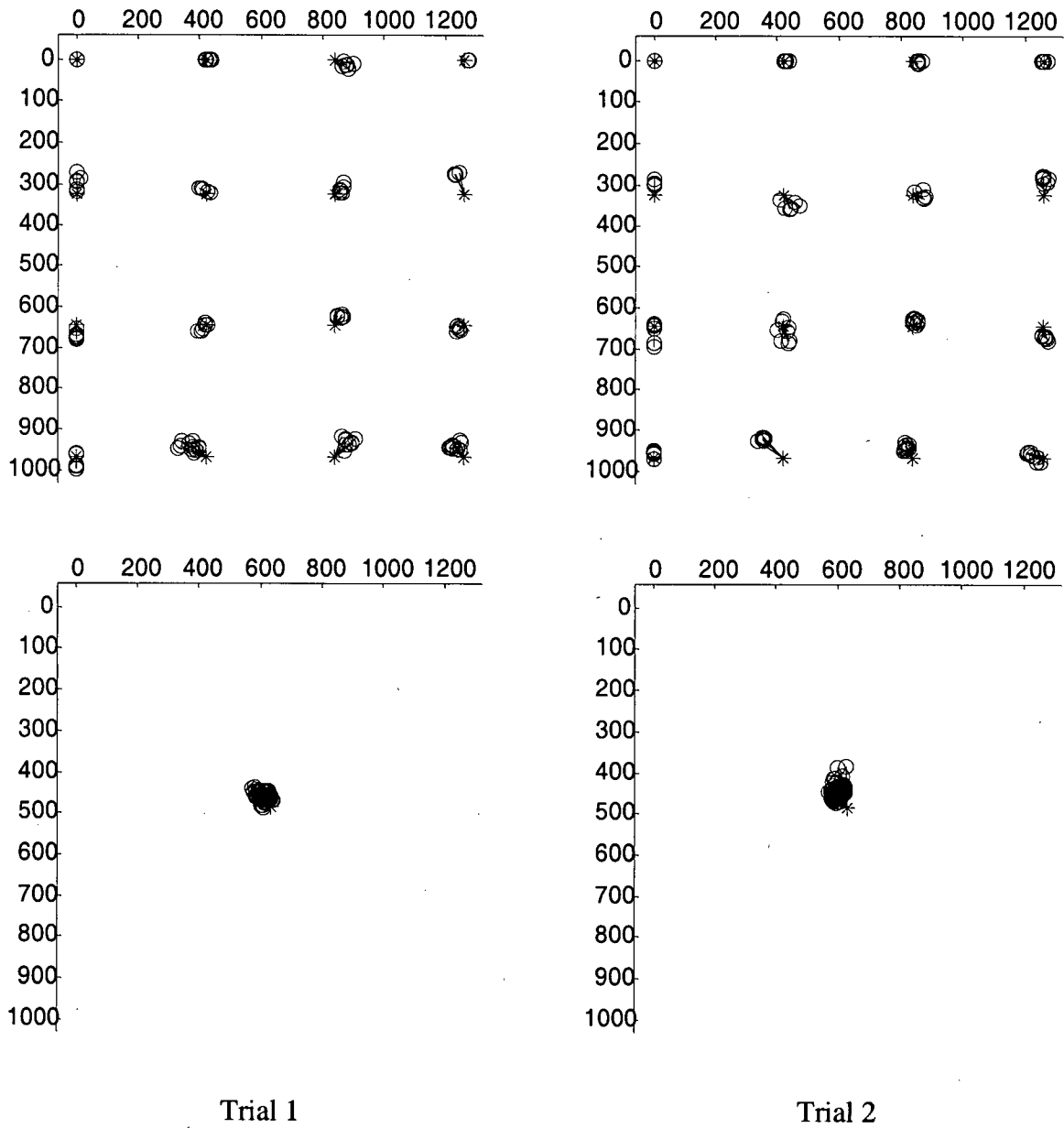
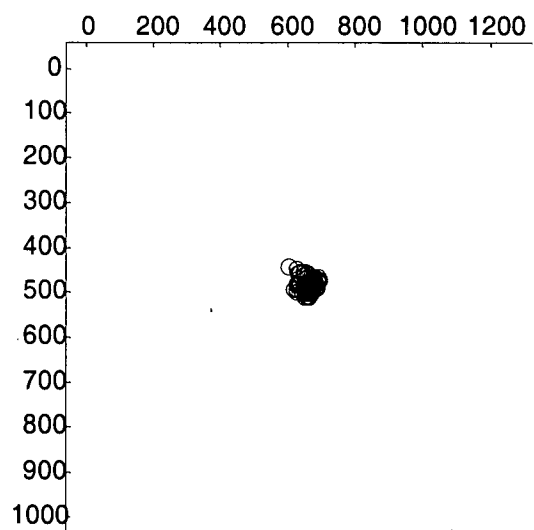
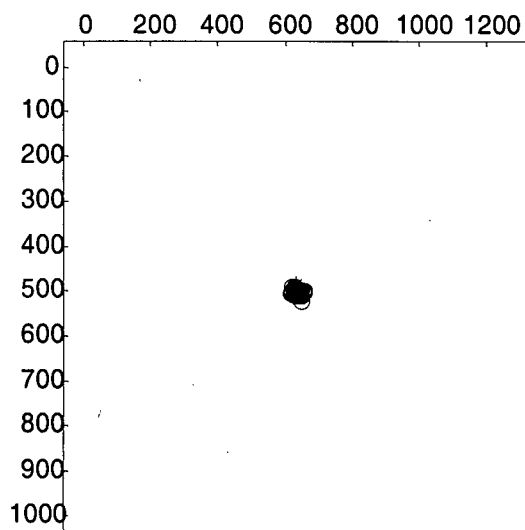
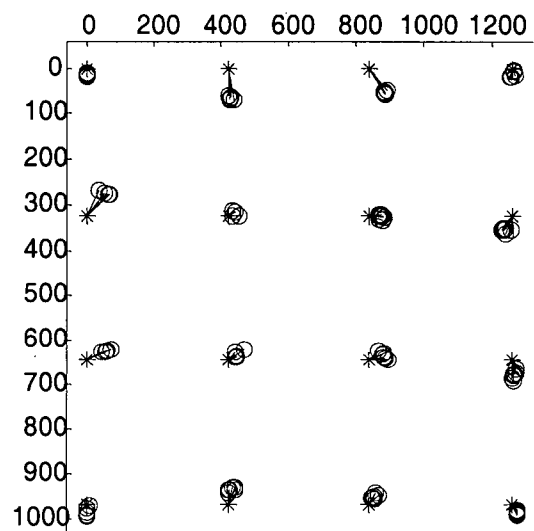
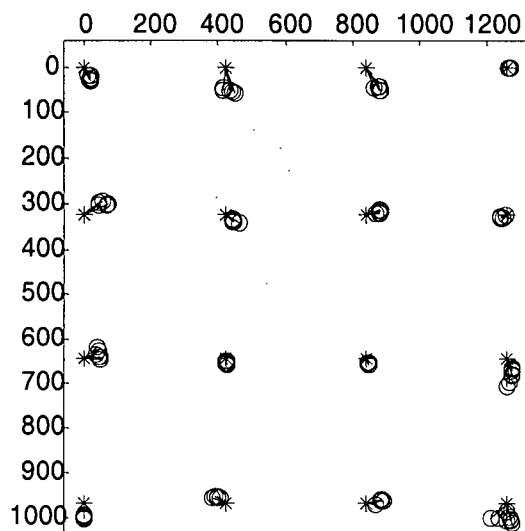


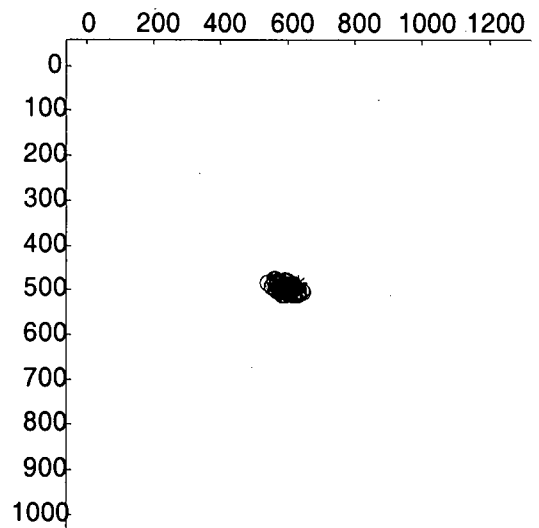
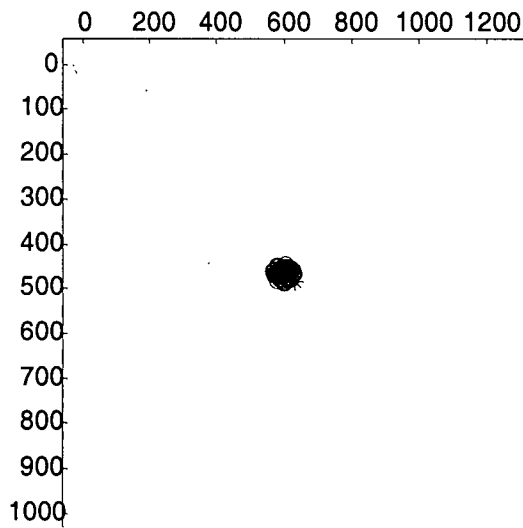
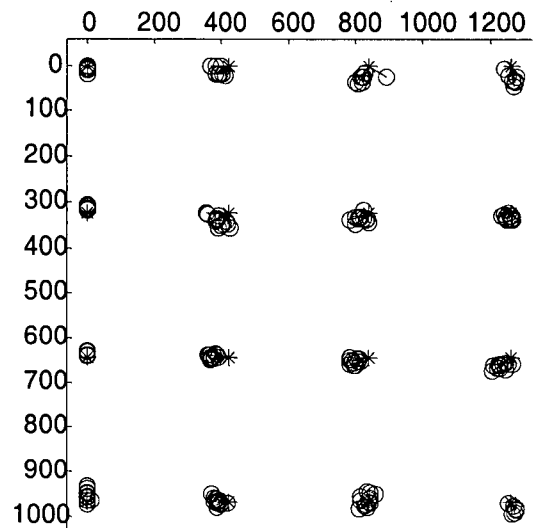
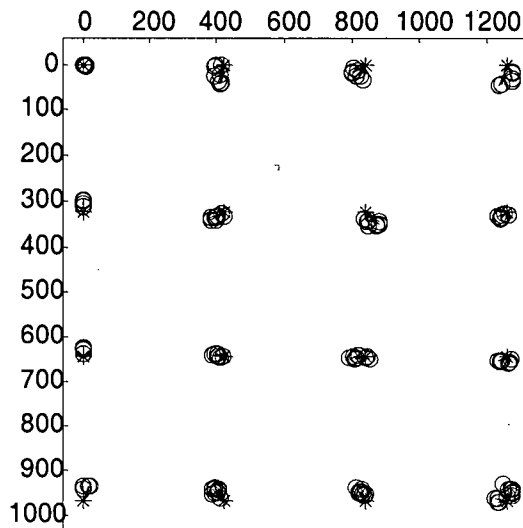
Figure 16-22 – Subject 1 – Trial 1 and 2



Trial 1

Trial 2

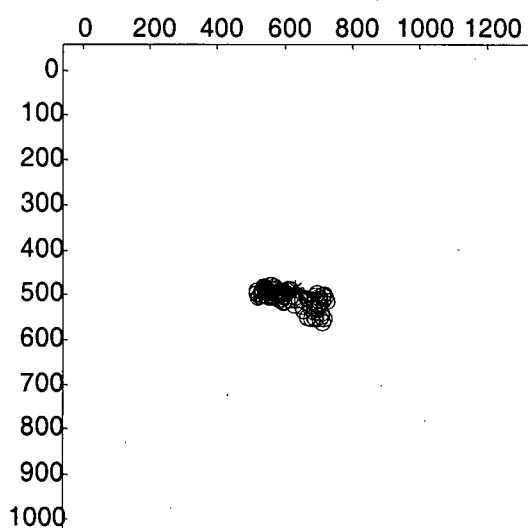
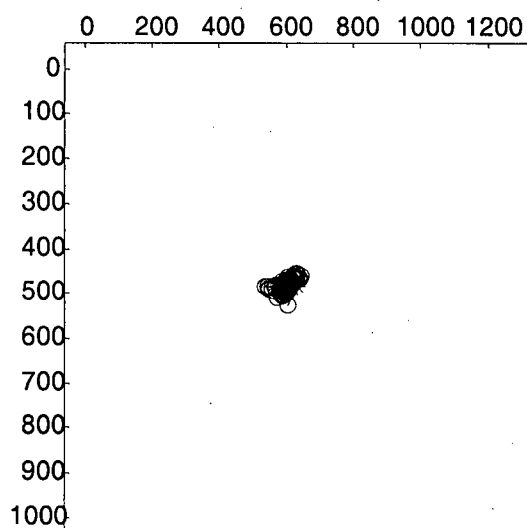
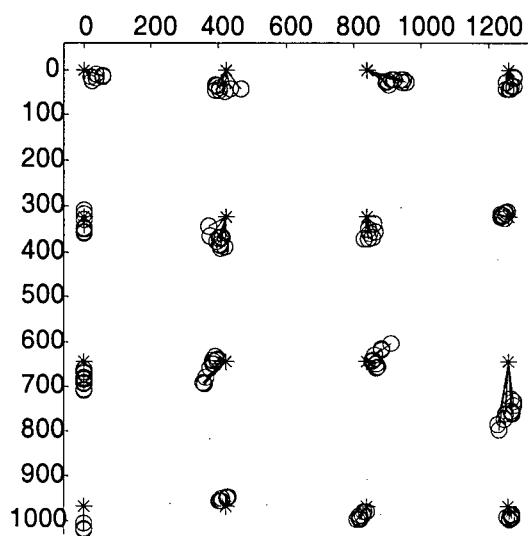
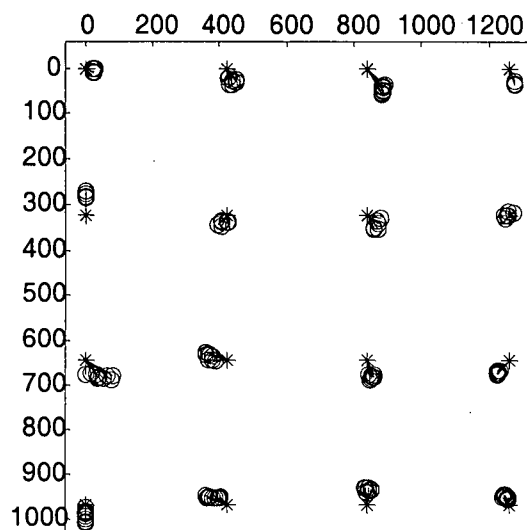
Figure 16-23 – Subject 2 – Trial 1 and 2



Trial 1

Trial 2

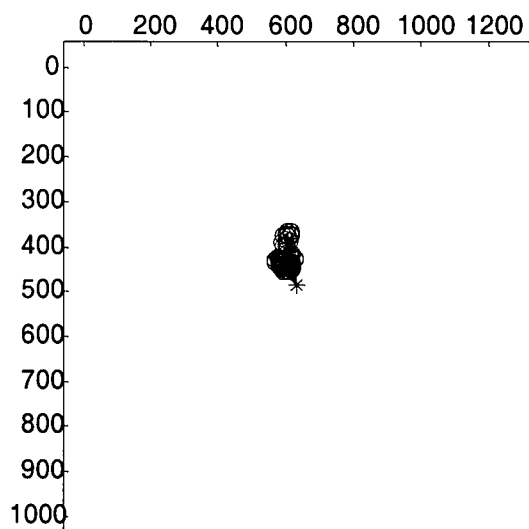
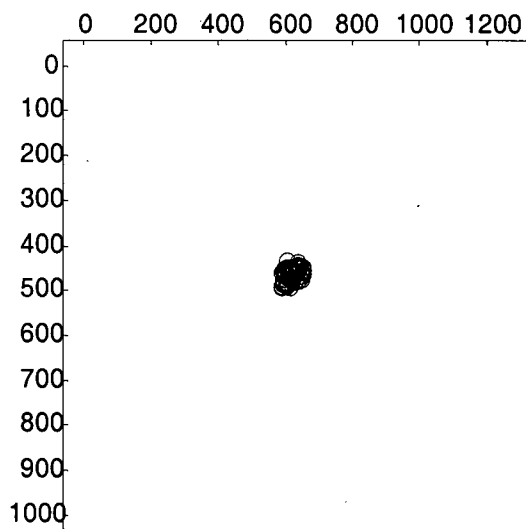
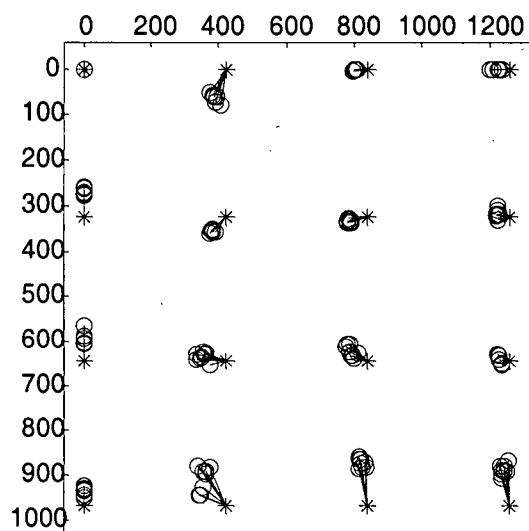
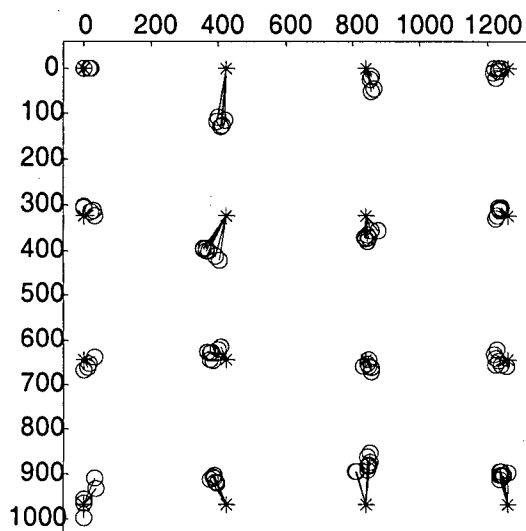
Figure 16-24 – Subject 3 – Trial 1 and 2



Trial 1

Trial 2

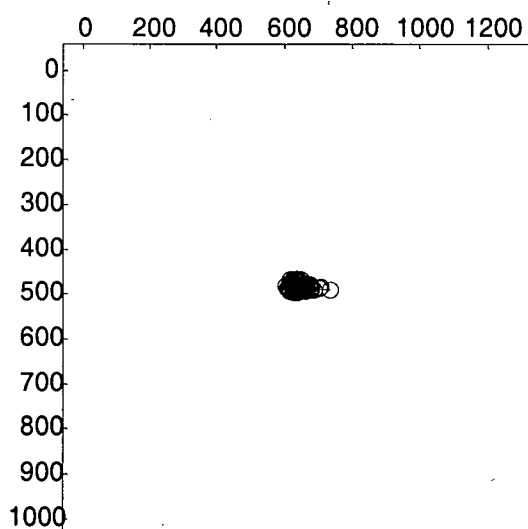
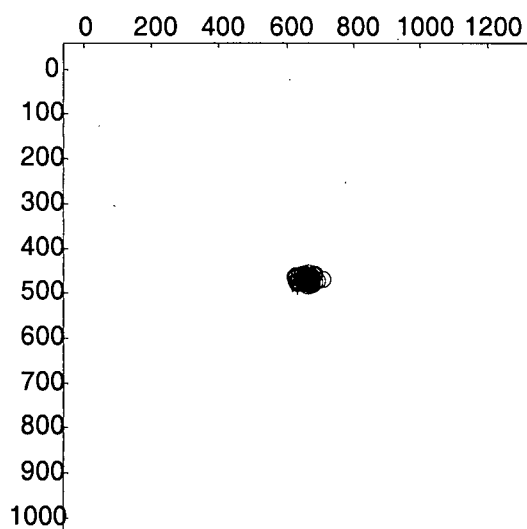
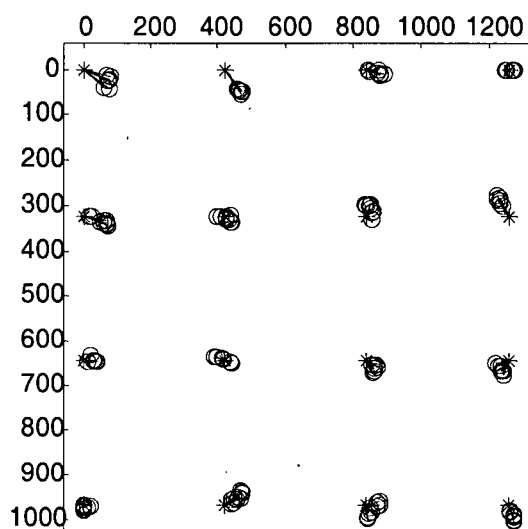
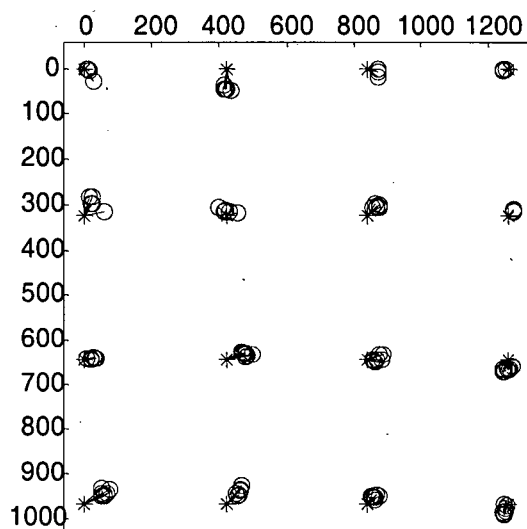
Figure 16-25 – Subject 4 – Trial 1 and 2



Trial 1

Trial 2

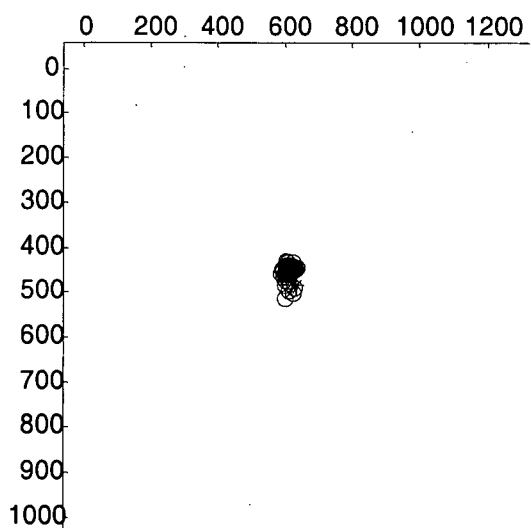
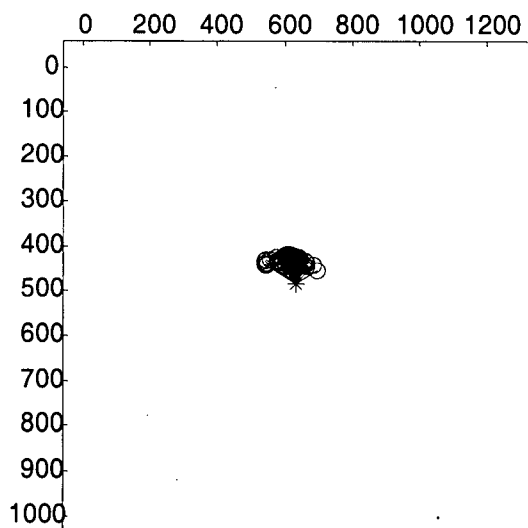
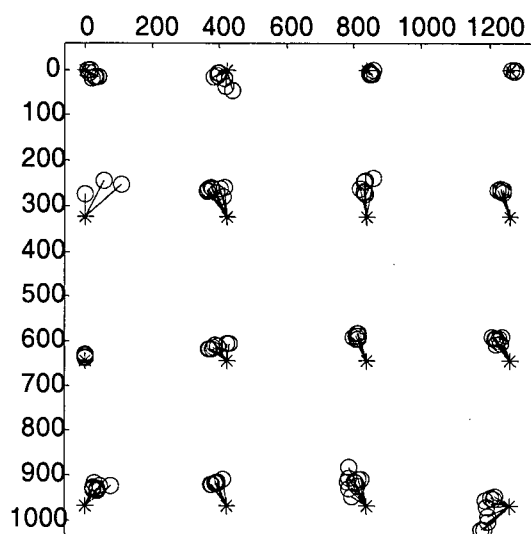
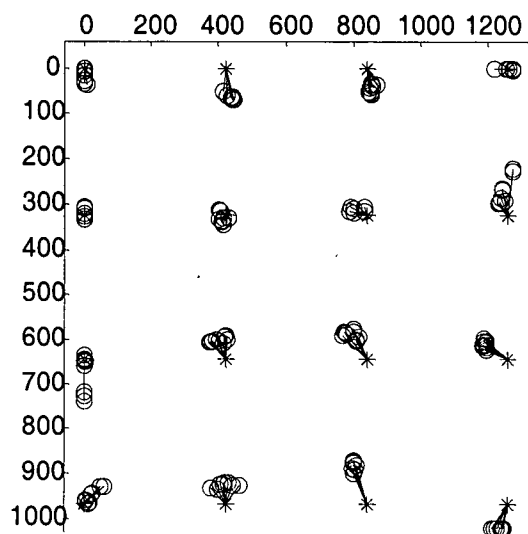
Figure 16-26 – Subject 5 – Trial 1 and 2



Trial 1

Trial 2

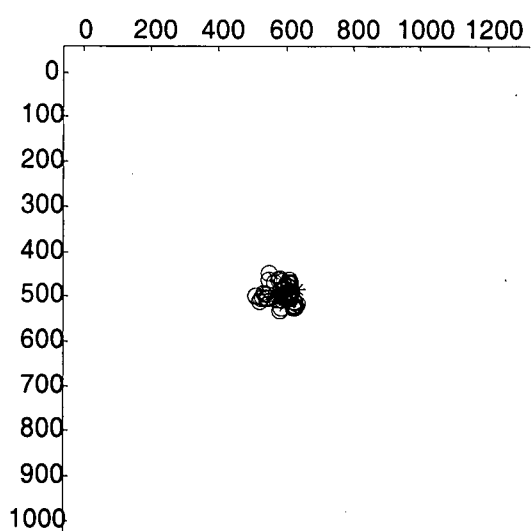
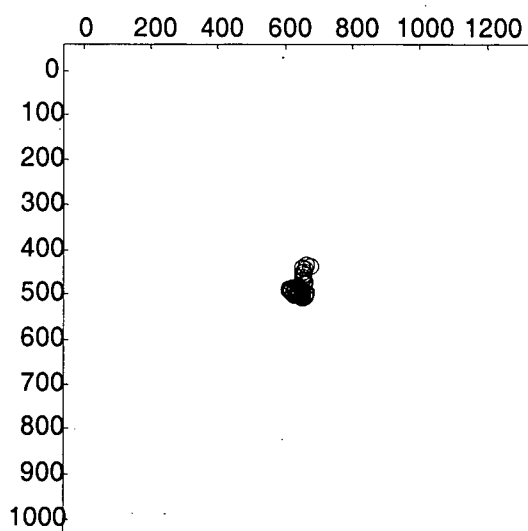
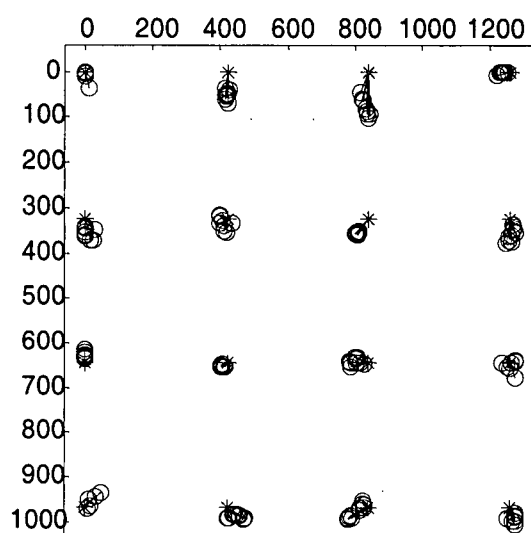
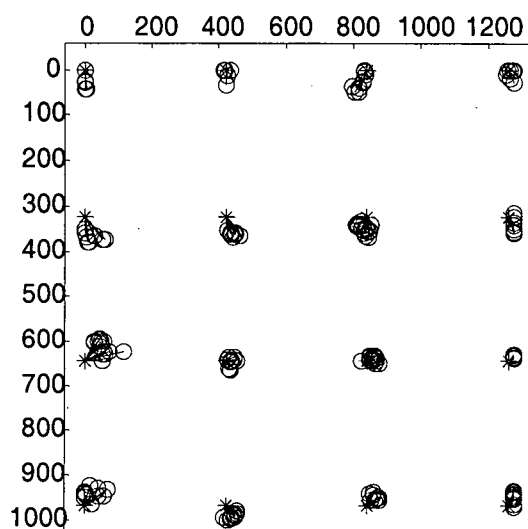
Figure 16-27 – Subject 6 – Trial 1 and 2



Trial 1

Trial 2

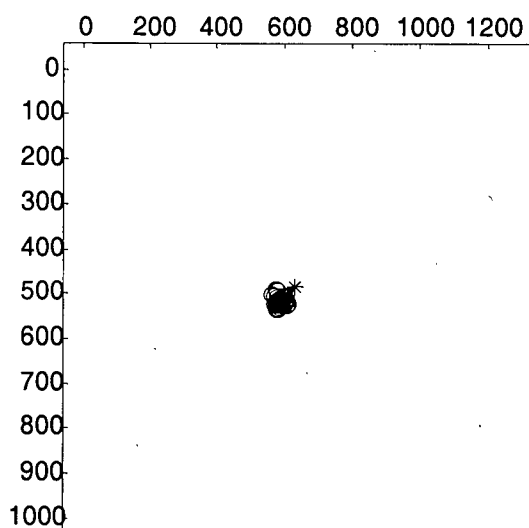
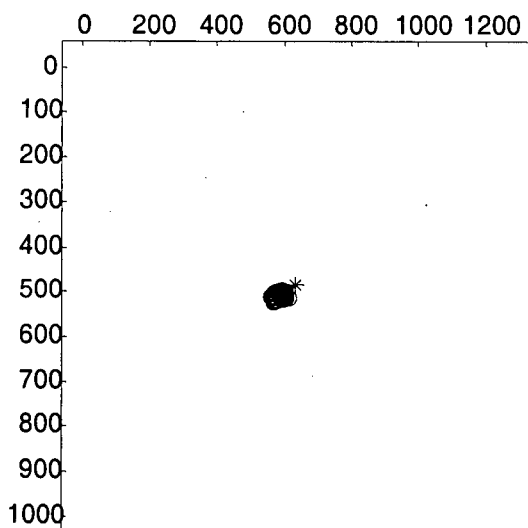
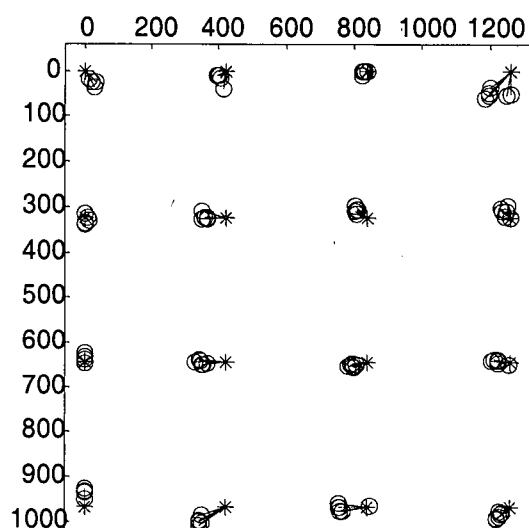
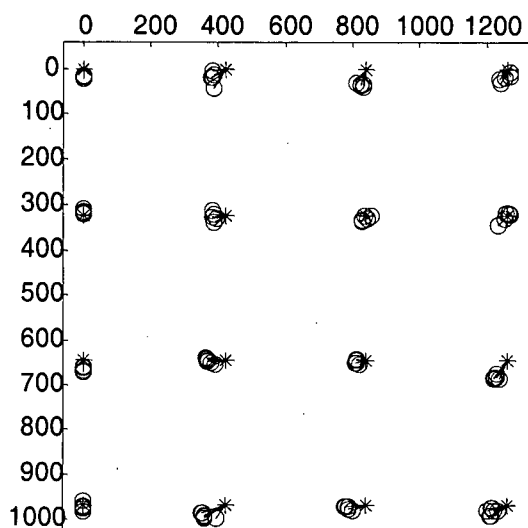
Figure 16-28 – Subject 7 – Trial 1 and 2



Trial 1

Trial 2

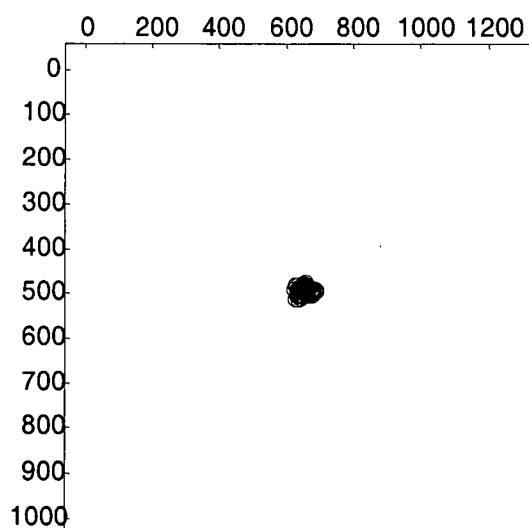
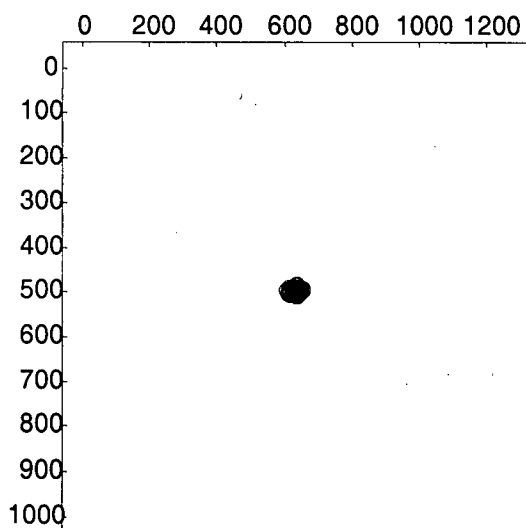
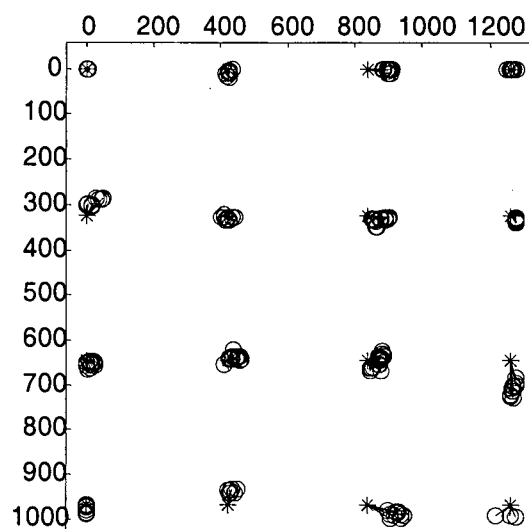
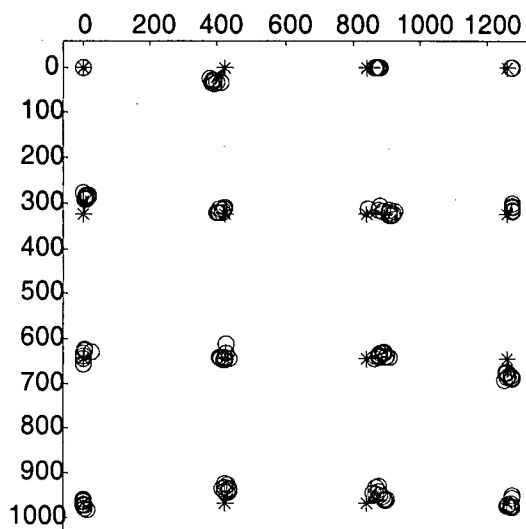
Figure 16-29 – Subject 8 – Trial 1 and 2



Trial 1

Trial 2

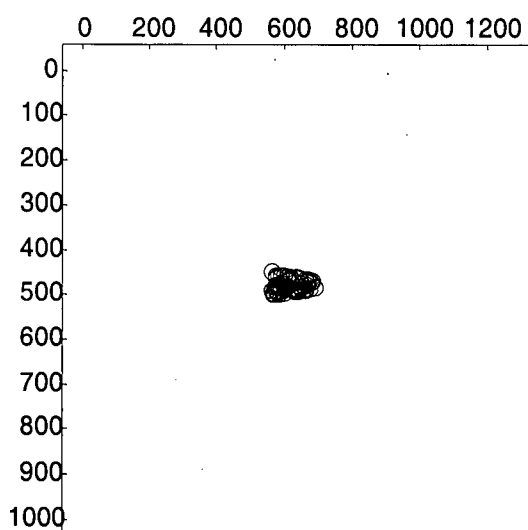
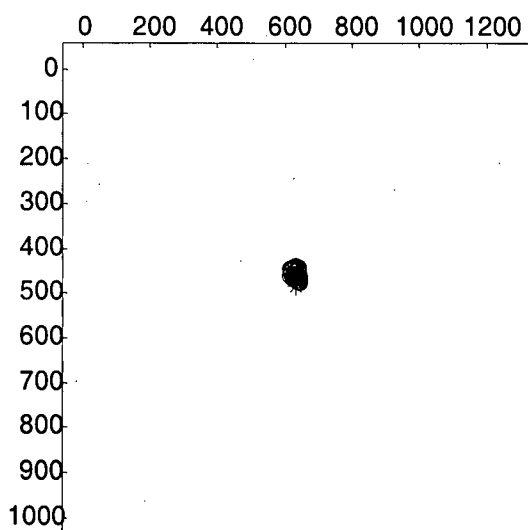
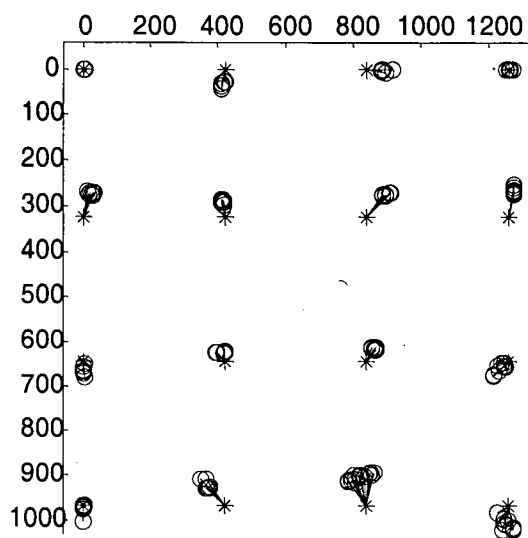
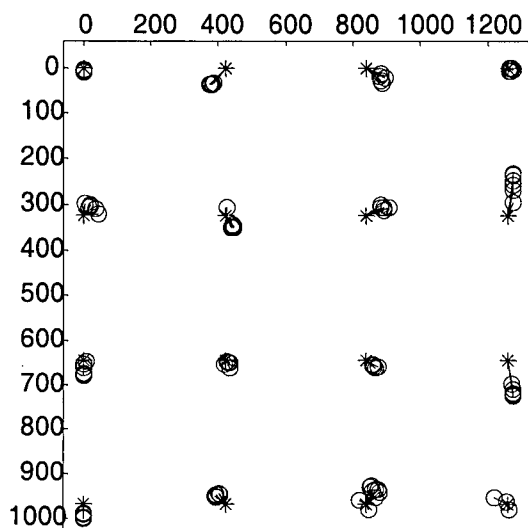
Figure 16-30 – Subject 9 – Trial 1 and 2



Trial 1

Trial 2

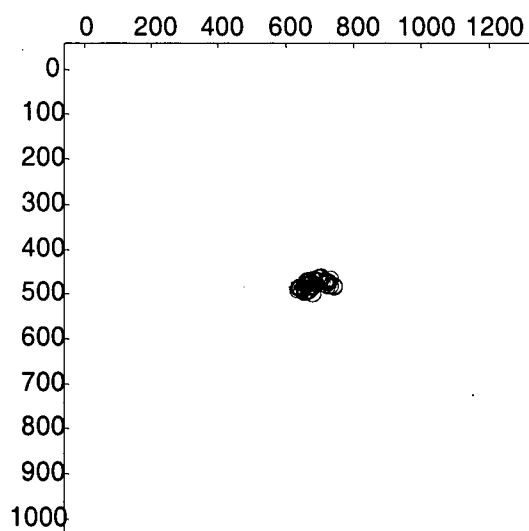
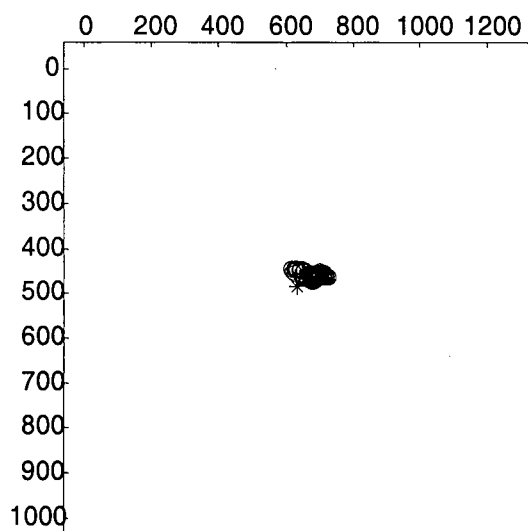
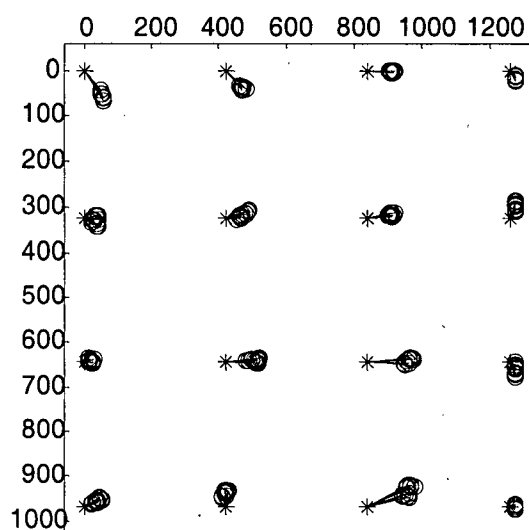
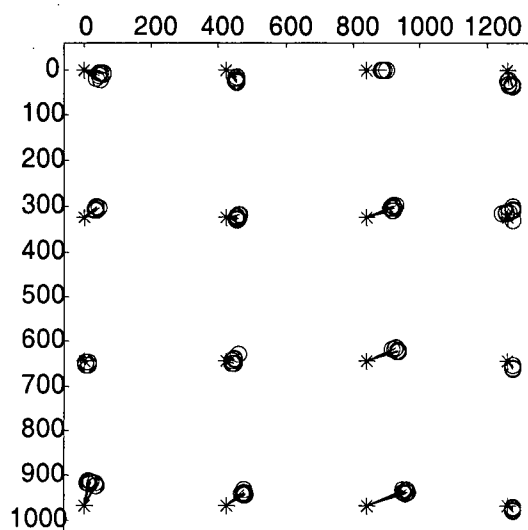
Figure 16-31 – Subject 10 – Trial 1 and 2



Trial 1

Trial 2

Figure 16-32 – Subject 11 – Trial 1 and 2



Trial 1

Trial 2

Figure 16-33 – Subject 12 – Trial 1 and 2