

A Non-Intrusive and Calibration-Free Gaze Tracking System

By

David Cang Trung Luong

A thesis submitted to the Faculty of Graduate Studies and Research in partial
fulfillment of the requirements for the degree of

Masters of Applied Science

in Biomedical Engineering

Ottawa-Carleton Institute for Biomedical Engineering

Department of Systems and Computer Engineering

Carleton University

Ottawa, Ontario, Canada

August 2012

Copyright © David C. Luong 2012



Library and Archives
Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

ISBN: 978-0-494-93515-6

Our file Notre référence

ISBN: 978-0-494-93515-6

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Abstract

In this thesis, a simple low cost eye gaze tracking system is presented which employs the cross ratios method for estimating eye gazes. A complete description of eye image collection and processing is explained. The effects of different eye image resolutions were examined using synthetically created 640x480 pixel eye images. There was some evidence that increasing eye image resolutions from 240x180 pixels to 640x480 pixels did not provide large gains in accuracy. Also examined was the effect of head position on the implemented system. It was observed that a right eye dominant user's head located 65cm from the monitor could move 10cm to the right, 17cm to the left, 5cm up and 5cm down and not greatly affect the accuracy of the eye gaze system. This movement translates to large head movements as the head can traverse across 63% of the monitor screen's width and 37% of the monitor screen's height. The limiting factor with head movement was the reliable extraction of eye features.

To facilitate the use of the system in non-cooperative environments, a simple bias compensation method is presented to eliminate a calibration routine from being preformed. The method reduced the average root mean square error of the estimated x-coordinate by 64.5 +/- 39.3 pixels (16.7 +/- 10.2mm) and the estimated y-coordinate by 53.1 +/- 14.7 pixels (13.7 +/- 3.8mm) on the monitor screen. This translates to a reduction of 28.5% +/- 17.2% for the x-coordinate and 25.7% +/- 6.6% for the y-coordinate of the eye gaze estimate. Current system limitations are also discussed that detail possible future work.

Acknowledgements

I would like to gratefully thank and acknowledge my supervisor, Dr. Adrian Chan, for his continued guidance and support throughout my graduate studies and completion of this thesis. His encouragement throughout this process has made the experience a worthwhile journey.

I would like to thank my wife, Lindsay, for her infinite patience over the past few years. She has been a constant source of support – emotional and moral – and it is because of her that I first became interested in the medical field.

Finally, I would like to thank my family and friends who have supported me through this process and helped push me to complete this thesis.

Table of Contents

ABSTRACT	III
ACKNOWLEDGEMENTS	IV
TABLE OF CONTENTS.....	V
LIST OF TABLES	VII
LIST OF FIGURES	VIII
LIST OF ABBREVIATIONS	XI
1 INTRODUCTION.....	1
1.1 INTRODUCTION TO EYE GAZE TRACKING AND ITS BIOMEDICAL APPLICATIONS	1
1.2 THESIS OBJECTIVES	2
1.3 CONTRIBUTIONS.....	2
1.4 ORGANIZATION OF THESIS.....	4
2 LITERATURE REVIEW.....	5
2.1 INTRODUCTION	5
2.2 EYE GAZE TRACKING IN LITERATURE	5
2.2.1 <i>Intrusive Eye Gaze Tracking Systems</i>	6
2.2.2 <i>Non-Intrusive Eye Gaze Tracking Systems</i>	7
2.2.3 <i>Commercial Eye Gaze Tracking Systems</i>	10
2.3 EYE GAZE ESTIMATION USING CROSS-RATIOS	11
2.4 CALIBRATION OF EYE GAZE TRACKING SYSTEMS.....	15
2.4.1 <i>Four Point Calibration</i>	15
2.4.2 <i>Single Point Calibration</i>	16
2.4.3 <i>Calibration Free</i>	16
2.5 BIOMEDICAL APPLICATIONS OF EYE GAZE TRACKING SYSTEMS	17
3 SYSTEM OVERVIEW	20
3.1 INTRODUCTION	20
3.2 SYSTEM HARDWARE SETUP	20
3.3 IMAGE ACQUISITION	22
3.4 IMAGE PROCESSING TO ESTIMATE EYE GAZE	25
3.4.1 <i>Estimating Pupil Center Location</i>	26
3.4.2 <i>Estimating Glint Locations</i>	31
3.4.3 <i>Estimating Gaze Point with Cross Ratios</i>	33
4 EFFECTS OF IMAGE RESOLUTION	34
4.1 INTRODUCTION	34
4.2 METHOD.....	34
4.2.1 <i>Data Collection</i>	34
4.2.2 <i>Creating Eye Images With Different Resolutions</i>	35
4.2.3 <i>Eye Gaze Estimation</i>	36
4.3 RESULTS	37
4.4 DISCUSSION	44
4.5 CONCLUSIONS	47

5	EFFECTS OF HEAD POSITION.....	49
5.1	INTRODUCTION	49
5.2	METHOD.....	49
5.2.1	<i>Eye gaze estimation</i>	51
5.3	RESULTS	51
5.4	DISCUSSION	56
5.5	CONCLUSIONS	58
6	BIAS COMPENSATION.....	60
6.1	INTRODUCTION	60
6.2	BIAS IN THE SYSTEM.....	60
6.3	BIAS COMPENSATION METHOD	62
6.4	RESULTS	65
6.5	DISCUSSIONS	74
6.6	CONCLUSIONS	77
7	THESIS SUMMARY AND FUTURE RECOMMENDATIONS	79
7.1	SUMMARY OF CONCLUSIONS	79
7.2	RECOMMENDATIONS FOR FUTURE WORK	80
	APPENDIX A: THE LETTER OF INFORMATION AND INFORMED CONSENT.....	83
	APPENDIX B: PORTA TEST PROTOCOL.....	92
	APPENDIX C: SYSTEM COMPARISON	93
	REFERENCES	95

List of Tables

TABLE 2-1: SUMMARY OF COMMERCIAL EYE GAZE TRACKING SYSTEMS	10
TABLE 3-1: SYSTEM COMPONENT COSTS.....	22
TABLE 3-2: PARTICIPANT INFORMATION	23
TABLE 4-1: RESIZE RESOLUTIONS OF EYE IMAGES.....	35
TABLE 4-2: THE NUMBER OF DISREGARDED 640X480 PIXEL EYE IMAGES FOR EACH TARGET FOR PARTICIPANTS P1 TO P5. A TOTAL OF 50 EYE IMAGES WERE COLLECTED FOR EACH TARGET..	38
TABLE 4-3: THE NUMBER OF EYE IMAGES WHERE A GHOST POINT WAS USED FOR THE EYE GAZE ESTIMATION USING 640X480 PIXEL EYE IMAGES FOR EACH TARGET FOR EACH PARTICIPANT P1 TO P5.	39
TABLE 4-4: THE NUMBER OF GHOST POINTS USED TO CALCULATE EYE GAZE ESTIMATION FOR EACH EYE IMAGE RESOLUTION USED. A TOTAL OF 400 EYE IMAGES WERE USED, 50 EYE IMAGES WERE COLLECTED FOR EACH TARGET AND INTERPOLATED INTO 7 OTHER DIFFERENT EYE IMAGE RESOLUTIONS.	40
TABLE 5-1: POSITIONS OF EACH PARTICIPANT'S HEAD RELATIVE TO THE CENTER OF THE MONITOR SCREEN.	51
TABLE 5-2: THE NUMBER OF DISREGARDED EYE IMAGES OF ALL TARGETS FOR EACH PARTICIPANT IN THE 9 DIFFERENT HEAD POSITIONS. A TOTAL OF 250 EYE IMAGES WERE COLLECTED IN EACH HEAD POSITION.	52
TABLE 5-3: THE NUMBER OF EYE GAZE ESTIMATES WHERE A GHOST WAS USED TO ESTIMATE ALL TARGETS FOR EACH PARTICIPANT IN EACH HEAD POSITION.....	52
TABLE 5-4: THE AVERAGE AND STANDARD DEVIATION OF THE RADIAL RMSE OF THE 9 TRIALS CONDUCTED FOR EACH PARTICIPANT FOR EACH TARGET.	54
TABLE 6-1: TABLE DISPLAYING THE AVERAGE REDUCTION IN PIXELS OF THE RMSE ALONG THE X-AXIS, Y-AXIS, AND RADIAL DISTANCE OF PARTICIPANTS P1 THROUGH P5 USING THE PROPOSED BIAS COMPENSATION METHOD USING 640X480 PIXEL EYE IMAGE RESOLUTION. A NEGATIVE VALUE EXPRESSES AN INCREASE IN ERROR.....	66
TABLE 6-2: TABLE SHOWING THE AVERAGE PERCENTAGE IN REDUCTION OF THE RMSE ALONG THE X-AXIS, Y-AXIS, AND RADIAL DISTANCE OF PARTICIPANTS P1 THROUGH P5 USING THE PROPOSED BIAS COMPENSATION METHOD FOR DIFFERENT HEAD POSITIONS. A NEGATIVE VALUE EXPRESSES AN INCREASE IN ERROR.....	74
TABLE C-1: TABLE COMPARING THE PROPERTIES OF THE IMPLEMENTED SYSTEM WITH THOSE THAT HAVE BEEN PREVIOUSLY BEEN IMPLEMENTED IN LITERATURE.	94

List of Figures

FIGURE 2-1: DIAGRAM OF EYE GAZE TRACKING SYSTEM IMPLEMENTATIONS AND THEIR CLASSIFICATIONS.6

FIGURE 2-2: DERIVATION OF THE GAZE DIRECTION BASED ON THE PUPIL CENTER AND PURKINJE IMAGE (BASED ON [19])8

FIGURE 2-3: THE RELATIONSHIP OF PLANES INVOLVED IN PROJECTIVE SPACE (BASED ON [6])..... 12

FIGURE 2-4: (A) POINTS USED FOR THE CROSS RATIO OF THE IMAGE PLANE, (B) POINTS USED FOR THE CROSS RATIO OF THE MONITOR PLANE. (BASED ON [6])..... 14

FIGURE 3-1: EYE GAZE SYSTEM COMPRISING OF A CANON PTZ CAMERA, NATION INSTRUMENTS USB DAC, 5 LED LIGHT SOURCES (L1 TO L5), AN LCD MONITOR AND PC (NOT SHOWN). . 21

FIGURE 3-2: SETTING UP THE CAMERA SO USER'S DOMINANT EYE FILLS BOUNDING BOX 23

FIGURE 3-3: CAPTURE IMAGES (A) DARK PUPIL, (B) BRIGHT PUPIL, AND (C) GLINT PUPIL 24

FIGURE 3-4: TARGET LOCATIONS PRESENTED TO USER DURING DATA COLLECTION (A) TARGET ORDER AND (B) SCREEN SHOT OF TARGETS AS THEY APPEARED TO THE PARTICIPANTS. 25

FIGURE 3-5: IMAGE PROCESSING DURING ESTIMATION OF APPROXIMATE PUPIL CENTER – (A) BACKGROUND REMOVAL, (B) BINARY THRESHOLDING, AND (C) MORPHOLOGICAL GRADIENT. . 27

FIGURE 3-6: HOUGH TRANSFORM DETECTING PUPIL. BLUE CIRCLES REPRESENT ALL THE CIRCLES FOUND BY THE HOUGH TRANSFORM IN THE IMAGE. THE RED RECTANGLE IS THE BOUNDING BOX USED IN SEARCHING FOR THE PUPIL CENTER. THE GREEN CROSS IS THE PUPIL CENTER FOUND 27

FIGURE 3-7: IMAGE PROCESSING STEPS EXECUTED TO DETERMINE THE TRUE PUPIL CENTER – (A) BACKGROUND REMOVAL, (B) BINARY THRESHOLDING, (C) MORPHOLOGICAL GRADIENT, AND (D) BRIGHT EYE IMAGE SHOWING APPROXIMATE PUPIL CENTER (GREEN CROSS) AND DETERMINED PUPIL CENTER (RED CROSS) 29

FIGURE 3-8: EYE GAZE ESTIMATES USING TWO DIFFERENT PUPIL DETECTION METHODS. THE RED X'S REPRESENT EYE GAZE ESTIMATES WHILE THE BLACK DIAMONDS REPRESENT THE TARGET POINTS. THE BLUE CROSSES SHOW THE MEAN EYE GAZE ESTIMATE OF THE CLUSTER, WHILE THE BLUE CIRCLES REPRESENT THE CLUSTER VARIANCE. (A) ESTIMATES USING THE HOUGH TRANSFORM, (B) ESTIMATES USING THE AREA OF INTEREST OPERATOR. 30

FIGURE 3-9: IMAGE PROCESSING STEPS TAKEN DURING ESTIMATION OF GLINT LOCATIONS – (A) BACKGROUND REMOVAL, (B) BINARY THRESHOLDING, (C) MORPHOLOGICAL GRADIENT, (D) FILLING OF "HOLES" IN IMAGE, (E) BOUNDING BOX, (F) ORGANIZATION OF GLINT LOCATIONS. 32

FIGURE 4-1: EYE IMAGE RESOLUTIONS WITH RESPECT TO ONE ANOTHER. THE LARGEST EYE IMAGE RESOLUTION USED WAS 1024x768 PIXELS WHILE THE SMALLEST EYE IMAGE RESOLUTION USED WAS 200x150..... 35

FIGURE 4-2: EXAMPLE OF AN IMAGE OF PARTICIPANT P1 WITH A BOUNDING BOX OUTLINING THE ROI. THE DIMENSION OF THE IMAGE IS 320 PIXELS WIDE BY 240 PIXELS TALL (AREA OF 76800 PIXELS). THE BOUNDING BOX IS 40 PIXELS WIDE BY 40 PIXELS TALL (AREA OF 1600 PIXELS OR $1600/76800 \times 100\% = 2.08\%$ OF THE EYE IMAGE)..... 36

FIGURE 4-3: GRAPH OF THE AVERAGE NUMBER OF DISREGARDED EYE IMAGES OF ALL PARTICIPANTS FOR ALL 15 TARGETS FOR EACH EYE IMAGE RESOLUTION. A TOTAL OF 750 EYE IMAGES WERE USED, 50 EYE IMAGES WERE COLLECTED FOR EACH TARGET IN 640x480 PIXEL RESOLUTION AND INTERPOLATED INTO ANOTHER 7 DIFFERENT EYE IMAGE RESOLUTIONS..... 39

FIGURE 4-4: BOX PLOTS SHOWING THE TREND OF THE AVERAGE RMSE OF THE RADIAL DISTANCE FROM THE TARGETS AS EYE IMAGE RESOLUTION INCREASES. DATA FOR EACH BOX PLOT CONSISTS OF THE AVERAGE RADIAL RMSE FOR EACH TARGET OF ALL PARTICIPANTS. EACH BOX REPRESENTS A DIFFERENT EYE IMAGE RESOLUTION USED. THE RED LINE IN THE BOX DENOTES THE MEDIAN OF THE DATA. THE TOP AND BOTTOM EDGES OF THE BLUE BOX REPRESENT THE 25TH AND 75TH PERCENTILES RESPECTIVELY. THE BLACK WHISKERS REPRESENT THE MAXIMUM AND MINIMUM DATA RANGE, WHILE THE RED CROSSES REPRESENT OUTLIERS. THE AVERAGE RMSE FOR EACH PARTICIPANT ARE REPRESENTED AN 'X' WHERE: GREEN = P1, BLACK = P2, CYAN = P3, MAGENTA = P4, AND BLUE = P5. 41

FIGURE 4-5: BOX PLOTS SHOWING THE TREND OF THE AVERAGE RMSE IN (A) X-AXIS AND (B) Y-AXIS AS EYE IMAGE RESOLUTION INCREASES. DATA FOR EACH BOX PLOT CONSISTS OF THE AVERAGE RMSE ALONG THE X-AXIS OR Y-AXIS FOR EACH TARGET OF ALL PARTICIPANTS. EACH BOX REPRESENTS A DIFFERENT EYE IMAGE RESOLUTION USED. THE RED LINE IN THE BOX DENOTES THE MEDIAN OF THE DATA. THE TOP AND BOTTOM EDGES OF THE BLUE BOX REPRESENT THE 25TH AND 75TH PERCENTILES RESPECTIVELY. THE BLACK WHISKERS REPRESENT THE MAXIMUM AND MINIMUM DATA RANGE, WHILE THE RED CROSSES REPRESENT OUTLIERS. THE AVERAGE RMSE FOR EACH PARTICIPANT ARE REPRESENTED AN 'X' WHERE: GREEN = P1, BLACK = P2, CYAN = P3, MAGENTA = P4, AND BLUE = P5. 42

FIGURE 4-6: THE VARIANCE OF RADIAL ERROR OF EYE GAZE ESTIMATES USING DIFFERENT EYE IMAGE RESOLUTIONS FOR (A) TARGET 1, (B) TARGET 5, (C) TARGET 8, (D) TARGET 11, AND (E) TARGET 15. THE RADIAL ERROR IS DEFINED BY THE CIRCLE THAT ENCOMPASSES THE MEAN EYE GAZE ESTIMATION (REPRESENTED BY A CROSS OR 'X'). THE FOLLOWING COLORS REPRESENT THE CORRESPONDING EYE IMAGE RESOLUTIONS IN PIXELS: SOLID BLUE = 200x150, DASH BLUE = 240x480, GREEN = 320x240, RED = 400x300, CYAN = 500x375, MAGENTA 640x480, YELLOW = 800x600, AND BLACK = 1024x768. 43

FIGURE 4-7: GRAPH SHOWING THE AVERAGE RADIAL VARIANCE OF ERROR OF EYE GAZE ESTIMATES FOR ALL TARGETS OF PARTICIPANTS P1 THROUGH P5 FOR EACH EYE IMAGE RESOLUTION USED. 44

FIGURE 5-1: COORDINATE SYSTEM USED WHEN DESCRIBING HEAD POSITION. 50

FIGURE 5-2: PLOTS (A) THROUGH (E) SHOW THE MEAN EYE GAZE ESTIMATES OF EACH TARGET FOR PARTICIPANTS P1 THROUGH P5 RESPECTIVELY OF ALL TRIALS 1 THROUGH 9. TARGET POSITIONS ARE REPRESENTED BY THE BLUE CIRCLES. EYE GAZE ESTIMATES FOR TRIALS 1 THROUGH 9 ARE REPRESENTED RESPECTIVELY AS FOLLOWS: BLUE X, GREEN X, RED X, CYAN X, MAGENTA X, YELLOW X, BLACK X, GREEN CROSS, AND RED CROSS..... 53

FIGURE 5-3: BOX PLOTS SHOWING THE AVERAGE RMSE IN THE X-AXIS AND Y-AXIS FOR DIFFERENT HEAD POSITIONS. DATA FOR EACH BOX PLOT CONSISTS OF THE AVERAGE RMSE ALONG THE X-AXIS OR Y-AXIS OF ALL PARTICIPANTS FOR ONE TARGET. THE RED LINE IN THE BOX DENOTES THE MEDIAN OF THE DATA. THE TOP AND BOTTOM EDGES OF THE BLUE BOX REPRESENT THE 25TH AND 75TH PERCENTILES RESPECTIVELY. THE BLACK WHISKERS REPRESENT THE MAXIMUM AND MINIMUM DATA RANGE, WHILE THE RED CROSSES REPRESENT OUTLIERS. 55

FIGURE 5-4: VENN DIAGRAMS SHOWING RELATIONSHIP BETWEEN THE LARGEST TRANSLATION RMS ERRORS (RED), SMALLEST TRANSLATION RMS ERRORS (GREEN) AND TARGETS (BLUE) OF (A) TARGET 1, (B) TARGET 5, (C) TARGET 8, (D) TARGET 11, AND (E) TARGET 15. 9x9 SQUARES AND CORRESPONDING NUMBERS REPRESENT THE PARTICIPANT'S HEAD POSITION RELATIVE TO THE SCREEN..... 57

FIGURE 6-1: SCHEMATIC DIAGRAM OF THE HUMAN EYE [45] 61

FIGURE 6-2: PICTURE OF PUPIL WITH FOUR PURKINJE REFLECTIONS. THE PARTICIPANTS' HEAD AND THE LIGHT SOURCES FOR THE REFLECTIONS ARE FIXED IN POSITION AND ONLY THE EYE IS ALLOWED TO MOVE. IN (A), THE PARTICIPANT'S GAZE IS ON A TARGET DIRECTLY AHEAD ON THE MONITOR SCREEN, WHILE IN (B) THE PARTICIPANT IS LOOKING AT A TARGET TO THE SIDE OF THE ORIGINAL TARGET. IN BOTH FIGURES: F REPRESENTS THE FOVEA, P_c REPRESENTS THE PUPIL CENTER, U_{v1} AND U_{v2} ARE TWO OF THE PURKINJE REFLECTIONS, G_e IS THE ESTIMATED GAZE POINT DETERMINED BY CROSS RATIOS, AND G_b IS THE TRUE GAZE POINT. 62

FIGURE 6-3: EYE GAZE DIAGRAM USED TO DETERMINE THE ROTATION θ OF THE EYE WHERE: R IS THE ESTIMATED RADIUS OF THE PUPIL, D_x IS THE DISTANCE ALONG THE X-AXIS BETWEEN THE X-COORDINATE OF THE CENTROID CREATED BY THE FOUR PURKINJE REFLECTIONS AND THE X-COORDINATE OF THE PUPIL CENTER, D IS THE DISTANCE BETWEEN THE USER AND THE MONITOR SCREEN, G_{ex} IS THE X-COORDINATE OF THE ESTIMATED GAZE POINT, G_{bx} IS THE X-COORDINATE OF THE ESTIMATED GAZE POINT WITH A BIAS CORRECTION ADDED, AND B_x IS THE BIASED DISTANCE NEEDED TO BE ADDED TO ESTIMATED GAZE POINT. THE EYE DIAGRAM USED TO DETERMINE THE ELEVATION ϕ IS SIMILAR TO THE ABOVE WITH D_x , G_{ex} , G_{bx} , AND B_x REPLACED WITH D_y , G_{ey} , G_{by} , AND B_y RESPECTIVELY TO REPRESENT THE Y-COORDINATE OF THE POINTS. . 65

FIGURE 6-4: EYE GAZE ESTIMATES FOR P1 USING 640x480 PIXEL EYE IMAGES. MAGENTA 'X' REPRESENT THE ORIGINAL EYE GAZE ESTIMATES. GREEN 'X' REPRESENT THE CORRECTED EYE GAZE ESTIMATES WITH BIAS COMPENSATION. 65

FIGURE 6-5: BOX PLOTS DISPLAYING THE TREND OF THE AVERAGE RMSE OF THE RADIAL DISTANCE FROM THE TARGETS AS EYE IMAGE RESOLUTION INCREASES. DATA FOR EACH BOX PLOT CONSISTS OF THE AVERAGE RADIAL RMSE FOR EACH TARGET OF ALL PARTICIPANTS. EACH BOX REPRESENTS A DIFFERENT EYE IMAGE RESOLUTION USED WITH (YELLOW BOX) OR WITHOUT (BLUE BOX) BIAS COMPENSATION WHERE 1 = 200x150, 2 = 240x180, 3 = 320x240, 4 = 400x300, 5 = 500x375, 6 = 640x480, 7= 800x600, AND 8 = 1024x768 PIXEL RESOLUTION. EYE IMAGE RESOLUTION BOX PLOTS FOR EYE GAZE ESTIMATES USING BIAS COMPENSATION ARE DENOTED WITH N (IE. 1B = BIAS COMPENSATION USING 200x150 PIXEL RESOLUTION). THE RED LINE IN THE BOX DENOTES THE MEDIAN OF THE DATA. THE TOP AND BOTTOM EDGES OF THE BOXES REPRESENT THE 25TH AND 75TH PERCENTILES RESPECTIVELY. THE BLACK WHISKERS REPRESENT THE MAXIMUM AND MINIMUM DATA RANGE, WHILE THE RED CROSSES REPRESENT OUTLIERS. 67

FIGURE 6-6: BOX PLOTS DISPLAYING THE TREND OF THE AVERAGE RMSE IN (A) X-AXIS AND (B) Y-AXIS AS EYE IMAGE RESOLUTION INCREASES. DATA FOR EACH BOX PLOT CONSISTS OF THE AVERAGE RMSE ALONG THE X-AXIS OR Y-AXIS FOR EACH TARGET OF ALL PARTICIPANTS. EACH BOX REPRESENTS A DIFFERENT EYE IMAGE RESOLUTION USED WITH (YELLOW BOX) OR WITHOUT (BLUE BOX) BIAS COMPENSATION WHERE 1 = 200x150, 2 = 240x180, 3 = 320x240, 4 = 400x300, 5 = 500x375, 6 = 640x480, 7= 800x600, AND 8 = 1024x768 PIXEL RESOLUTION. EYE IMAGE RESOLUTION BOX PLOTS FOR EYE GAZE ESTIMATES USING BIAS COMPENSATION ARE DENOTED WITH N (IE. 1B = BIAS COMPENSATION USING 200x150 PIXEL RESOLUTION). THE RED LINE IN THE BOX DENOTES THE MEDIAN OF THE DATA. THE TOP AND BOTTOM EDGES OF THE BOX REPRESENT THE 25TH AND 75TH PERCENTILES RESPECTIVELY. THE BLACK WHISKERS REPRESENT THE MAXIMUM AND MINIMUM DATA RANGE, WHILE THE RED CROSSES REPRESENT OUTLIERS. 68

FIGURE 6-7: THE AVERAGE PIXEL REDUCTION OF RMSE ALONG THE X-AXIS (BLUE), Y-AXIS (RED) AND RADIAL DISTANCE (GREEN). A NEGATIVE NUMBER DENOTES THAT THERE WAS AN INCREASE IN RMSE..... 69

FIGURE 6-8: GRAPH DISPLAYING THE AVERAGE RADIAL VARIANCE OF ERROR OF ALL TARGETS FOR ALL PARTICIPANTS P1 THROUGH P5 FOR DIFFERENT EYE IMAGE RESOLUTIONS. BLUE LINE REPRESENTS THE ORIGINAL EYE GAZE ESTIMATES WHILE THE RED LINE REPRESENTS THE BIAS COMPENSATED EYE GAZE ESTIMATES. 69

FIGURE 6-9: PLOTS (A) THROUGH (E) SHOW THE MEAN EYE GAZE ESTIMATES FOR EACH TARGET FOR PARTICIPANTS P1 THROUGH P5 RESPECTIVELY FOR ALL TRIALS 1 THROUGH 9. TARGET POSITIONS ARE REPRESENTED BY THE BLACK CIRCLES. ORIGINAL EYE GAZE ESTIMATES FOR TRIALS 1 THROUGH 9 ARE REPRESENTED AS A BLUE 'X' WHILE EYE GAZE ESTIMATES WITH BIAS COMPENSATION ARE REPRESENTED BY A RED 'X'. 71

FIGURE 6-10: BOX PLOTS DISPLAYING THE AVERAGE RMSE IN THE X-AXIS AND Y-AXIS FOR DIFFERENT HEAD POSITIONS WITH (YELLOW BOX) AND WITHOUT (BLUE BOX) BIAS COMPENSATION. DATA FOR EACH BOX PLOT CONSISTS OF THE AVERAGE RMSE ALONG THE X-AXIS OR Y-AXIS OF ALL PARTICIPANTS FOR EACH TARGET. THE RED LINE IN THE BOX DENOTES THE MEDIAN OF THE DATA. THE TOP AND BOTTOM EDGES OF THE BOXES REPRESENT THE 25TH AND 75TH PERCENTILES RESPECTIVELY. THE BLACK WHISKERS REPRESENT THE MAXIMUM AND MINIMUM DATA RANGE, WHILE THE RED CROSSES REPRESENT OUTLIERS. 73

List of Abbreviations

3D	Three Dimensional
ADC	Analog-to-Digital Converters
CCD	Charge Coupled Device
DAC	Data Acquisition
HCI	Human-Computer Interaction
IDE	Integrated Development Environments
IR	Infrared
JPEG	Joint Photographic Experts Group
LED	Light Emitting Diode
LCD	Liquid Crystal Display
PC	Personal Computer
PCI	Peripheral Component Interconnect
RMSe	Root Mean Square Error
USB	Universal Serial Bus
USD	United States Dollars

1 Introduction

1.1 Introduction to Eye Gaze Tracking and its Biomedical Applications

Eye gazes can provide a large amount of information about a person's attention and intention in everyday life. This information can be extracted and, in turn, utilized in different fields of research such as human-computer interaction (HCI), cognitive science, psycholinguistics, marketing research, and medical research. Because of these applications, there has been a large amount of interest in eye gaze tracking systems.

There are a number of potential biomedical applications for eye gaze systems which include: communication systems for people with high level physical disabilities or (loss of) muscle control [1], diagnosis and measurement of strabismus (also known as "lazy eye") in patients [2][3], and clinical user interface designs such as radiologist workstations [4] and control of endoscopic cameras [5].

A variety of different techniques, intrusive and non-intrusive, exist for estimating an individual's eye gaze. There are a wide range of hardware costs with differing levels of achievable accuracy and precision. Traditional intrusive techniques include equipment such as special contact lenses, electrodes with data acquisition systems, and head mounted devices, all of which provide accurate results even with user head movements. However, with the advancement of camera technology and computing power, non-intrusive vision-based techniques have evolved to allow user head movements while maintaining a level of accuracy.

The majority of commercial eye gaze tracking systems have been developed with human computer interaction in mind. However, these systems can be

expensive and deter their adoption into other applications. There exists a potential to develop a simpler, and therefore more cost effective system that would allow it to be tailored to more specific biomedical environments. This could, in turn, help expand the use of eye gaze tracking systems in the health care field.

1.2 Thesis Objectives

The overall motivation for this research is the implementation of a simple eye gaze tracking system for biomedical purposes, with the underlying goals of the system being cost effective and quickly adaptable for all users. We decided to focus on a non-intrusive eye gaze estimation technique based upon the use of cross-ratios. The system does not require any equipment to be worn by the user, nor will it produce any distracting light that may be seen. In this thesis, our objectives were to explore the effects of eye image resolutions and head movement on the system, and determine a method to eliminate the use of a personal calibration process that requires active user participation.

1.3 Contributions

The following is a list of thesis contributions with a brief description of each.

1. Implementation of a simple, low-cost non-intrusive eye gaze tracking system using a single camera:

This thesis presents an implementation of a simple and low-cost non-intrusive eye gaze tracking system based on the cross ratios method for eye gaze estimation. The system is comprised of a single camera instead of the two cameras that are more commonly found. Eye feature extraction methods are presented that include a comparison of the use of a Hough transform with an Area of Interest operator for pupil detection. This implementation shows that

it is a feasible alternative to more expensive commercially available eye gaze tracking systems.

2. Examination of the effect of eye image resolution on eye gaze accuracy and precision:

In this thesis, the effects of using different eye image resolutions on the eye gaze estimates are presented. Eye image resolutions were synthetically created using bicubic interpolation; however, using the eye image resolutions that were effectively downsampled, there was some evidence to suggest that increasing the eye image resolutions from 240x180 pixels to 640x480 pixels may not provide large gains in accuracy.

3. Examination of the effect of head position on eye gaze accuracy and precision:

In this thesis, the effects of different head positions on the eye gaze estimates are presented. It was observed that when the top bridge of the user's nose was located 65cm from the monitor, relatively large head movements, anywhere from 10cm to the right, 17cm to the left, 5cm up, and 5cm down, when facing the monitor did not greatly affect the accuracy of the eye gaze estimation system. It was also observed that the limiting factor of the system when accounting for head movement was the reliable extraction of eye features. With larger distances from the center of the monitor, eye features became more difficult to extract, resulting in the estimation of eye features or not estimating a gaze point all together.

4. Proposal and evaluation of a method for bias removal from eye gaze estimates to eliminate a personal calibration routine:

This thesis presents a method to eliminate the bias seen in the eye gaze estimates that would allow the eye gaze tracking system to be used without conducting a calibration routine. The method involved determining the rotation and elevation of a user's eye to account for the fact that the gaze line originated from the fovea and not the surface of the pupil. The method was able to reduce the average root mean square error (RMSe) of eye gaze estimates, but not completely eliminate the error. The bias compensation reduced the average RMSe of the estimated x-coordinate by 64.5 ± 39.3 pixels ($16.7 \pm 10.2\text{mm}$) and of the y-coordinate by 53.1 ± 14.7 pixels ($13.7 \pm 3.8\text{mm}$).

1.4 Organization of Thesis

This thesis consists of seven chapters including the current one. Chapter 2 presents a high-level review of the literature in the field of eye gaze tracking systems, specifically those based on feature-based approaches. Chapter 3 presents an overview of the system implemented, detailing the hardware setup and image acquisition and processing methods. In Chapter 4, the effects of eye image resolutions on the eye gaze estimates made using the cross-ratios method are discussed, while in Chapter 5, the effects of different head positions on the eye gaze estimates are examined. Chapter 6 presents the implementation and results of a bias compensation method for the system. Conclusions and recommendations for future work are discussed in Chapter 7.

2 Literature Review

2.1 Introduction

This chapter provides a high-level overview of eye gaze tracking systems currently implemented and described in the literature. Section 2.2 describes the basic principles of current methods that exist for eye gaze tracking, while Section 2.3 provides a detailed background on the theory of using cross ratios to estimate these eye gazes. Next, Section 2.4 looks at different methods of calibration and the limits they impose on users. Finally, Section 2.5 provides examples of where eye gaze tracking systems have been proven useful in biomedical applications.

2.2 Eye Gaze Tracking in Literature

Many different techniques currently exist for estimating eye gaze. These techniques have a wide range of hardware costs, accuracy, and limitations. Figure 2-1 summarizes the most common classification of eye gaze tracking technologies. At the most general level, the eye tracking technologies can be categorized as either being an intrusive or non-intrusive system. Intrusive technologies, where part of the system makes physical contact with the individual, include those that use sensors for electrooculography, head mounted systems, and contact lenses. Non-intrusive technologies, where no physical contact is made, are centered around remote base systems that have a camera or multiple cameras fixed to a location away from the user. Light sources may be used to produce reflections in the eye to allow for extraction of eye features.

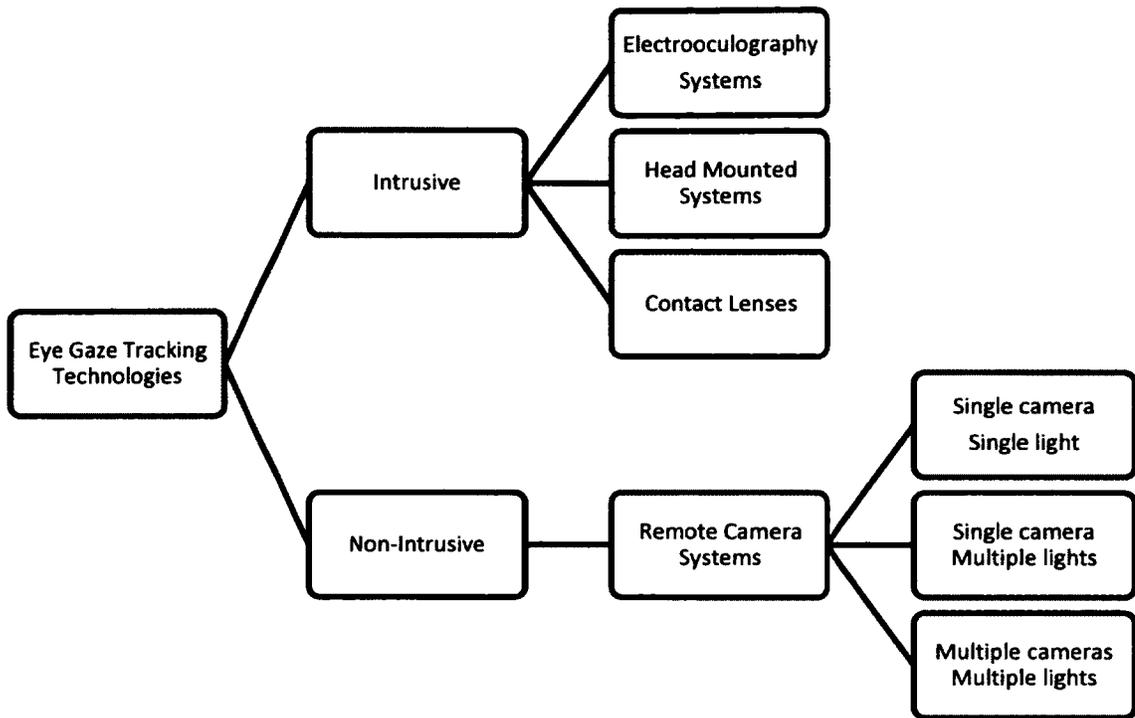


Figure 2-1: Diagram of eye gaze tracking system implementations and their classifications.

2.2.1 Intrusive Eye Gaze Tracking Systems

There has been much research into developing eye gaze systems that are considered intrusive. The three main systems are based on: 1) electrooculography, 2) head mounted devices, and 3) contact lenses. These eye gaze systems have been reported to be more accurate than non-intrusive methods as there is minimal effect from head and body movement.

Electrooculography research has been used to develop eye gaze estimation systems similar to those in [7]-[9]. These systems consist of electrodes being placed on the skin around the eye to measure the electrical potential generated by eye movements. This information can be used to form x-axis and y-axis vectors,

similar to the output of a computer mouse [7]. Head mounted systems have been developed mostly as a method for human computer interaction ([10]-[12]). The head mounted devices are video based and function similarly to remote camera systems. With remote camera systems, the camera is placed in a defined location, which allows the eye to move relative to it. However, with a head mounted system, the camera stays in a fixed position relative to the eye, thus eliminating any effects of head and body position. Contact lens based systems, initially developed by Robinson [13], function on the principle of electromagnetic induction. The contact lens is embedded with a coil (called a search coil) that induces a voltage from an externally driven electromagnetic field. This voltage is then measured through analog-to-digital converters (ADCs) to produce measurements for all three degrees of freedom, making the system very accurate [13]. Unfortunately, the major drawback with all these systems is that they can cause discomfort to the user, especially if the system is to be used for an extended period of time [6].

2.2.2 Non-Intrusive Eye Gaze Tracking Systems

Non-intrusive methods for eye gaze tracking are frequently remote camera based systems that rely on characteristics and properties of the eye. This is generally referred to as feature-based gaze estimation as the systems are extracting local features of the eye such as its contours and reflections from eye images. With this information, two main methods exist to process the information: 1) a model-based approach and 2) an interpolation-based approach [14]. Model-based approaches [1], [6], and [15] compute gaze direction by determining the intersection of the direction of gaze with the target that is being focused on. Interpolation-based approaches instead map eye features to gaze coordinates to parametric forms such as polynomials [16] or into nonparametric forms such as neural networks [17]. Both approaches are initialized into three types of hardware

systems for extracting eye features: 1) a system with a single camera and single light, 2) a system with a single camera and multiple lights, and 3) a system with multiple cameras and multiple lights.

2.2.2.1 Single camera and single light source systems

Early eye gaze tracking systems were composed of a single calibrated camera and a single light source due to its simplicity. Most interpolation-based approaches employ this configuration; however model-based approaches such as [18] and [19] have also used this configuration. To compute the eye gaze, these systems would capture pupil and Purkinje reflection images then use an eye model to compute the gaze direction. Purkinje reflections are several reflections that occur on the boundary between the cornea and lens of the eye [14]. These eye models consisted of population averages for eye size, shape, and position of the fovea. Personal calibration routines would also need to be conducted prior to use to obtain accurate eye gaze estimates. Referring to Figure 2-2, Ohnu et al. [19] describe that one could compute the center of cornea curvature from the Purkinje image. This, coupled with the center of the pupil, would be able to compute the gaze direction which would then be able to translate to the gaze position.

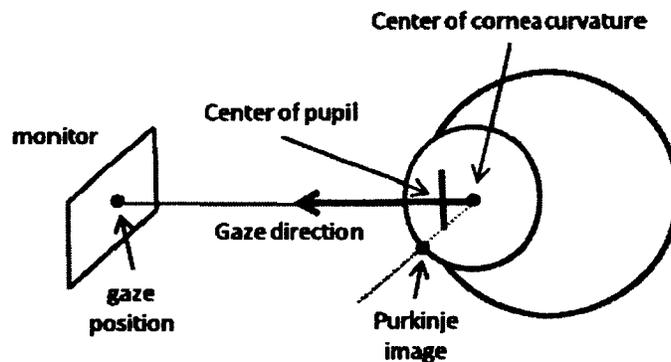


Figure 2-2: Derivation of the gaze direction based on the pupil center and Purkinje image (based on [19])

Unfortunately, the limitation of this method was that the head could not move during use. This was to ensure that the difference vector between the pupil and Purkinje reflection remained constant as the location of the Purkinje image would change when the head position would change.

2.2.2.2 Single camera and multiple light source systems

Similar to the previous method, adding light sources to the setup greatly improved the head position invariance for eye gaze estimations. Implementations from [15], [18], [20], [21], [22], and [23] use a model where the cornea center is better estimated from two or more light sources and allow for small amounts of head movement. Users still need to go through a calibration routine; however, it has been simplified to the point where only one calibration point, sometimes none, may be needed to estimate certain properties of the eye. These methods sometimes still require a calibrated camera and have been found to be difficult to remain accurate when changing geometric properties of the system such as its light sources, camera, or monitor screen.

2.2.2.3 Multiple camera and multiple light source systems

The limitation of a single camera system is that there is a trade off between the amount of head movement that is tolerable and the resolution of the eye images captured. To allow for large amounts of head movement, the camera requires a large field of view, while for high resolution images of the eye, a narrower field of view is used to enable the camera to zoom in on the user's eye. To overcome this limitation, some systems such as [1] and [6], have employed multiple cameras where one camera could be used to track the head while a second camera would be used for close-up eye images. One of the most popular methods was proposed by Yoo and Chung [6], which made use of cross ratios. In their implementation, one

camera was used to track the head in order to control a second camera, which would obtain a high resolution image of the eye. The first camera was not they were able to use the cross-ratio of the four light sources from the captured eye image to estimate gaze points. This will be explained in more detail in Section 2.3.

Other systems, such as [24] and [25], make use of multiple cameras as a stereo camera configuration to use a 3D model of the eye to estimate eye gaze. Based on fitting projected model features to extracted image features from the cameras, the visual axis can be estimated. The intersection of the visual axis onto the monitor screen is then calculated to produce eye gaze estimation coordinates.

2.2.3 Commercial Eye Gaze Tracking Systems

There are a number of commercially available vision-based eye gaze tracking systems on the market. These systems are used mainly for assistive technology; however, they contain integrated development environments (IDE) that can be used to tailor the device for any desired application on a computer platform. These systems range in price from \$5,000 USD to \$16,000 USD. The exact methods of eye gaze estimation of the commercial systems listed are unavailable from the manufacturers, but appear to be vision-based eye gaze trackers.

Table 2-1: Summary of Commercial Eye Gaze Tracking Systems

Company	Model	Type	Cost
Mirametrix [26]	S2 Eye Tracker	Non-intrusive	\$5,000 USD
LC Technologies Inc [27]	Eyegaze Edge Desktop	Non-intrusive	\$8,700 USD
Tobii [28]	Tobii C15	Non-intrusive	\$8,395 USD
DynaVox [29]	Vmax + EyeMax	Non-intrusive	\$8,470 USD
H.K. Eye Can Ltd. [30]	VisionKey 7.3	Intrusive	\$6,695 USD
EyeTech Digital Systems [31]	Gaze Tracker	Non-intrusive	\$5,000 USD
Arrington Research [32]	Turnkey HMD (B-HMZ1-U220)	Intrusive	\$15,998 USD

2.3 Eye Gaze Estimation Using Cross-Ratios

Yoo and Chung [6] proposed a non-intrusive method for eye gaze estimation based on the reflection of light. This method involved extracting eye features from video images and using them to estimate eye gaze with the cross-ratio property of projective space.

The user's eye is modeled as a sphere with the population average diameter of the eye and with the cornea being a segment of this sphere with a defined radius. The range of head movement is assumed to not exceed 30° to either side of the body's center line and to be located 400 to 600mm from the monitor.

The implemented system consisted of five infrared (IR) light sources and two cameras attached to a computer system and monitor. Users would focus on points located on the monitor while the system would estimate these gaze locations. One of the two cameras was used to track the user's head, while the other camera was used to focus on the user's eye. One IR light source was mounted to each of the four corners of the monitor to generate four Purkinje reflections in the eye image. The fifth light source was mounted to the camera that focused on the eye to generate bright pupil images, similar to the red-eye effect in flash photography. The four Purkinje images would represent the reflection of the monitor while the center of the pupil could be extracted from the bright pupil image. The center of the pupil is assumed to be the projection of the eye gaze onto the monitor plane. Another projection takes place on the image plane when the camera captures the reflections in the user's eye. Using the reflections in the eye and knowing that the composition of the two projections previously described is another projection, we can compute the cross-ratio. This cross-ratio will effectively do a reverse projection of the pupil on the image plane into the eye gaze location on the monitor plane.

Figure 2-3 illustrates the concept of the cross ratios property in more detail. The groups of IR LEDs attached to the corners of the monitor are: **L1**, **L2**, **L3**, and **L4**. They produce light which creates points v_1 , v_2 , v_3 , and v_4 , respectively, through a virtual tangent plane before reaching the pupil. These points are translated as u_{v1} , u_{v2} , u_{v3} , and u_{v4} onto the camera image plane when images of the pupil are captured.

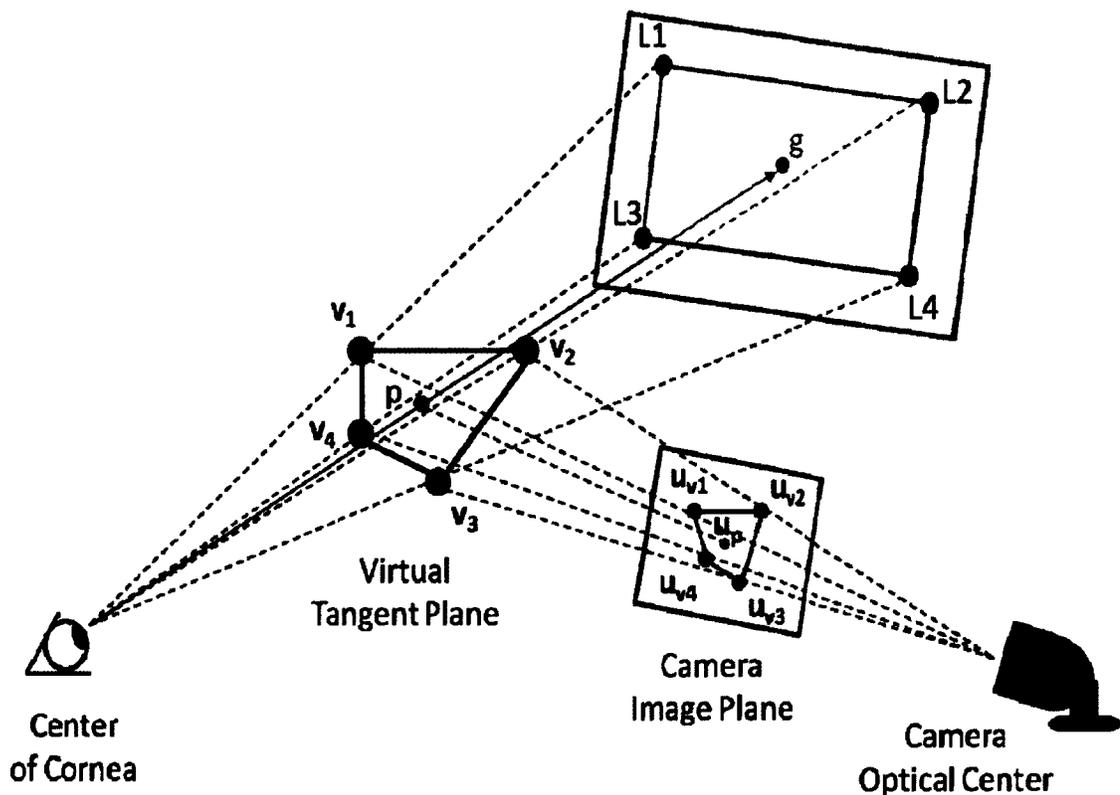


Figure 2-3: The relationship of planes involved in projective space (based on [6])

The center of the pupil, p , is projected onto the camera image plane as u_p , and onto the monitor plane as g , the gaze point. Since the pupil projection points are approximately coplanar, one can estimate g with the projective invariant, the cross ratio.

Yoo and Chung [6] presented a geometric derivation that simplifies the expressions for the cross ratio of the monitor plane and the cross ratio of the image

plane. Referring to Figure 2-4(a), the point **e** is the cross-point of the line $\overline{u_{v1}u_{v3}}$ and line $\overline{u_{v2}u_{v4}}$. The point \mathbf{u}_{m1} is the cross point of the line $\overline{u_{v1}u_{v2}}$ and the line that connects point **e** with the vanishing point (cross point of $\overline{u_{v1}u_{v4}}$ and $\overline{u_{v2}u_{v3}}$). The point \mathbf{u}_{m2} is the cross point of the line $\overline{u_{v1}u_{v2}}$ and the line that connects point \mathbf{u}_p with the vanishing point. Therefore the equation for the cross-ratio of the four points on line $\overline{u_{v1}u_{v2}}$ is:

$$CR_{image}^x = \frac{(x_{v1}y_{m1} - x_{m1}y_{v1})(x_{m2}y_{v2} - x_{v2}y_{m2})}{(x_{v1}y_{m2} - x_{m1}y_{v2})(x_{m1}y_{v2} - x_{v1}y_{m2})} \quad (1)$$

Where: $\mathbf{u}_{vi} = (x_{vi}, y_{vi})$ and $\mathbf{u}_{mi} = (x_{mi}, y_{mi})$ and $(i=1, 2, 3, 4)$ are the coordinates of each point in the image.

Similarly, the cross-ratio of the monitor plane in Figure 2-4(b) can be derived as follows:

$$CR_{screen}^x = \frac{\left(w - \frac{w}{2}\right)\hat{x}_g}{\left(w - \hat{x}_g\right)\frac{w}{2}} = \frac{\hat{x}_g}{w - \hat{x}_g} \quad (2)$$

Where: w is the width of the monitor and \hat{x}_g is the x-coordinate of the estimated gaze point **g**.

The cross-ratios of the image plane and monitor plane are equal because the cross-ratio is invariant in projective space [6]. This equality allows us to simplify the x-coordinate of the estimated gaze point as:

$$\hat{x}_g = \frac{wCR_{image}^x}{1 + CR_{image}^x} \quad (3)$$

We can derive the cross-ratio equations for the y-coordinate in Equations 4 to 6 with the same process as is shown in Equations 1 to 3.

$$CR_{image}^y = \frac{(x_{v2}y_{m3} - x_{m3}y_{v2})(x_{m4}y_{v3} - x_{v3}y_{m4})}{(x_{v2}y_{m4} - x_{m4}y_{v2})(x_{m3}y_{v3} - x_{v3}y_{m3})} \quad (4)$$

$$CR_{screen}^y = \frac{\left(h - \frac{h}{2}\right)\hat{y}_g}{\left(h - \hat{y}_g\right)\frac{h}{2}} = \frac{\hat{y}_g}{h - \hat{y}_g} \quad (5)$$

$$\hat{y}_g = \frac{hCR_{image}^y}{1 + CR_{image}^y} \quad (6)$$

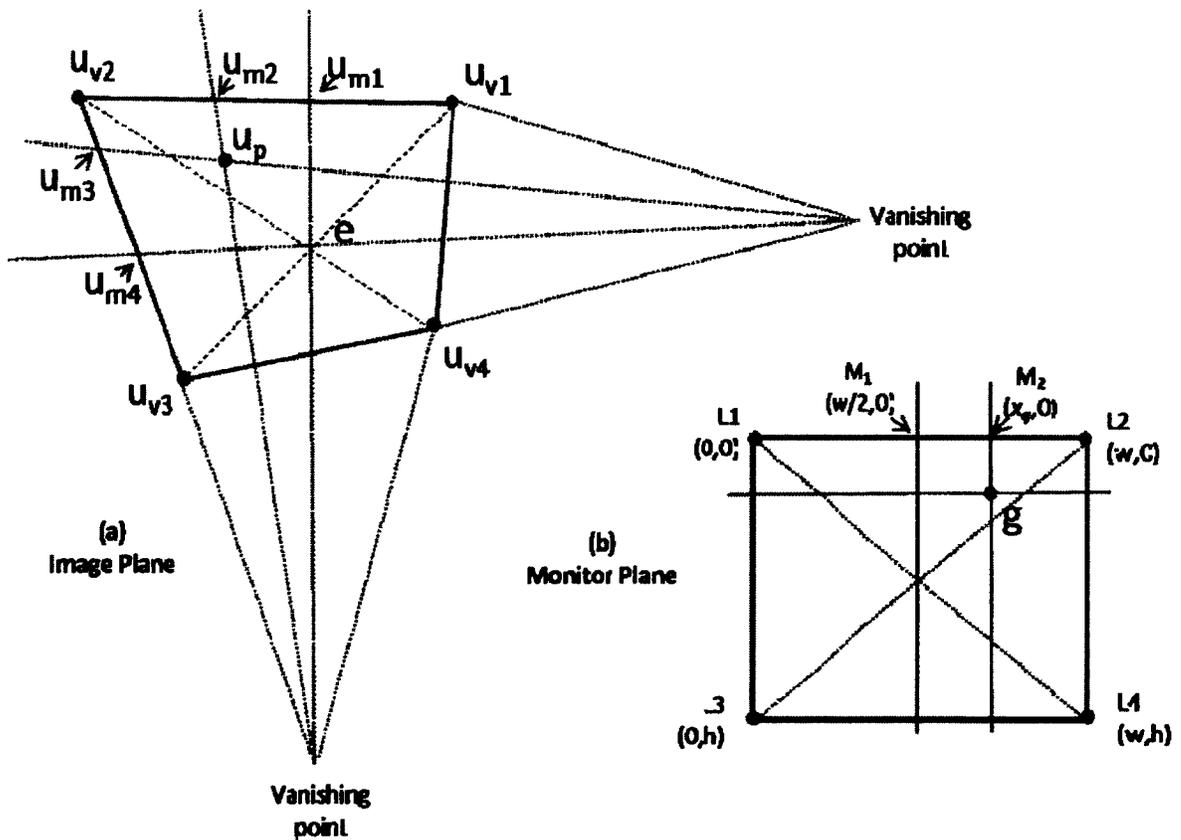


Figure 2-4: (a) Points used for the cross ratio of the image plane, (b) Points used for the cross ratio of the monitor plane. (based on [6])

2.4 Calibration of Eye Gaze Tracking Systems

One of the limitations of vision based eye gaze tracking systems is that most require the user to perform a personal calibration. This is to compensate for any nonlinear displacements of the projected eye features in the image due to the different eye shapes of each individual. Reducing the calibration process, while maintaining a high level of accuracy, is one of the most sought after objectives in developing eye gaze tracking systems.

2.4.1 Four Point Calibration

For the eye gaze system using the cross-ratios method, Yoo and Chung [6] had to first estimate the image coordinates of the virtual projection points in order to utilize the cross ratio property since they cannot be directly observed. After simplification, the virtual projection point can be estimated with the following equation:

$$u_{vx} = u_c + \alpha(u_{rx} - u_c) \quad (7)$$

Where: u_{vx} is the virtual projection point, u_c is the coordinate of the pupil center, u_{rx} is the coordinate of the glint from the corner of the monitor, and α is a constant coefficient approximately equal to 2.

To increase the accuracy of the system, a calibration process is used to compute the exact coefficient of α . This is done by having the user look at each of the light sources (L1 through L4) one at a time. In theory, the virtual projection point of the glint should now be lined up with the center of the pupil. The coefficient can then be computed with the following equation:

$$\alpha = \frac{d(u_p, u_c)}{d(u_{rx}, u_c)} \quad (8)$$

Where: $d(u_p, u_c)$ is the Euclidean distance between the image coordinates of the pupil center and the glint made by the LED of the camera, and $d(u_{rx}, u_c)$ is the Euclidean

distance between the image coordinates of the glint made by the LED on the monitor and the glint made by the LED of the camera. This process is repeated four times to solve the coefficient for the four virtual projection points.

2.4.2 Single Point Calibration

It has been found that some eye gaze systems such as [33] and [34] have been able to reduce the number of calibration points to one. Guesterin and Eizenmann [33] implemented a two camera system where the cameras were operated as a stereo configuration. By having the user gaze at a known location, the horizontal and vertical angles from the optic axis and the visual axis can be solved and applied to eye gaze estimates for the rest of the tracking session. Villanueva and Cabeza [34] simplified the setup by determining that the 1-point calibration could be used with a single camera setup. Their method was based on a geometric analysis that calculated the visual and optical offset, corneal radius, and distance between pupil and cornea centers to improve eye gaze estimates. In both methods, it was found that the single calibration point was feasible, but that an increase in points would make the system more robust.

2.4.3 Calibration Free

Model and Eizenman [35] proposed an "automatic" calibration system where the user did not have to fixate their gaze on any specific calibration point for a determined amount of time. Instead, this method employed a stereo camera setup and relied on the assumption that the visual axes of both eyes would intersect on the monitor. This would allow the system to calculate subject specific angles between optical and visual axes. This assumption holds true for most people; however, it can fail for certain individuals, such as those with strabismus. Although a novel concept,

the system unfortunately was prone to large, subject dependent, gaze estimation errors.

Other calibration free eye gaze tracking methods include [36], [37], and [38]. Shih *et al.* [36] proposed a stereo camera system that estimated the gaze direction based on the orientation of the eye's 3D line of sight. The method determined the 3D position of the eye based on calibrated transformation matrices and determined the eye gaze by solving linear equations. Nouredin *et al.* [37] developed an eye gaze tracking system that used a wide angle and narrow angle camera. The device was set up so that the wide angle camera would focus on the user, while the narrow angle camera focused on a mirror that rotated to reflect light into the camera. The system would approximate the distance of the user from the camera to position the narrow angle camera and mirror as well as determine a scaling factor to be used in the eye gaze estimation. Park [38] presented a method that used three cameras where one of the cameras had a wide angle view to detect eye center positions. These positions were used to control the pan and tilt function of the other two cameras with narrow angle views to accurately capture magnified eye images. These images were then used to determine the 3D position of the eye to calculate the eye gaze position. In these last three methods, the user calibration procedure was eliminated because a 3D eye gaze vector could be determined that was made possible by using two or more cameras.

2.5 Biomedical Applications of Eye Gaze Tracking Systems

The most common function of an eye gaze tracking system is as a human-computer interface device. This interface allows for the design of applications tailored to any particular need. There are a number of applications in medicine that have utilized this technology. Most applications are for patient assistive care where the eye gaze tracking systems do not have to be as accurate; however, with higher

accurate eye gaze tracking systems, they can be utilized in clinical diagnosis and research work.

Traditionally, in medicine, the most common application of eye gaze tracking systems is as an assistive device for people with high level physical disabilities or muscle control. The system allows an individual to interact with a computer without the use of traditional input devices such as a keyboard and mouse or voice commands. Commercial systems have been developed to help individuals who suffer from disabilities such as amyotrophic lateral sclerosis (ALS), cerebral palsy, and aphasia/stroke. These patients may have an impaired ability to speak and move so these systems can provide them with a tool that can output speech sounds to communicate with others or execute control commands for other assistive devices they may be using. The system can help patients control their environment, improving their quality of life. Commercial systems have also been designed to help those with neurological disorders such as developmental delay, autism, and Rett Syndrome. These systems can assist individuals with social integration, helping them develop language and other communication skills [28].

Highly accurate eye gaze tracking systems can be used in different clinical diagnostic tools. One application is in the diagnosis and measurement of strabismus in patients. Strabismus is a condition in which a patient has an ocular misalignment more commonly referred to as "lazy eye" or "crossed eye". An eye gaze system would measure the gaze in each eye in conjunction with Hess-Lancaster tests to measure the ocular deviation. This measurement would provide quantifiable results to help determine proper corrective action [2], [3]. Another application for very accurate eye gaze systems is the positioning control of an endoscopic camera during surgery. Previously, the controls of a camera position platform were performed with voice commands from a surgeon. Unfortunately, the limitation was the system had poor voice recognition and fixed step size for camera movement, sometimes causing

the camera to overshoot a desired target. With an accurate eye gaze tracking system, surgeons can guide the camera to an exact region of interest based on their gaze point [5].

Other applications for highly accurate eye gaze trackers have come in the form of clinical research work. For example, in one study [4], researchers used an eye tracking system to collect data to optimize radiology workstations. Radiologists interpret patient images to diagnose potential problems. The task is often repetitive but must be done accurately, to prevent errors, and efficiently, to prevent a backlog of workload that delays patient care. Researchers analyzed gaze information to explain how a radiologist's visual attention was allocated and any eye gaze patterns that may have occurred during visual searches of patient images. The information was then used to propose an optimized ergonomic workstation setup for hospitals.

3 System Overview

3.1 Introduction

This chapter presents the proposed implementation of an eye gaze estimating system based on a property of projective space cross-ratios [6]. The system uses an IR camera to detect reflections of IR light in the eye. These reflections provide the necessary information to estimate the user's eye gaze. This chapter will describe the hardware and software components used in the system. Next, the process of acquiring images for the system is explained. Finally, the image processing techniques needed to determine the pupil center, the Purkinje reflections, and cross ratios are presented. The system implemented in this thesis work was not set up to run in real-time but rather to collect information and process it offline.

3.2 System Hardware Setup

The system presented herein (Figure 3-1) consists of a personal computer (PC) with a liquid crystal display (LCD) monitor, a Canon (Japan) VC-C50i pan-tilt-zoom (PTZ) IR charge-coupled device (CCD) camera with a resolution of 640x480 pixels and five IR light sources.

The computer system is used to control the LEDs, capture images, and process images offline. The computer system is an Intel Core 2 Duo 3.0Ghz system running Windows XP SP3. Matlab R2010a is used for offline processing of the images and data. A data acquisition (DAC) device (model USB-6008, National Instruments, Austin TX, USA) is connected to the computer via the universal serial bus (USB) to control the IR LEDs. A single-channel color image acquisition board (model PXI-1411, National Instruments, Austin TX, USA) is connected to the computer via the peripheral component interconnect (PCI) bus to capture images from the camera.

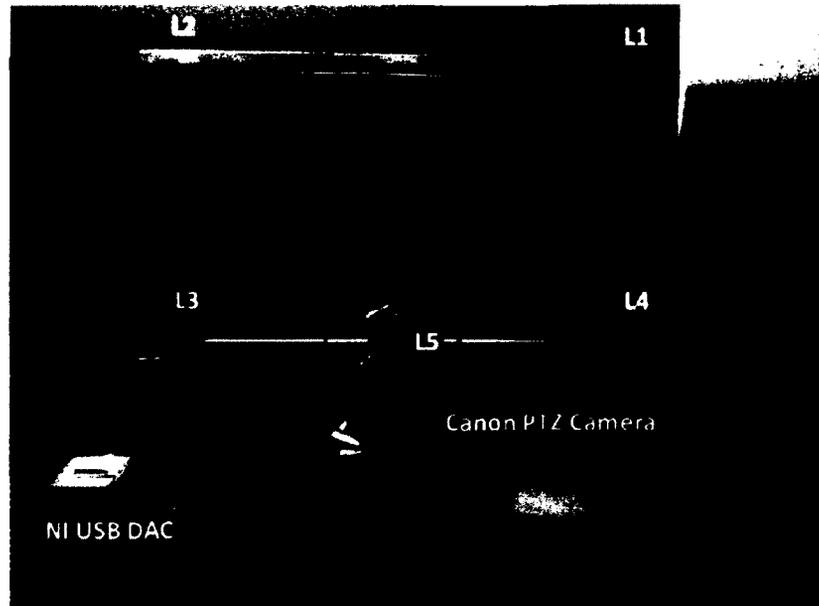


Figure 3-1: Eye gaze system comprising of a Canon PTZ camera, Nation Instruments USB DAC, 5 LED light sources (L1 to L5), an LCD monitor and PC (not shown).

The LCD monitor is a 20" widescreen monitor (model UltraSharp2009W, Dell, Round Rock TX, USA) with a max resolution of 1680x1050 pixels, which is centered and placed 65 centimetres in front of the user's head. Dividing the resolution of the monitor with the dimensions of the active area of the monitor (435mm x 270mm), each pixel in the horizontal direction is 0.2589mm while each pixel in the vertical direction is 0.2571mm. The camera is positioned and centered directly in front and underneath the monitor.

There are five IR light sources (L1 to L5, Figure 3-1). Four of the IR light sources (L1 to L4) are made from five Fairchild Semiconductor QED223 Plastic IR LEDs connected together. These four light sources are attached to each of the four corners of the LCD monitor to produce glints (the four Purkinje reflections). The fifth IR light source (L5) is made from six Fairchild Semiconductor QEE123 Plastic IR LEDs connected together. The fifth light source is mounted directly around the lens of the camera to produce the bright-eye effect. The costs of the main individual components are listed below in Table 3-1. The implemented system is

approximately half the present cost of the least expensive commercial system presented in Table 2-1.

Table 3-1: System component costs

Part	Approximate Cost (\$CAD)
National Instruments PXI-1411 Frame Grabber	\$1,200
Canon VC-C50i PTZ Camera	\$1,000
National Instruments USB-6008 DAQ	\$190
Fairchild Semiconductor QED223 LEDs	\$10
Fairchild Semiconductor QEE123 LEDs	\$5
Total	\$2,405

It is important to note that the system can be made much cheaper using any one of the USB or wireless cameras that are currently available. An equivalent camera from Lightinthebox.com (model EasyN, China) can be purchased for approximately \$70 USD. This could replace the current Canon PTZ camera and eliminate the need for the National Instruments frame grabber, significantly reducing the cost of the system to approximately a fifth of the cost of the least expensive commercial system. To further reduce the cost of the system, replacing the USB DAQ with a custom hardware circuit board containing a microcontroller could reduce the overall cost of the system to \$100 to \$150 CAD, significantly cheaper than a commercial system.

3.3 Image Acquisition

This research was reviewed and approved by the Carleton University Research Ethics Committee. The letter of information was presented to each participant and informed consent was received prior to data collection. Both forms can be seen in Appendix A. Data collection was performed on five participants that comprised of four males and one female. Three males had no known vision disorders, while one male and the female participant wore corrective contact lenses for myopia. A summary of the participant details can be seen in Table 3-2.

Table 3-2: Participant information

Subject	Sex	Age	Dominant Eye	Vision Disorder
P1	M	29	Right	None
P2	F	28	Right	Myopic
P3	M	28	Right	None
P4	M	30	Right	None
P5	M	29	Right	Myopic

Each of the participant's heads were positioned into a head fixation device (Figure 3-2a). The device was comprised of a Victory 3001 camera tripod with a custom head mount attached. The head mount was made up of two vertical padded metal bars that the user placed their head between, a chin rest, and a forehead strap. With the user's head in the mount, the amount of head movement was minimized while the system captured images. The camera's pan, tilt, and zoom functions were then manually controlled to obtain images of the user's dominant eye. The user's dominant eye was determined prior to data collection using the procedures outlined by the Porta Test (see Appendix B). The dominant eye is the eye that will gaze directly at the target, while the non-dominant eye's gaze is at a slight angle, providing the user depth perception. Since we want the user to gaze exactly on targets, the camera was adjusted until an image of the dominant eye would fill a 280 x 150 pixel bounding box that was located in the center of the 640x480 pixel image plane (Figure 3-2b).



Figure 3-2: Setting up the camera so user's dominant eye fills bounding box

The camera's frame rate was set to 12 hertz. This allowed the system to collect four sets of three images. Controlling the amount of IR light reflecting onto the user's eye produced a different type of eye image. The three types of images captured (Figure 3-3) were a dark eye image, a bright eye image, and a glint image. The dark eye images (Figure 3-3a) were images of the eye when all IR light sources were turned off. The bright eye images (Figure 3-3b) were images of the eye when only the IR light source that was mounted on the camera was turned on (L5). This produced an image in which the pupil was reflected, much like the red-eye effect produced in flash photography. The glint images (Figure 3-3c) were images of the eye that were produced when only the IR light sources located on each corner of the monitor were turned on (L1 to L4). This produced an image containing the four Purkinje reflections that were seen as bright points on the user's eye. All images were saved in joint photographic experts group (JPEG) format to be processed offline.

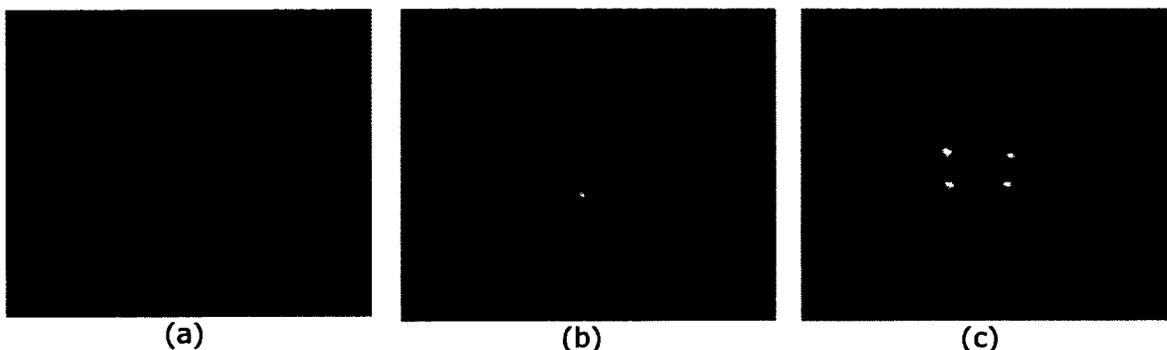


Figure 3-3: Capture images (a) dark pupil, (b) bright pupil, and (c) glint pupil

A program written in Matlab was used to control when the 15 targets appeared on the monitor screen. The default display on the monitor was a flat black background. When each user was ready, targets, made up of white circles (26 pixels or 10mm) with black crosses and smaller black circles, appeared one at a time in locations across the screen. The order and location of the targets is presented in

Figure 3-4. The user was directed to fixate their gaze at each target point while the system captured images for 50 eye gaze estimates (approximately 35 s). Details about the collection of images for experiments involving image resolution and head position are further described in Chapter 4.2 and Chapter 5.2, respectively.

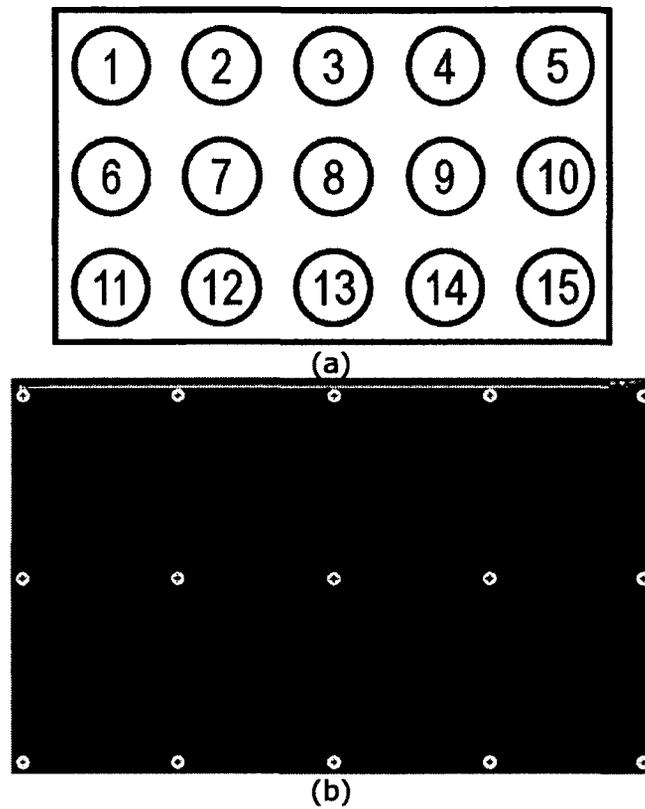


Figure 3-4: Target locations presented to user during data collection (a) target order and (b) screen shot of targets as they appeared to the participants.

3.4 Image Processing to Estimate Eye Gaze

This section will present the data processing that was done offline with Matlab to estimate the user's eye gaze. There are three main steps involved in estimating the user's eye gaze which include estimating the pupil center location, estimating glint locations, and finally, calculating the cross ratios. Each step is further discussed in the following subsections.

3.4.1 Estimating Pupil Center Location

The eye gaze system estimated the pupil center in a two-stage process. The first stage in the process was to estimate approximately where the pupil was in the image. The second stage involved forming a boundary box to isolate the pupil and extract the pupil center.

To estimate the approximate location of the pupil in the image, the location of the bright glint on the user's pupil in the bright eye image was extracted. The first step involved eliminating the background of the bright eye image by finding the difference between it and the dark eye image (Figure 3-5a). The image was then converted into a binary black and white image by applying a threshold (Figure 3-5b). Finally, a morphological gradient operator was applied to the binary image to enhance segmentation of blobs (Figure 3-5c). Unfortunately, this process produced "holes" in the image. Holes are defined as "a set of background pixels that cannot be reached by filling in the background from the edge of the image" [39]. Using the Matlab image processing command `imfill`, these holes were eliminated. The centroid of the largest blob left in the image was then taken as the approximate location of the pupil.

The next stage in estimating the pupil center involved forming a bounding box around the approximate location found in the previous step. After a bounding box was formed, the true pupil center can be determined using one of the following two methods: a Hough transform or an Area of Interest (AOI) operator.

The first method uses a standard Hough transform to detect circles [40]. Detecting the circles involved computing the gradient of the image and the magnitude of the gradient. The Hough transform was then applied with a range of radiuses as an input, returning the centroids and radiuses of circles it had detected. The centroids were then arranged in descending order based on size and then

descending order based on proximity to the approximate pupil location determined earlier. Starting with the first centroid in the list, each was checked to see if it was within the area of the bounding box. If so, the centroid was designated the best match for the pupil with the centroid representing the pupil center (Figure 3-6).

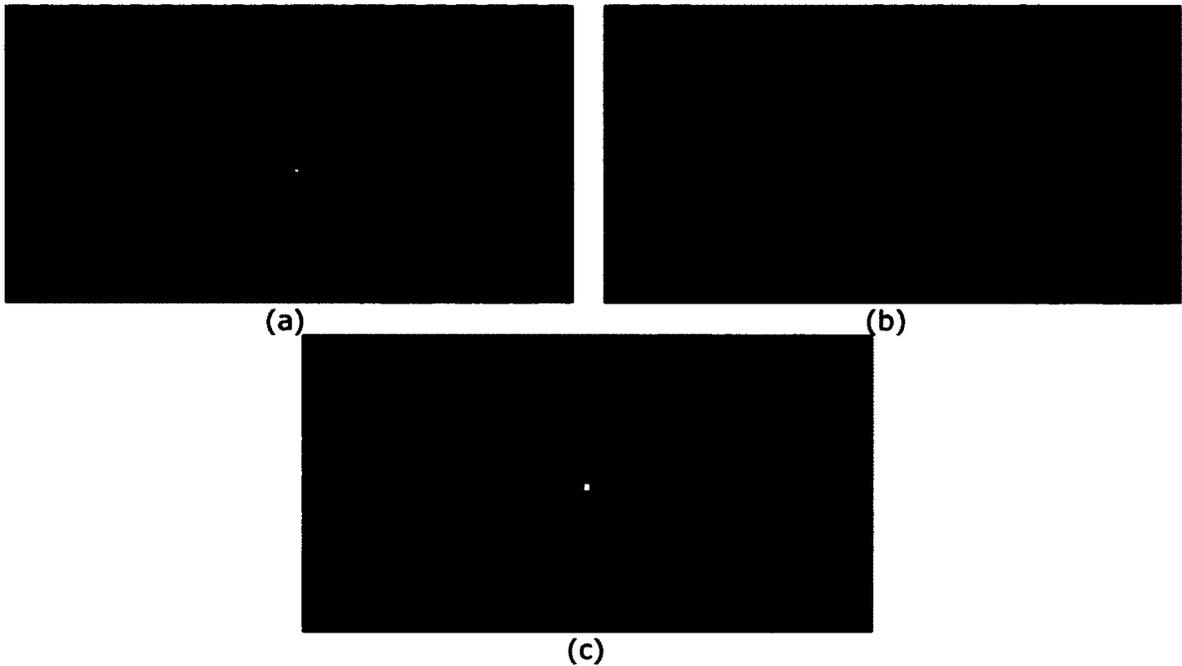


Figure 3-5: Image processing during estimation of approximate pupil center – (a) background removal, (b) binary thresholding, and (c) morphological gradient.



Figure 3-6: Hough transform detecting pupil. Blue circles represent all the circles found by the Hough transform in the image. The red rectangle is the bounding box used in searching for the pupil center. The green cross is the pupil center found

The second method used an Area of Interest operator to detect the pupil region. The first step was to eliminate the background in the bright eye image by finding its difference from the glint image (Figure 3-7a). This background elimination was more effective in removing the face from the image compared to the first stage of the pupil estimation described earlier. A threshold was then applied to create a binary image (Figure 3-7b). Next, a morphological gradient was applied with any holes filled, similar to the steps taken previously (Figure 3-7c). The centroids and sizes of areas of interest in the image were then determined. After, the AOIs were sorted by size in descending order. Starting with the largest AOI, their respective centroids were checked to see if they fell within the area of the bounding box. If so, the AOI was deemed to represent the pupil and its centroid was defined as the pupil center (Figure 3-7d).

Eye gaze estimates using the two pupil detection methods for one participant yielded the following plots that can be seen in Figure 3-8. From the plots, the radial mean distances of the clusters from the target points were similar; however, the cluster variances were greater for the estimates using the Hough transform than they were for the estimates using the AOI operator. Based on this empirical data, the use of the AOI operator in the eye gaze estimates was chosen.

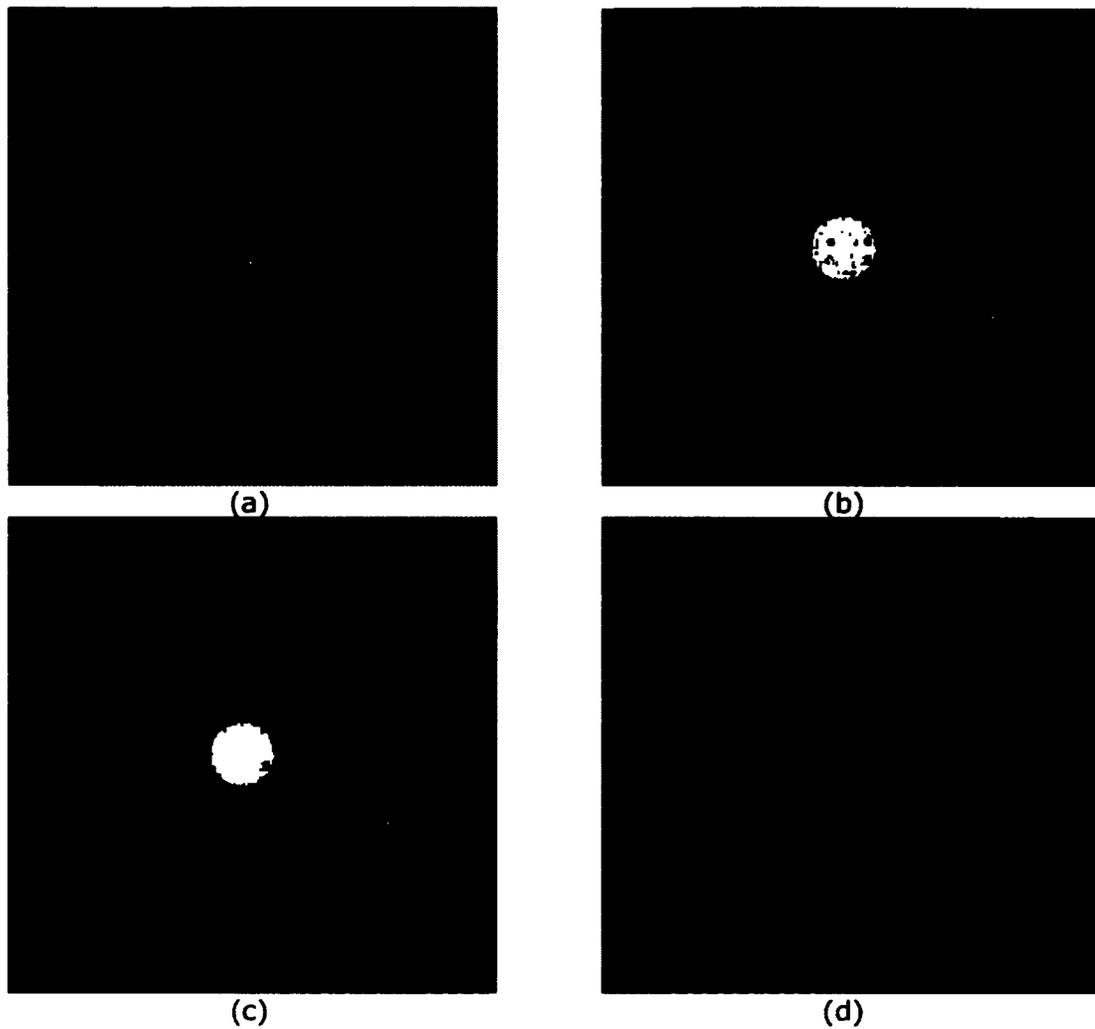
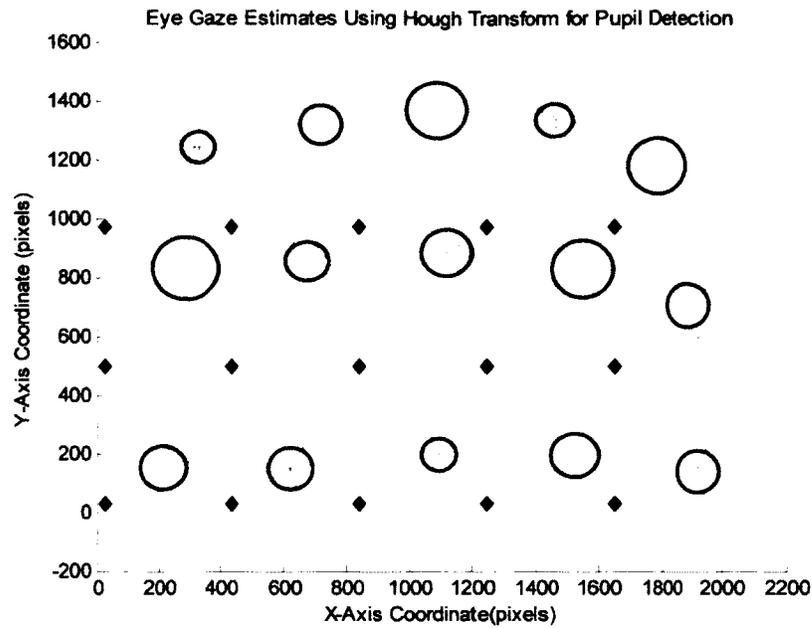
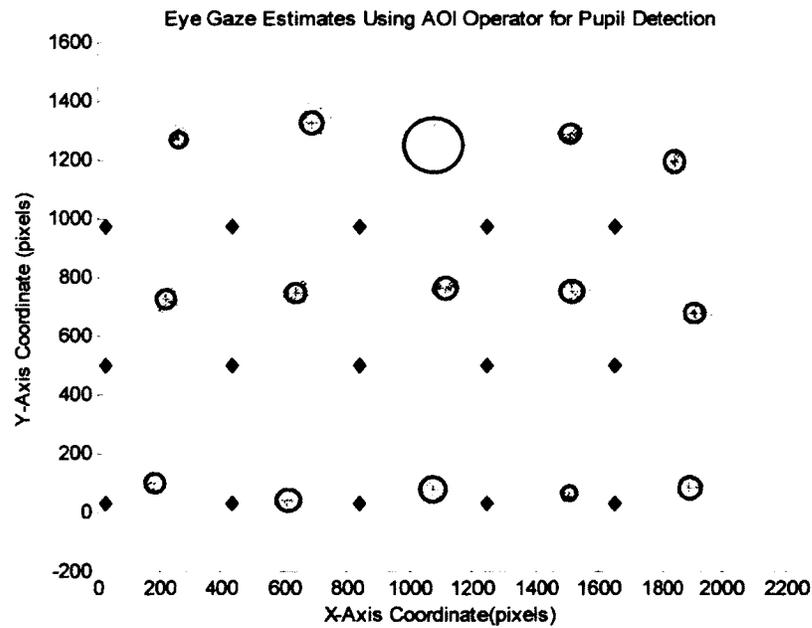


Figure 3-7: Image processing steps executed to determine the true pupil center – (a) background removal, (b) binary thresholding, (c) morphological gradient, and (d) bright eye image showing approximate pupil center (green cross) and determined pupil center (red cross)



(a)



(b)

Figure 3-8: Eye gaze estimates using two different pupil detection methods. The red x's represent eye gaze estimates while the black diamonds represent the target points. The blue crosses show the mean eye gaze estimate of the cluster, while the blue circles represent the cluster variance. (a) Estimates using the Hough Transform, (b) Estimates using the Area of Interest Operator.

3.4.2 Estimating Glint Locations

Estimating the glint locations was similar to the steps taken to estimate the pupil center. First, the background of the glint image was removed by finding the difference between it and the dark eye image (Figure 3-9a). A threshold was then applied to create a binary image (Figure 3-9b). After, a morphological gradient was applied (Figure 3-9c) and any holes were filled in (Figure 3-9d). Areas and centroids of blobs were then computed and sorted by size in descending order. Blobs larger or smaller than defined sizes were removed under the assumption that they were not glints but rather from face reflections or noise, respectively. Using the pupil center, a bounding box was defined (Figure 3-9e). The four largest remaining blobs within the bounding box were taken to be the four glints (Purkinje reflections). The four glints were then classified as u_{v1} , u_{v2} , u_{v3} , and u_{v4} (Figure 3-9f) representing light sources L1 through L4 respectively.

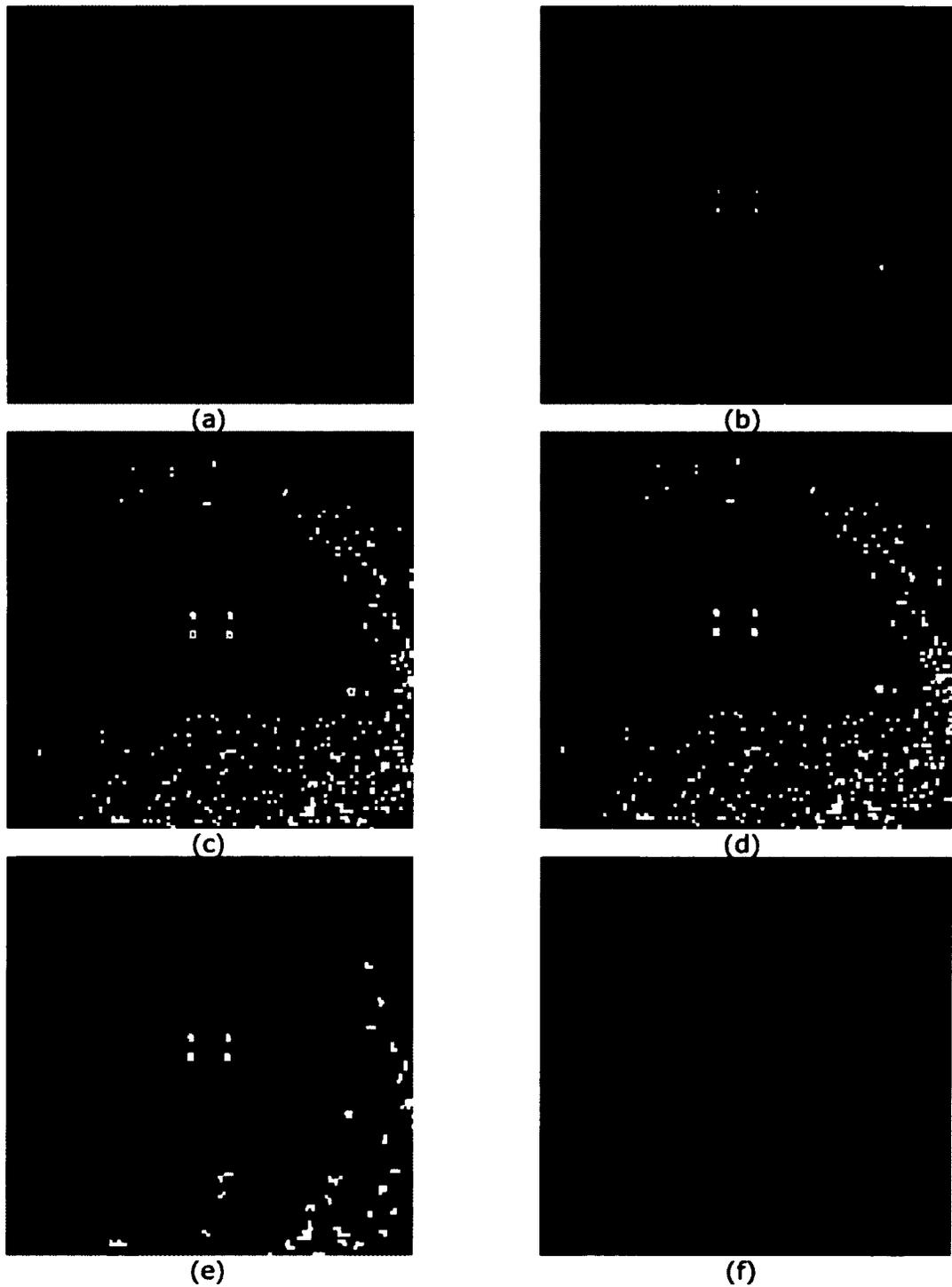


Figure 3-9: Image processing steps taken during estimation of glint locations – (a) background removal, (b) binary thresholding, (c) morphological gradient, (d) filling of “holes” in image, (e) bounding box, (f) organization of glint locations.

Having all four Purkinje reflections detected at all times was difficult, as they would sometimes vary in strength. When only three reflections were detected, the system determined a "ghost point" by estimating the fourth reflection so that it could be used with the cross ratios formula. The fourth point was estimated as follows:

1. Assigning the three known points as u_{v1} , u_{v2} , and u_{v3}
2. Determining the x-axis distance of $u_{v1}u_{v2}$, $u_{v1}u_{v3}$, $u_{v2}u_{v3}$
3. Determining the line segment with the minimum x-axis distance.
4. Between the 2 points of this line segment, assigning P1 as the point with the maximum y-axis coordinate and P2 as the point with the minimum y-axis coordinate. Assigning P3 as the remaining point
5. Assign x-axis coordinate of P4 as x-axis coordinate of P3
6. Determine the y-axis distance of P1P3 and P2P3. If $P1P3 > P2P3$, then assign the y-coordinate of P4 as the y-coordinate of P1; else, assign the y-coordinate of P4 as the y-coordinate of P2.
7. Finally, P1, P2, P3, and P4 are rearranged to form u_{v1} , u_{v2} , u_{v3} , and u_{v4} .

If the process finds less than three glints, the system is unable to estimate the user's gaze point.

3.4.3 Estimating Gaze Point with Cross Ratios

The constant coefficient, α , was set to 2 to approximately estimate the virtual projection points u_{v1} , u_{v2} , u_{v3} , and u_{v4} . Using the location of the pupil center as p and the four projection points, the system was able to estimate the user's gaze point using the cross ratios formula as described in Section 2.3.

4 Effects of Image Resolution

4.1 Introduction

This chapter presents the effects of eye image resolution on eye gaze estimates when using the implemented system based on cross-ratios. This chapter will describe the method of data collection. Results of the eye gaze estimate accuracy are then presented when using different eye image resolutions. A discussion is presented to explain the results acquired. The chapter will conclude with the findings from experimental results using different eye image resolutions.

4.2 Method

4.2.1 Data Collection

Data were collected from five participants (4 males and 1 female) who were right-eye dominant. The female (P2) and one male (P5) wore corrective contact lenses while the 3 other males (P1, P3, and P4) had no known vision disorders. As described in detail in Chapter 3.3, each participant sat 65 centimetres in front of the monitor screen with their head attached to the fixation device to minimize head movement. Participants' heads were placed so that the top bridge of their nose was aligned with the center of the monitor screen. Fifteen targets appeared on the monitor screen one at a time. The location and order of the targets on the monitor screen can be seen in Figure 3-4a. Each participant was asked to focus their gaze at each of the fifteen target points as they appeared on the monitor screen until they disappeared. During this time, eye images were captured from the camera at a resolution of 640x480 pixels. This resolution was the maximum eye image resolution the camera was able to capture.

4.2.2 Creating Eye Images With Different Resolutions

Images were resized in Matlab offline using bicubic interpolation into seven other 4:3 aspect ratio eye image resolutions (Table 4-1). Eye image resolutions can be seen with respect to one another in Figure 4-1.

Table 4-1: Resize resolutions of eye images

Resolution (pixel)	Magnification	Approximate ROI (pixel)
<i>200x150</i>	$5/16 \times 5/16$	25x25
<i>240x180</i>	$3/8 \times 3/8$	30x30
<i>320x240</i>	$1/2 \times 1/2$	40x40
<i>400x300</i>	$5/8 \times 5/8$	50x50
<i>500x375</i>	$25/32 \times 25/32$	62x62
<i>640x480</i>	0	80x80
<i>800x600</i>	$5/4 \times 5/4$	100x100
<i>1024x768</i>	$8/5 \times 8/5$	128x128

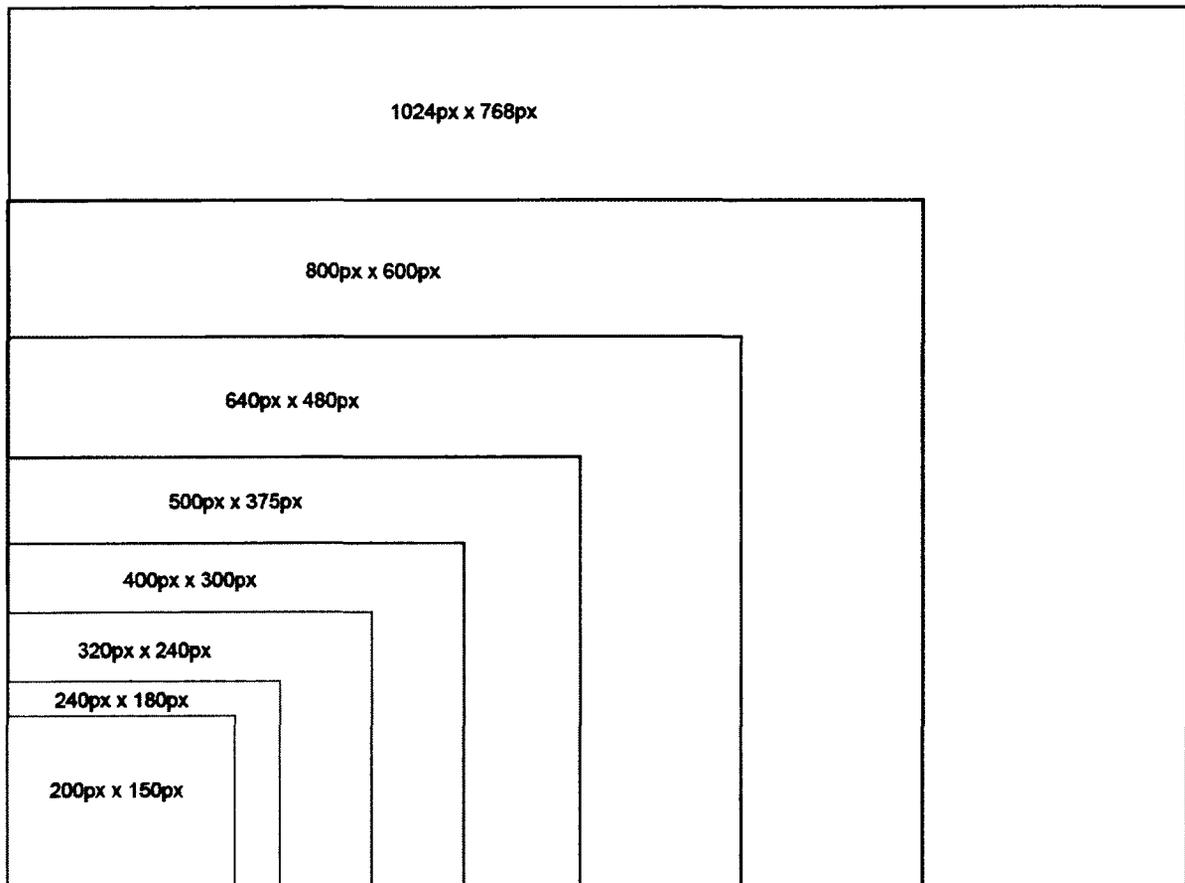


Figure 4-1: Eye image resolutions with respect to one another. The largest eye image resolution used was 1024x768 pixels while the smallest eye image resolution used was 200x150.

The region of interest (ROI), which contains the user's iris and pupil, is approximately 2% of each eye image. Figure 4-2 is an example of an eye image showing the ROI bounding box. The area in the red bounding box is where the system will search for eye features to use for eye gaze estimation.

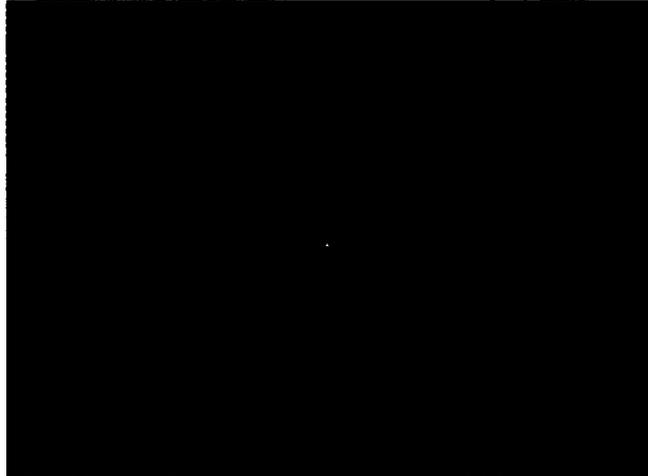


Figure 4-2: Example of an image of participant P1 with a bounding box outlining the ROI. The dimension of the image is 320 pixels wide by 240 pixels tall (area of 76800 pixels). The bounding box is 40 pixels wide by 40 pixels tall (area of 1600 pixels or $1600/76800 \times 100\% = 2.08\%$ of the eye image).

4.2.3 Eye Gaze Estimation

Eye gaze estimates were performed for all 15 target points using the proposed eye image resolutions as described in Section 3.4. The average root mean square error (RMSE) was computed between the location of the target on the computer monitor and the estimated gaze point. Eye images where a pupil was not detected with the method described in Section 3.4.1 were automatically discarded during processing. Manually checking a small sample of the discarded eye images showed that either the camera had captured the participant blinking or that it had captured the participant's head or pupil moving, resulting in a blurred image that could not be used. To avoid distortion of the eye gaze estimate sample, outliers in the data for each target were removed with the fourth-spread outlier test. This test identifies

the boundaries of the data and divides it into quartiles, measures the distance f_s between the upper and lower quartile, then sets the upper and lower outlier boundaries as a function of this distance (median of data $\pm f_s * 1.5$). The advantage of using this test for outlier detection is that there was no requirement for sample size. The test also helps to remove the effects of saccades, which are quick eye movements to other stimuli while trying to maintain a fixed gaze. As the eye image resolution decreased, eye gaze estimates became increasingly difficult for the system to compute consistently (this will be explained in Section 4.4). This resulted in sample sizes ranging from 20 to 50 estimates. With the outlier test, any sample size used would detect extreme outliers that may influence the mean of the sample [42].

4.3 Results

A total of 50 eye image sets were captured for each target for each eye image resolution. Table 4-2 shows the number of disregarded points of each participant for each target using 640x480 pixel eye images. The average number of disregarded 640x480 pixel eye images after removal of blinks and outliers was 14.59 \pm 5.36 eye images. Table 4-3 shows the number of eye images where a ghost point was used to calculate the eye gaze estimate for a target using 640x480 pixel eye images. The average number of eye estimates where a ghost point was used was 9.19 \pm 7.82 for each target. For targets 12, 14, and 15, the number of ghost points used was much larger compared to the number of ghost points used to estimate all other targets.

Figure 4-3 is a graph of the average number of disregarded eye images for each resolution after interpolation. A total of 750 eye images were captured for each participant. There was a decrease in disregarded eye images when there was an increase in the eye image resolution from 200x150 pixels up to 500x375 pixels.

When the eye image resolution increased from 500x375 pixels to 1024x768 pixels, the average number of disregarded eye images did not vary greatly.

Table 4-2: The number of disregarded 640x480 pixel eye images for each target for participants P1 to P5. A total of 50 eye images were collected for each target.

Target	Disregarded Eye Images						Avg	Std
	P1	P2	P3	P4	P5			
1	12	12	43	17	19	20.60	12.90	
2	10	12	11	13	8	10.80	1.92	
3	14	13	6	13	13	11.80	3.27	
4	7	15	11	15	10	11.60	3.44	
5	9	14	12	15	12	12.40	2.30	
6	9	16	7	15	11	11.60	3.85	
7	9	11	12	13	23	13.60	5.46	
8	6	14	6	30	15	14.20	9.81	
9	11	13	11	15	11	12.20	1.79	
10	10	7	9	10	16	10.40	3.36	
11	16	18	9	9	18	14.00	4.64	
12	13	29	11	37	13	20.60	11.70	
13	19	20	12	13	12	15.20	3.96	
14	28	18	14	14	18	18.40	5.73	
15	22	30	24	17	14	21.40	6.23	
Average of All Targets						14.59	5.36	

Table 4-3: The number of eye images where a ghost point was used for the eye gaze estimation using 640x480 pixel eye images for each target for each participant P1 to P5.

Target	Ghost Points						Avg	Std
	P1	P2	P3	P4	P5			
1	1	5	1	0	10	3.4	4.16	
2	6	12	5	0	11	6.8	4.87	
3	5	7	0	0	5	3.4	3.21	
4	0	14	0	0	5	3.8	6.10	
5	41	0	2	3	5	10.2	17.31	
6	6	7	0	0	3	3.2	3.27	
7	17	5	0	1	4	5.4	6.80	
8	6	11	3	0	15	7	6.04	
9	0	9	27	0	7	8.6	11.06	
10	0	8	27	1	1	7.4	11.41	
11	27	13	3	6	6	11	9.67	
12	19	20	39	9	2	17.8	13.99	
13	15	12	11	1	15	10.8	5.76	
14	21	19	32	11	24	21.4	7.64	
15	19	10	24	13	22	17.6	5.94	
Average of All Targets						9.19	7.82	

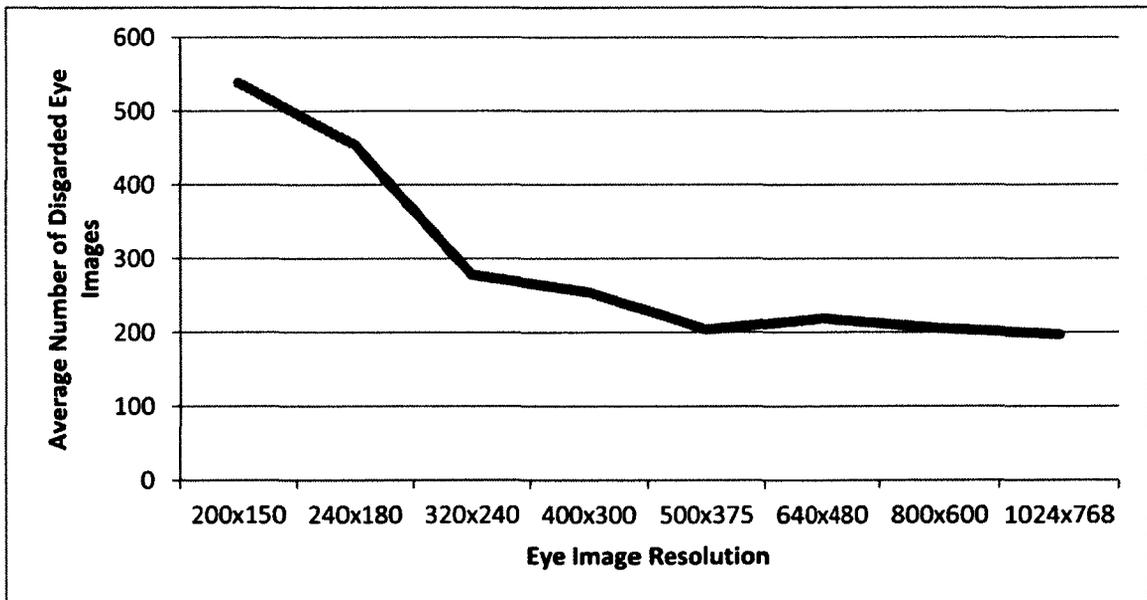


Figure 4-3: Graph of the average number of disregarded eye images of all participants for all 15 targets for each eye image resolution. A total of 750 eye images were used, 50 eye images were collected for each target in 640x480 pixel resolution and interpolated into another 7 different eye image resolutions

Table 4-4 is the number of eye images where a ghost point was used in the eye gaze estimation for each resolution after interpolation.

Table 4-4: The number of ghost points used to calculate eye gaze estimation for each eye image resolution used. A total of 400 eye images were used, 50 eye images were collected for each target and interpolated into 7 other different eye image resolutions.

Resolution	Ghost Points						
	P1	P2	P3	P4	P5	Avg	Std
1	64	128	55	35	101	76.60	37.39
2	107	133	99	134	161	126.80	24.62
3	113	120	124	47	188	118.40	50.01
4	135	159	119	60	127	120.00	36.73
5	135	99	125	48	108	103.00	33.82
6	183	152	174	45	135	137.80	55.17
7	149	126	161	45	187	133.60	54.19
8	144	111	143	46	116	112.00	39.87

Plotting the estimated gaze points for each participant relative to the targets on the monitor screen where they were to focus produced plots similar to Figure 3-8(b). This figure was generated with eye gaze estimates for P1 using eye images with a resolution of 640x480 pixels. The estimated gaze points were shifted to the right and up from the target points. Figure 4-4 is a box plot of the RMSe of the radial distance of the estimate gaze point to the targets. The data used for the box plot contained the average eye gaze estimate for all 15 targets for each participant. Each box represents the data using one particular eye image resolution. Taking the mean of the RMSe in the x-axis and y-axis for all 15 targets for each participant and plotting graphs similar to that previously presented yields Figure 4-5.

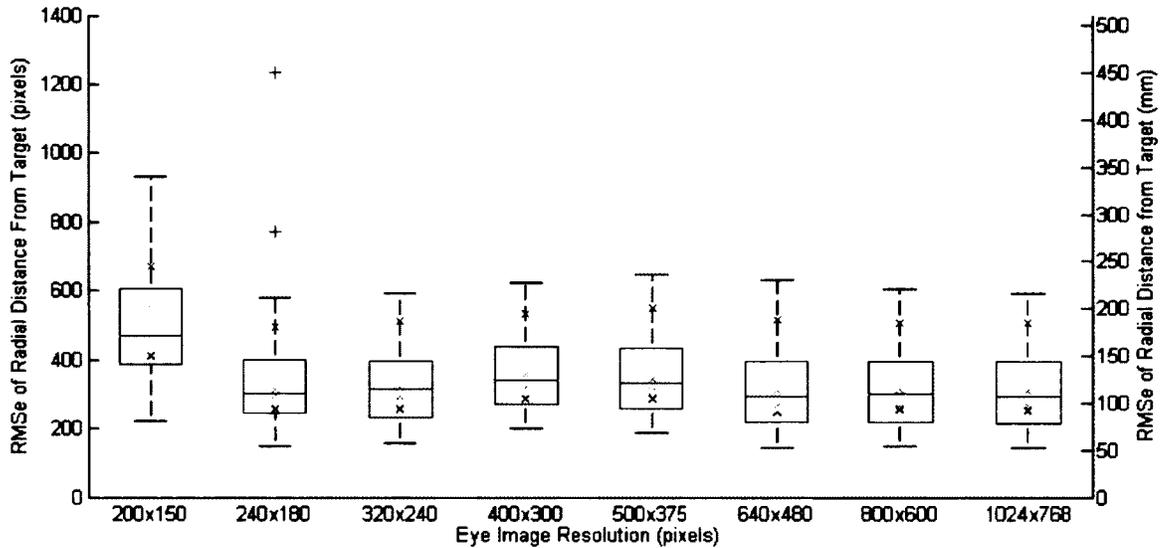
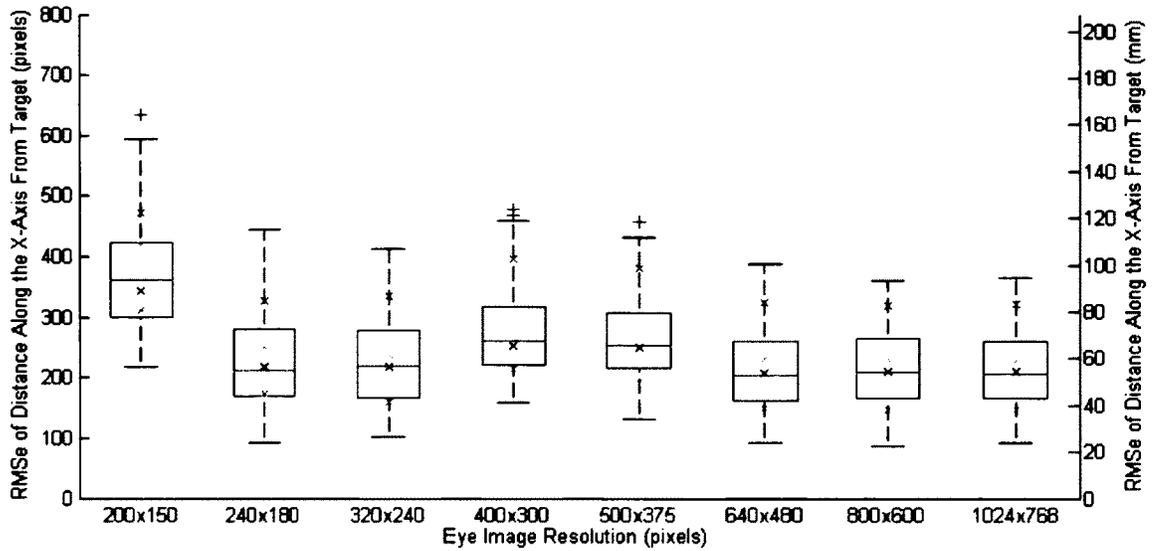
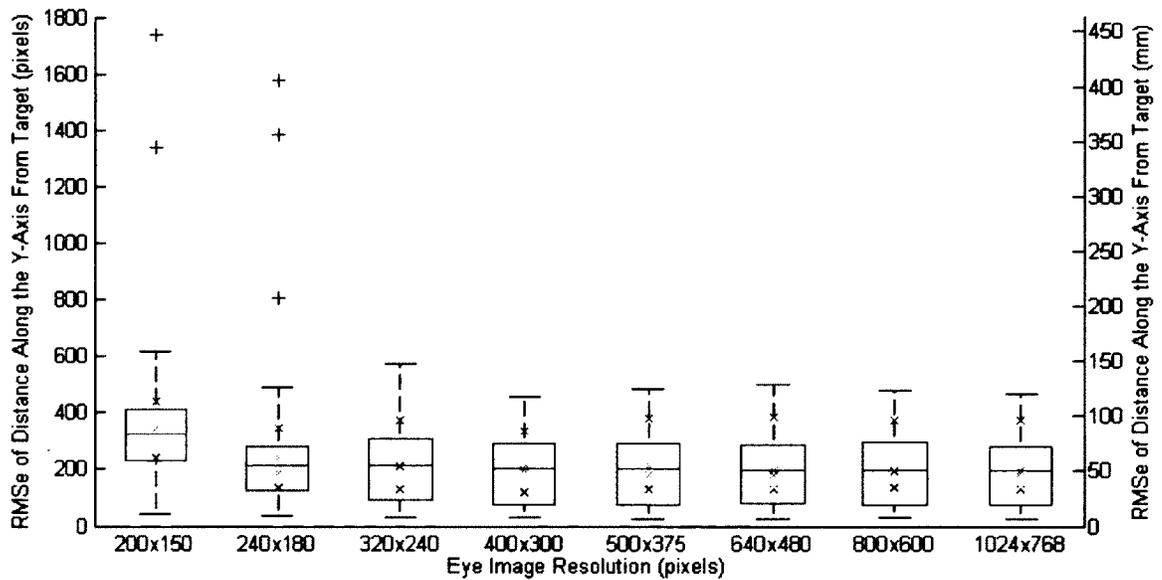


Figure 4-4: Box plots showing the trend of the average RMSe of the radial distance from the targets as eye image resolution increases. Data for each box plot consists of the average radial RMSe for each target of all participants. Each box represents a different eye image resolution used. The red line in the box denotes the median of the data. The top and bottom edges of the blue box represent the 25th and 75th percentiles respectively. The black whiskers represent the maximum and minimum data range, while the red crosses represent outliers. The average RMSe for each participant are represented an 'x' where: green = P1, black = P2, cyan = P3, magenta = P4, and blue = P5.

Plots of the variance of the eye gaze estimates using different eye image resolutions can be seen in Figure 4-6 for participant P2. The estimates are for five of the fifteen targets that participants were asked to focus on during the data collection. The figure consists of plots for eye gaze estimates when participants were focusing on targets 1, 5, 8, 11, and 15. These plots are shown as they represent the general area of the monitor. Targets 1, 5, 11, and 15 appeared in the top left corner, top right corner, bottom left corner and bottom right corner of the monitor screen respectively, while target 8 appeared at the center of the monitor screen.



(a)



(b)

Figure 4-5: Box plots showing the trend of the average RMSe in (a) x-axis and (b) y-axis as eye image resolution increases. Data for each box plot consists of the average RMSe along the x-axis or y-axis for each target of all participants. Each box represents a different eye image resolution used. The red line in the box denotes the median of the data. The top and bottom edges of the blue box represent the 25th and 75th percentiles respectively. The black whiskers represent the maximum and minimum data range, while the red crosses represent outliers. The average RMSe for each participant are represented as 'x' where: green = P1, black = P2, cyan = P3, magenta = P4, and blue = P5.

Figure 4-7 is the average variance of error for eye gaze estimates of all 15 targets for all participants P1 through P5 for each eye image resolution.

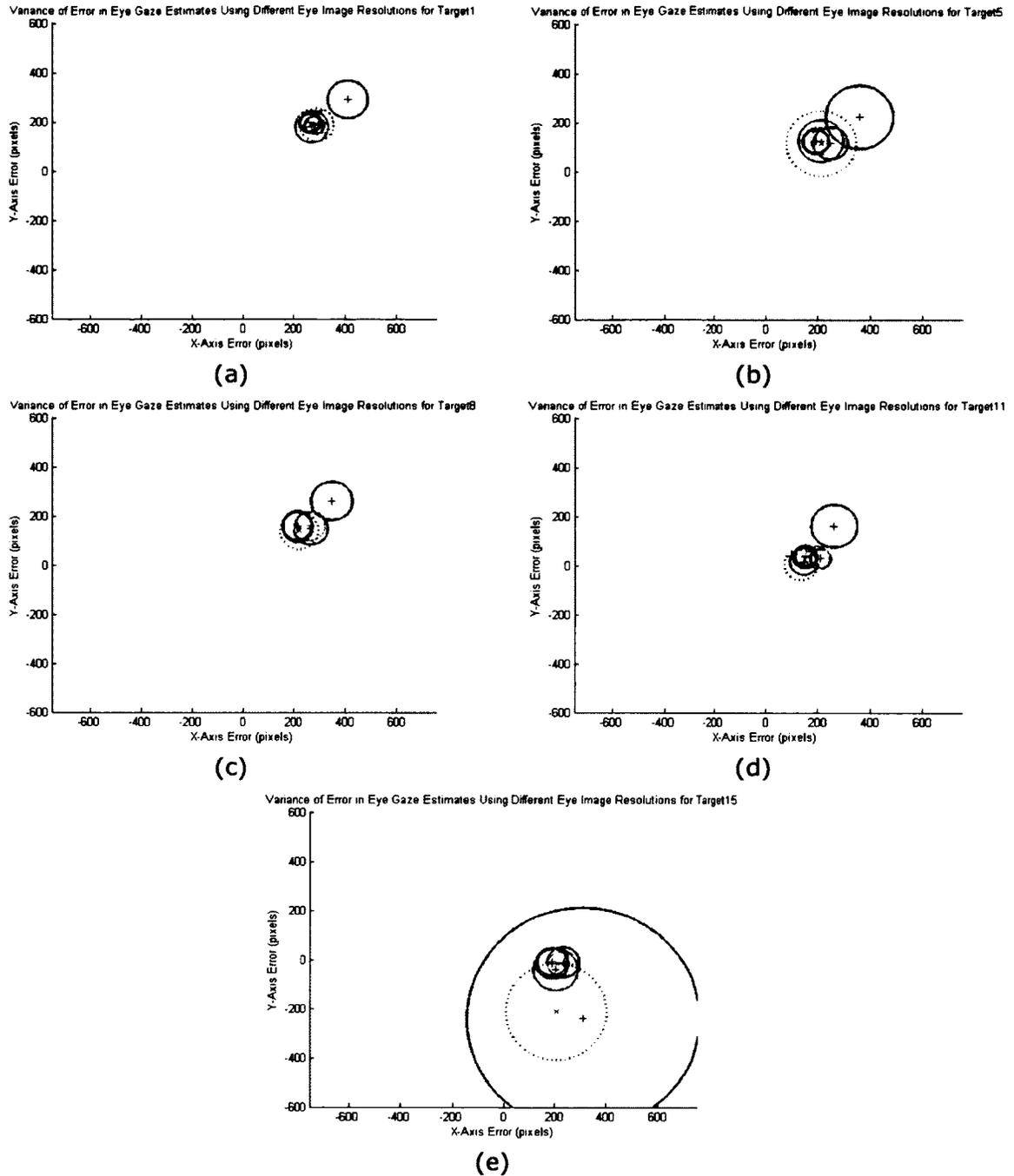


Figure 4-6: The variance of radial error of eye gaze estimates using different eye image resolutions for (a) target 1, (b) target 5, (c) target 8, (d) target 11, and (e) target 15. The radial error is defined by the circle that encompasses the mean eye gaze estimation (represented by a cross or 'x'). The following colors represent the corresponding eye image resolutions in pixels: solid blue = 200x150, dash blue = 240x480, green = 320x240, red = 400x300, cyan = 500x375, magenta 640x480, yellow = 800x600, and black = 1024x768.

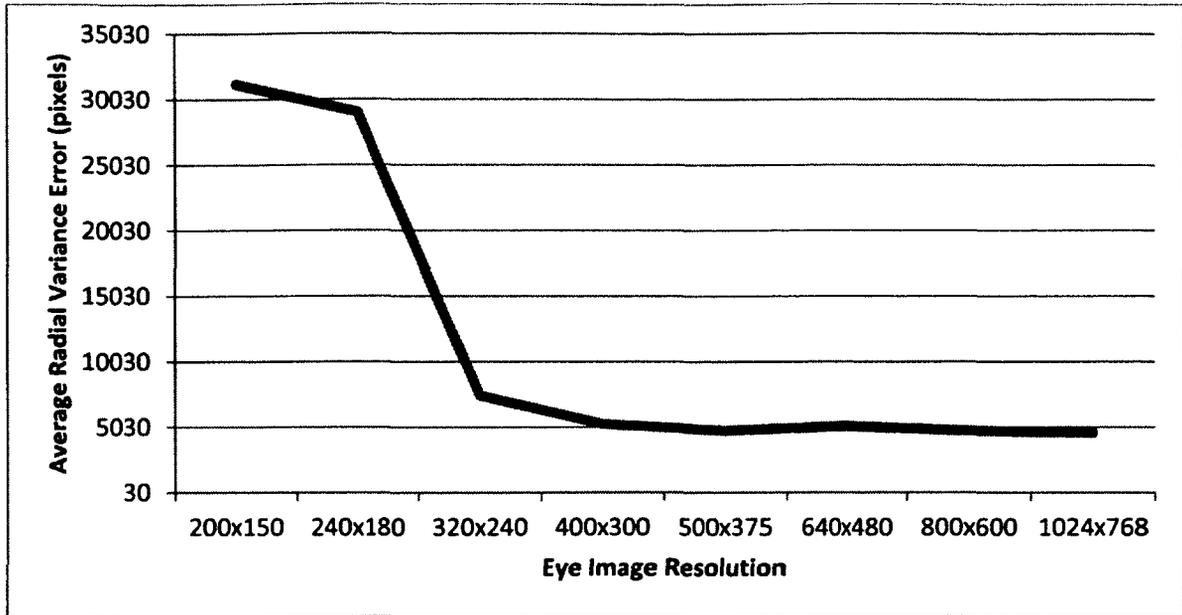


Figure 4-7: Graph showing the average radial variance of error of eye gaze estimates for all targets of participants P1 through P5 for each eye image resolution used.

4.4 Discussion

A large part of the error seen in the results can be attributed to a bias in the system. Recall that the system was being used without any calibration routine. If a calibration routine was performed for each participant prior to data collection, the bias could be eliminated; however, it would stand to reason that the trends mentioned hereafter would remain the same.

From the results, there was neither a large increase nor decrease in estimation error when increasing the eye image resolution from the camera default of 640x480 pixels up to 1024x768 pixels as seen in Figure 4-4. One would have expected that when using greater eye image resolution, the accuracy of the system would have increased; hence the estimation error would have decreased. The difference between the average radial RMSe of eye gaze estimates using 640x480 pixels and 1024x768 pixels was 4 pixels (1.5mm on the monitor screen). Looking at the average variance of the radial error for all the targets between the two eye image resolutions was approximately 7 pixels (2.5mm). The low average radial RMSe and

variance would suggest that there was no noticeable improvement in accuracy and precision.

In Figure 4-4, it shows that there was an increase in average RMSe (14.9 +/- 25.9 pixels / 5.43 +/- 9.46 mm) of the eye gaze estimate when decreasing the resolution of the images from 640x480 pixels to 240x180 pixels and a relatively large increase in average RMSe (188.6 pixels / 68.82mm) when using 200x150 pixel eye images. This behaviour was consistent across all five participants and appears to be largely influenced by the estimation of the x-coordinate (Figure 4-4a) as both have a similar trend. The radial RMSe of eye image resolutions 500x375 pixels to 240x180 pixels did not vary greatly. One would expect that the error would continue to increase relative to image size used. This provides some substantiating evidence that an increase in eye image resolution may not greatly increase accuracy in the system. It is important to note that experiments with the use of a higher resolution camera would need to be conducted to conclusively determine any effects of increasing eye image resolutions.

With the increase in eye image resolution, the distances between pupil features would be represented with more accuracy allowing a more accurate result when using the cross ratios to determine eye gaze. For instance, using a 640x480 pixel eye image, 1 pixel represented 0.125mm while in a 1024x768 pixel eye image, 1 pixel represented 0.833mm. The higher eye image resolution image would have been prone to less rounding error when measuring pupil features, which in turn would have resulted in a more accurate estimate when computing the x- and y-coordinate of the cross ratios.

It is important to note that the increase and decrease in eye image resolution in this experiment was synthetic, as it was achieved using bicubic interpolation. Effectively, the eye images were being upsampled for higher eye image resolutions, therefore, there was no real new information being added to the images. However,

by doing so, there was a possibility of seeing if the error would have increased. To truly see if the system can achieve more accurate eye gaze estimates using higher resolution eye images, a higher resolution camera would be needed. Even with the higher resolution camera, we would not expect there to be any large improvements since the relative improvements one could see would be small compared to the variability between different eye images captured. For lower resolution eye images, artifacts may be introduced in the eye images that could cause some error during eye gaze estimation. However, once again, these errors would be considered small compared to the variability of different eye images captured.

The eye gaze estimation system appears to be sensitive to feature detection found in the eye image plane. A pixel in an eye image will translate to several tens of pixels on the monitor plane. The following relationships can be seen between the pupil center in the eye image (u_{px}, u_{py}) and the estimated gaze point on the monitor plane (x_g, y_g) for 640x480 eye image pixel resolution images:

$$(u_{px}, u_{py}) \rightarrow (x_g, y_g)$$

$$(u_{px} \pm 1, u_{py}) \rightarrow (x_g \pm 70, y_g)$$

$$(u_{px}, u_{py} \pm 1) \rightarrow (x_g, y_g \pm 75)$$

If the pupil center estimate is one pixel off in the x-axis of the eye image, then there is an offset in the gaze estimate of approximately 70 pixels (18.12mm) in the x-axis on the monitor plane. Similarly, one pixel off in the y-axis of the eye image results in an offset in the gaze estimate of approximately 75 pixels (19.28mm) in the y-axis on the monitor plane. This would contribute to the decreasing estimation error seen for 320x240 pixel and 240x180 pixel eye image resolutions, where one would have expected it to increase.

In Figure 4-6, the radius of the variance circles increase with eye image resolution. Graphing for all participants, Figure 4-7 shows that the average radial

variance error is larger for smaller eye image resolutions compared to greater eye image resolutions used in this study. Variances are largest for the two lowest eye image resolutions tested, 200x150 pixels and 240x480 pixels. As the eye image resolution decreases, the offset due to incorrect pupil center estimates will increase because the relative distances between eye features will be decreased. Feature estimation errors for lower eye image resolutions will translate to larger errors on the monitor plane compared to feature estimation errors for higher resolution images. Because of this, the variance for eye gaze estimates will be larger for lower eye image resolutions.

4.5 Conclusions

Eye image resolution was not a large factor in the accuracy of eye gaze estimation until reducing the eye image resolution where the ROI was lower than 45x38 pixels. With lower resolution eye images, feature estimation translated to larger amounts of pixels on the monitor compared to feature estimation when using higher resolution eye images. If there was an error in feature estimation, the error of the eye gaze estimate on the monitor was magnified. During testing, when the ROI was larger than 45x38 pixels, the accuracy of the eye gaze estimation did not largely change. However, to conclusively determine the effects of higher resolution eye images, experiments with a higher resolution camera will need to be conducted.

When decreasing eye image resolution, even though the mean of the eye gaze estimate did not largely vary, the variance of eye gaze estimates increased. This could be problematic for applications that require a high level of precision. However, if an application can tolerate a lower level of precision, lower resolution eye images could be utilized. This would enable a system to be designed using less expensive imaging technology. A typical system would utilize a minimum of two cameras where one camera would focus on tracking a user's head to enable control

for another camera to focus on the user's pupil. If the system can tolerate a lower level of precision, then the complexity of the system can be reduced so that there is only a single camera tracking a user's head. Eye features could then be extracted from a zoomed in portion of that image. In essence, the effective pixels between features collected of the eye in the image would be reduced, similar to reducing the eye image resolution in the current implementation. Therefore depending on the available ROI, there is a trade off between the simplicity of the system and the accuracy and precision of the system.

5 Effects of Head Position

5.1 Introduction

This chapter presents the effects of different head positions on eye gaze estimates when using the implemented system based on cross-ratios. The method of data collection is first described. Results of the estimation accuracy are then presented when the head is in different positions. Next, a discussion is presented to explain the results acquired. The chapter will finally conclude with the findings from the experimental results.

5.2 Method

Data were collected from five participants (4 males and 1 female) who were right-eye dominant. The female (P2) and one male (P5) wore corrective contact lenses while the 3 other male participants (P1, P3, and P4) had no known vision disorders. As described in detail in Chapter 3.3, each participant sat 65 centimetres in front of the monitor screen with their head attached to the fixation device to minimize any head movement. Nine trials were performed with the head positioned in a different location relative to the monitor. Referring to the coordinate system in Figure 5-1, each participant's head was fixed along the z-axis (65cm from the monitor screen), while the position along the x- and y-axis changed for each trial, with the coordinate (0,0,0) being the physical location of the center of the monitor screen. Head movement to the right, relative to the user, was considered movement along the positive x-axis, while movement to the left was considered movement along the negative x-axis. Head movement upwards towards the ceiling was considered movement along the positive y-axis while movement downwards towards the floor was considered movement along the negative y-axis.

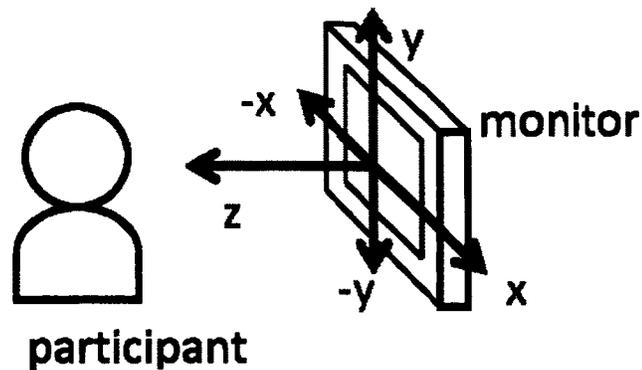


Figure 5-1: Coordinate system used when describing head position.

Participants' heads were placed in the head fixation device so the top bridge of their nose was located in space at the locations outlined in Table 5-1. These head positions would represent head movement of 10cm vertically (± 5 cm in the y-plane) and approximately 27cm laterally ($+10$ cm and -17 cm in the x-plane). Since the participants were right eye dominant, the head movements in the positive x-direction in trials 4 through 6 were shorter compared to the head movements in the negative x-direction for trials 7 through 9. This was to ensure that the camera would be able to capture usable eye images where eye features could be extracted. It also allowed the camera to have an even degree of rotation when panning from side to side since it focused on the dominant eye. The typical distance between the centers of the pupils in each eye for an adult male is 7cm, while for an adult female is 6.5cm [43]. Therefore, during trials 4 through 6, the right eye was positioned at approximately 13.5cm for males, or 13.25cm for females from the center of the monitor. With the increased distance in head movement for trials 7 through 9, the right eye was positioned 13.5cm for males or 13.75cm for females in the negative x-axis direction from the center of the monitor.

Five targets were presented individually on the monitor screen where each participant was asked to fix their gaze on. Referring previously to Figure 3-4, targets

1, 5, 8, 10, and 15 were presented respectively. This represented the upper left, upper right, center, lower left, and lower right regions of the monitor screen as seen by the participant. Eye images were captured from the camera at a resolution of 640x480 pixels.

Table 5-1: Positions of each participant’s head relative to the center of the monitor screen.

Trial#	Location the top bridge of each participants nose in space (x,y,z)
1	(0cm, 0cm, 65cm)
2	(0cm, 5cm, 65cm)
3	(0cm, -5cm, 65cm)
4	(10cm, 0cm, 65cm)
5	(10cm, 5cm, 65cm)
6	(10cm, -5cm, 65cm)
7	(-17cm, 0cm, 65cm)
8	(-17cm, 5cm, 65cm)
9	(-17cm, -5cm, 65cm)

5.2.1 Eye gaze estimation

Eye gaze estimation was computed offline using Matlab. For each participant, the gaze point estimation was made for each of the 5 target points for each trial as described in Chapter 3.4. The average root mean square error (RMSe) was computed between the location of the target on the computer monitor and the estimated gaze point. Eye images where a pupil could not be detected were discarded. Outliers in the data eye gaze estimates were removed with the fourth-spread method. More information on the discarding of samples can be found previously in Chapter 4.2.3.

5.3 Results

Table 5-2 displays the number of disregarded eye images of all targets of all participants P1 through P5 for each of the 9 trials. There were 50 eye images

captured for each target for a total of 250 eye images captured in each trial. The average number of discarded eye images after removal of blinks and outliers for each trial was 32.36 +/- 11.37. Table 3-1 displays the number of eye gaze estimates where a ghost point was used. The average number of ghost points used was 62.31 +/- 20.69 in each trial for all targets.

Table 5-2: The number of disregarded eye images of all targets for each participant in the 9 different head positions. A total of 250 eye images were collected in each head position.

Position	Disregarded Eye Images						
	P1	P2	P3	P4	P5	Avg	Std
1	35	35	34	35	43	36.40	3.71
2	43	30	28	59	51	42.20	13.33
3	45	40	42	46	65	47.60	10.01
4	29	29	3	29	50	28.00	16.67
5	31	29	11	41	7	23.80	14.32
6	36	21	11	24	44	27.20	12.95
7	30	33	20	36	28	29.40	6.07
8	33	33	9	33	31	27.80	10.55
9	30	24	9	31	50	28.80	14.75
Average of All Positions						32.36	11.37

Table 5-3: The number of eye gaze estimates where a ghost was used to estimate all targets for each participant in each head position.

Position	Ghost Points						
	P1	P2	P3	P4	P5	Avg	Std
1	42	82	56	81	81	68.40	18.39
2	20	74	74	59	65	58.40	22.39
3	71	102	56	58	71	71.60	18.39
4	44	61	27	45	104	56.20	29.30
5	62	83	30	56	18	49.80	25.95
6	52	95	48	76	101	74.40	24.15
7	70	46	53	55	61	57.00	9.03
8	73	72	51	72	111	75.80	21.74
9	20	57	54	52	63	49.20	16.84
Average of All Positions						62.31	20.69

Plots of the average estimated gaze points relative to target points for participants P1 through P5 can be seen in Figure 5-2 while Table 5-4 presents the

average and standard deviation of the radial RMSe of the clusters seen in each plot. Similar to the results presented in Chapter 4.3, estimated gaze points were shifted to the right and up from the target points.

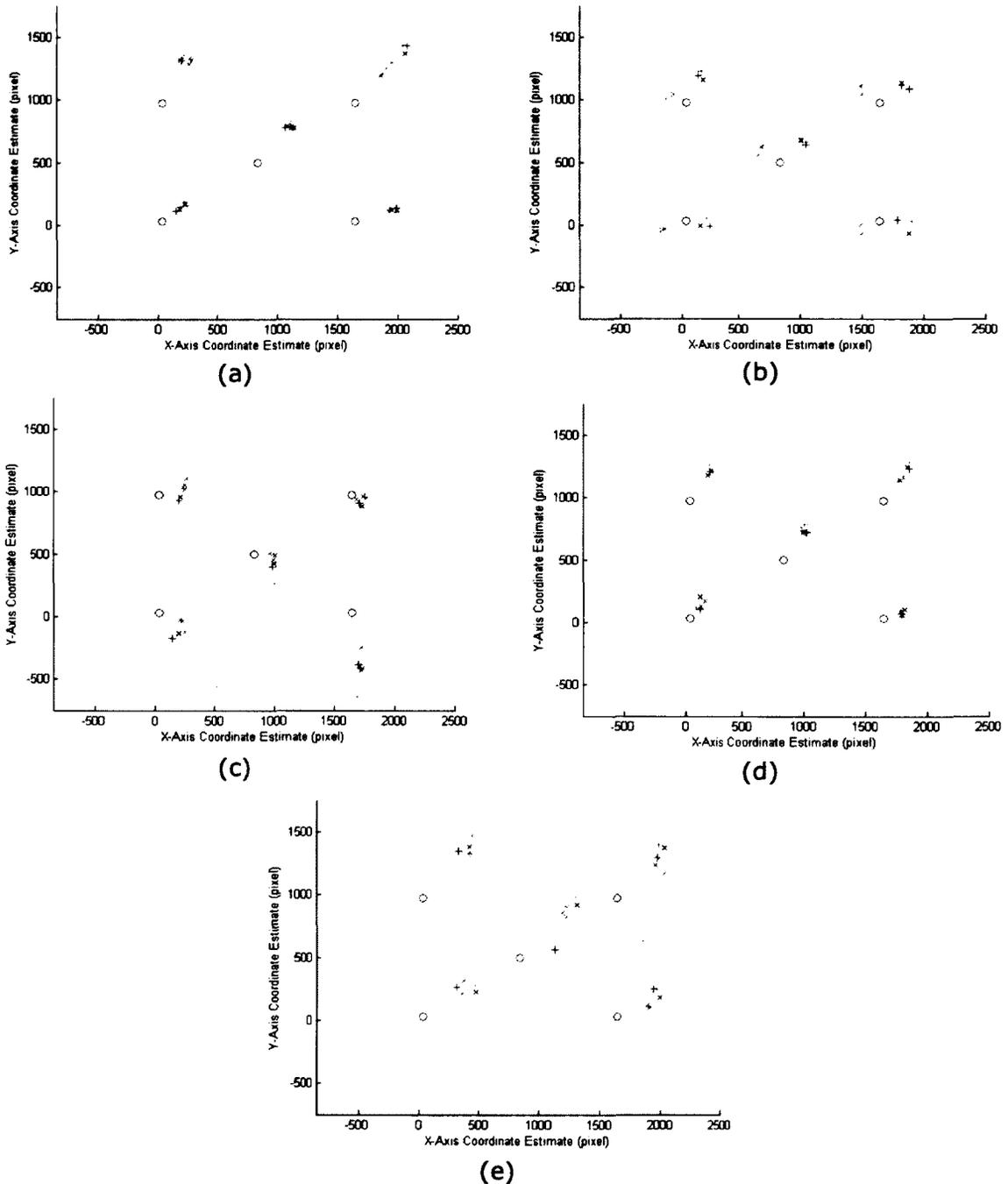
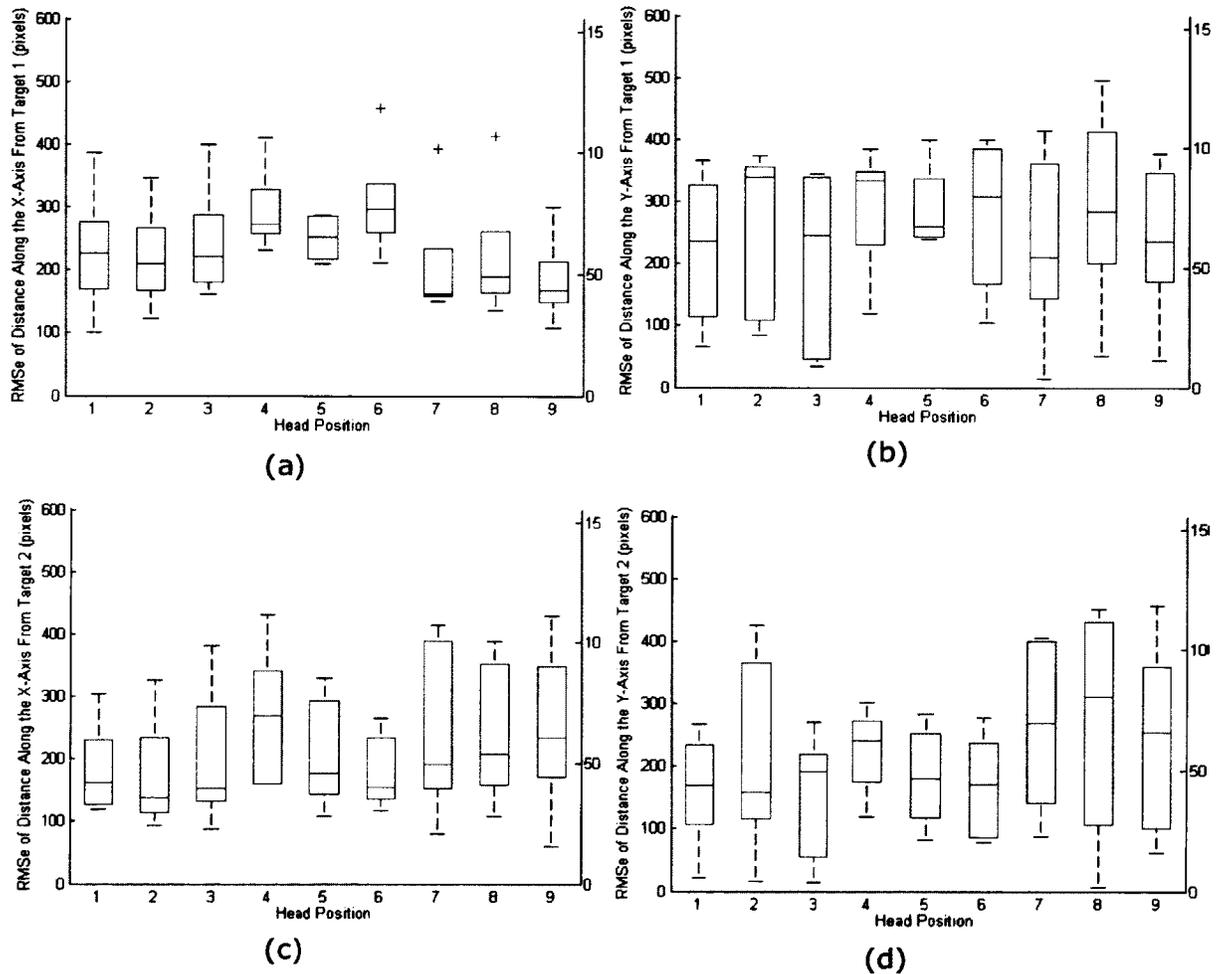


Figure 5-2: Plots (a) through (e) show the mean eye gaze estimates of each target for participants P1 through P5 respectively of all trials 1 through 9. Target positions are represented by the blue circles. Eye gaze estimates for trials 1 through 9 are represented respectively as follows: blue x, green x, red x, cyan x, magenta x, yellow x, black x, green cross, and red cross.

Table 5-4: The average and standard deviation of the radial RMSe of the 9 trials conducted for each participant for each target.

Target	P1		P2		P3		P4		P5	
	Avg	Std	Avg	Std	Avg	Std	Avg	Std	Avg	Std
1	427.75	44.85	254.83	93.34	247.20	62.35	340.94	54.36	547.41	56.49
2	443.27	126.71	209.87	30.32	157.97	139.33	266.15	61.90	454.45	60.37
3	396.61	25.98	253.89	62.62	179.86	47.82	308.26	36.08	564.23	145.62
4	235.87	53.80	225.11	72.65	282.55	182.35	212.56	116.43	456.95	65.42
5	312.93	32.63	208.80	45.85	470.58	113.32	178.59	73.37	350.20	116.65

Figure 5-3 displays the translational error characterized by the average RMSe of the estimated eye gaze in radial distance, along the x-axis and y-axis of all participants for each target for each of the 9 trials. Each trial will from here on after be referred to as a head position.



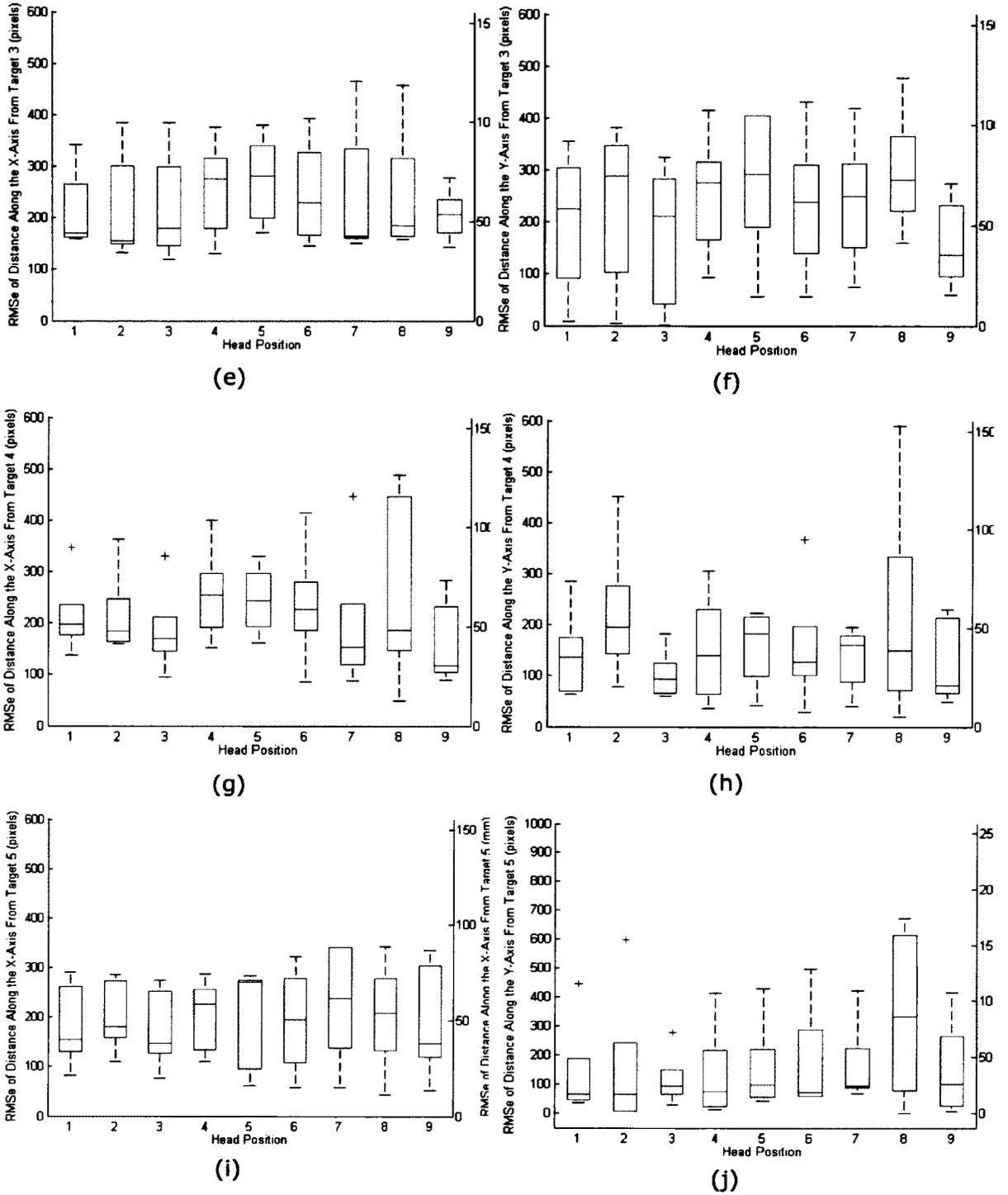


Figure 5-3: Box plots showing the average RMSe in the x-axis and y-axis for different head positions. Data for each box plot consists of the average RMSe along the x-axis or y-axis of all participants for one target. The red line in the box denotes the median of the data. The top and bottom edges of the blue box represent the 25th and 75th percentiles respectively. The black whiskers represent the maximum and minimum data range, while the red crosses represent outliers.

5.4 Discussion

Table 5-2 shows that the average number of disregarded images for each target did not increase significantly between different head positions. This would suggest that when the head was displaced as described in Section 5.2, the implemented system was capable of capturing eye images where eye features could be extracted. However, in

Table 5-3, the variance of the number of ghost points used when estimating eye gaze was large, suggesting not all four Purkinje reflections could be conclusively found in every head position for every target point. This led to some error with eye gaze estimation, as the fourth Purkinje reflection needed to be estimated by the system. Also, the average number of ghost points did not decrease as the dominant eye moved towards the center of the monitor as one would expect. This could result from a number of factors such as the LED sources used to create the Purkinje reflections not being powerful enough, or the glint detection algorithm not being robust enough to avoid discarding faint glints.

Similar to Chapter 4.4, a large part of the error seen in the results can be attributed to a bias in the system that could be removed with a calibration routine; however this does not affect the general trends observed. We can see in Figure 5-2(a)-(e) that plotting the mean eye gaze estimates for each trial created a cluster close to each target. The average standard deviation of the RMSE, which describes the radius of the cluster, was 76.82 pixels (28.03mm) where the minimum value was 25.98 pixels and the maximum value was 182.35 pixels. This translates to a minimum error of 0.95 cm and a maximum error of 6.65cm from the target on the monitor plane. These results show that the implemented system has a degree of precision that may be acceptable for applications that do not require a high degree of accuracy.

Analyzing Figure 5-3(b) and (c), we can create Karnaugh map-like diagrams to show possible relationships between the translational error in the x-axis and y-axis for each target and the head position participants are in. These are shown in Figure 5-4 (a) through (e) for targets 1, 5, 8, 11, and 15 respectively.

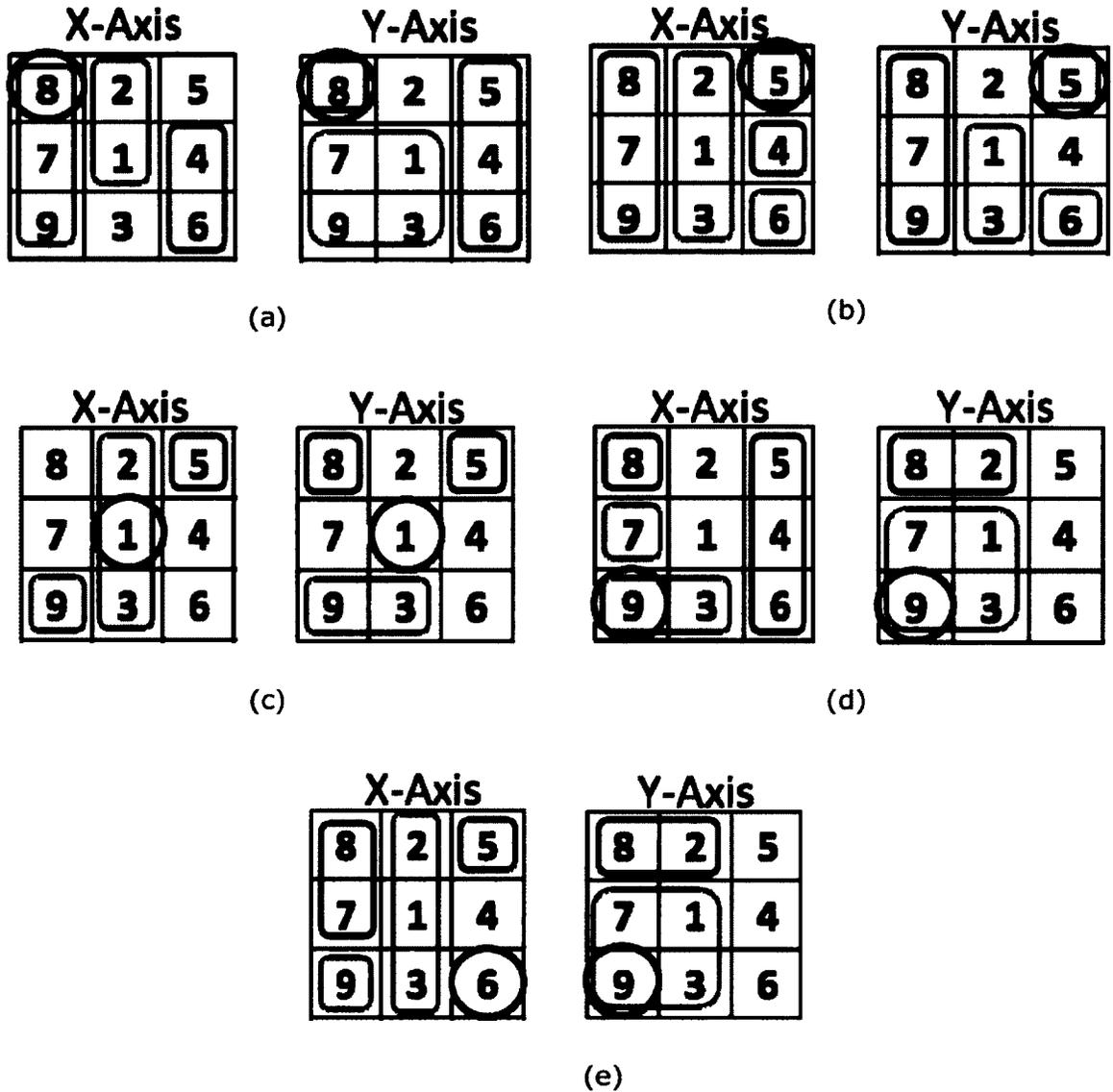


Figure 5-4: Venn diagrams showing relationship between the largest translation RMS errors (red), smallest translation RMS errors (green) and targets (blue) of (a) Target 1, (b) Target 5, (c) Target 8, (d) Target 11, and (e) Target 15. 9x9 squares and corresponding numbers represent the participant's head position relative to the screen.

Looking at the Karnaugh-map like diagrams, we notice the following general trends:

- For targets 1, 5, 11, and 15 representing the 4 corners of the monitor, the translational error along the x-axis appeared to be largest when the head position was located furthest away from the target along the x-axis.
- For target 8 representing the center of the monitor, the translational error along the x-axis appeared to be largest when the participant's head was in head position 5 (upper right corner).
- For targets 1 and 5 (upper corners), the translational error along the y-axis appeared to be largest when the participants' heads were located in head positions furthest from the target along the x-axis.
- For targets 8, 11, and 15 (center and bottom corners respectively), the translational error along the y-axis appeared to be largest when the participants' heads were located above the target along the y-axis.

Based on the small sample size ($n=5$), these findings would suggest that the eye gaze estimation error increased as the distance between the target and dominant eye increased.

5.5 Conclusions

In the implemented system, head positions where the top bridge of a user's nose was located anywhere between +10cm and -17cm along the x-axis, +5cm and -5cm along the y-axis, and 65cm along the z-axis from the center of the monitor, did not significantly affect the accuracy of the eye gaze estimation. When a user's head was placed in different locations, the system was able to capture the minimum number of eye features to provide an eye gaze estimation result; however the system was unable to capture all four Purkinje reflections for all head positions which led to some estimation error as the system estimated the fourth reflection.

For each participant in the experiment, the average eye gaze estimate in each head position formed a cluster near each target. Each cluster was offset from the target because of a bias in the system. The radiuses of the cluster varied slightly for each target and participant. Experimental results demonstrated that the system could have good eye gaze estimation accuracy, although it would suffer from poor precision when a user's head is not relatively fixed to the center of the monitor. With poorer precision, the implemented system could still be useful in human computer interaction (HCI) applications that employed an icon menu system. Icon widths and lengths could be made large enough where the variance of eye gaze estimates would become irrelevant if the user is to fix their gaze to the center of the icon.

The experimental results also suggested that the eye gaze estimation error usually increased when the distance between the target and dominant eye of a user increased in the x-direction and y-direction. Because of the small sample size, there was a large variance in the RMSE measured. A larger sample of data will need to be collected in order to conclusively determine the relationship.

6 Bias Compensation

6.1 Introduction

In Chapter 4 & 5, there was a large bias present in the data that could have been attributed to the omission of a calibration routine for each participant prior to the data collection. Many implementations ([1], [6], [15], [18], [33], [34]) require that each user calibrates the eye gaze system by focusing on targets on a monitor screen before the system estimates gaze locations. This allows the system to determine the bias and correct for it automatically when in use.

This chapter proposes a method to correct for the bias in the implemented eye gaze system without performing a calibration prior to use. Possible explanations for the bias seen in the data are first explained. The method for bias compensation is then presented. Afterwards, results of the estimation accuracy are shown with the previously collected data from the experiments performed in Chapter 4 & 5. This is followed by a discussion explaining the results acquired. The chapter will conclude with the findings from the proposed bias compensation method.

6.2 Bias in the System

One possibility for the bias in the system could be attributed to the finite size of each user's eye. Figure 6-1 shows a schematic diagram of the eye, where the major parts of the eye are labeled. Using Figure 6-1 as a reference, the eye's basic function is described as follows. The eye first allows light through the cornea and pupil. This light is then concentrated through the lens and focused onto the retina in the back of the eye. Located at the back of the eye directly behind the lens is the most sensitive part of the retina, the fovea. It is at the fovea where the eye translates light to electrical impulses that are then carried up to the brain [44].

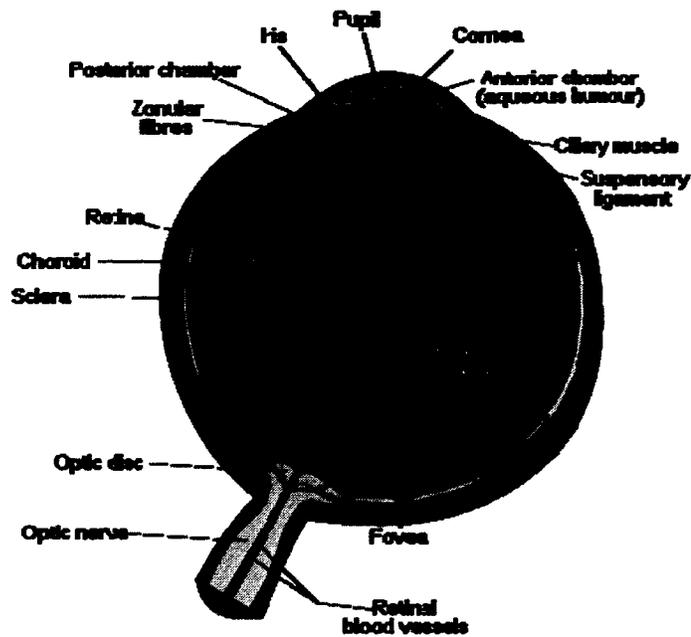


Figure 6-1: Schematic diagram of the human eye [45]

The implemented eye gaze system extracted the pupil location and the four Purkinje reflections from captured eye images. These images were captured from a camera located underneath the monitor positioned in the center of the bottom edge. Therefore, these images were approximately coplanar with the monitor screen, which allowed the system to make use of the cross-ratios method.

When the user's gaze is perpendicular to the monitor, the system should estimate eye gaze accurately, since a perpendicular line can be drawn from the monitor through the center of the pupil to the center of the fovea. However, when the eye is rotated in the eye socket, the information extracted does not accurately represent the location of the fovea, which is where the true eye gaze occurs. The gaze system will instead estimate the gaze perpendicular to the center of the pupil region it sees in the 2-dimensional eye image. We can see in Figure 6-2(a) that when a user's gaze is perpendicular to the screen, the pupil is approximately

centered to the centroid of the four Purkinje reflections, indicating that the eye is perpendicular to the screen. In Figure 6-2(b), the user's eye has rotated slightly and their gaze is no longer perpendicular with the screen. However, with the cross-ratio's method using a 2-dimensional image, the system will return an estimated gaze point as if the user's eye was perpendicular to the screen.

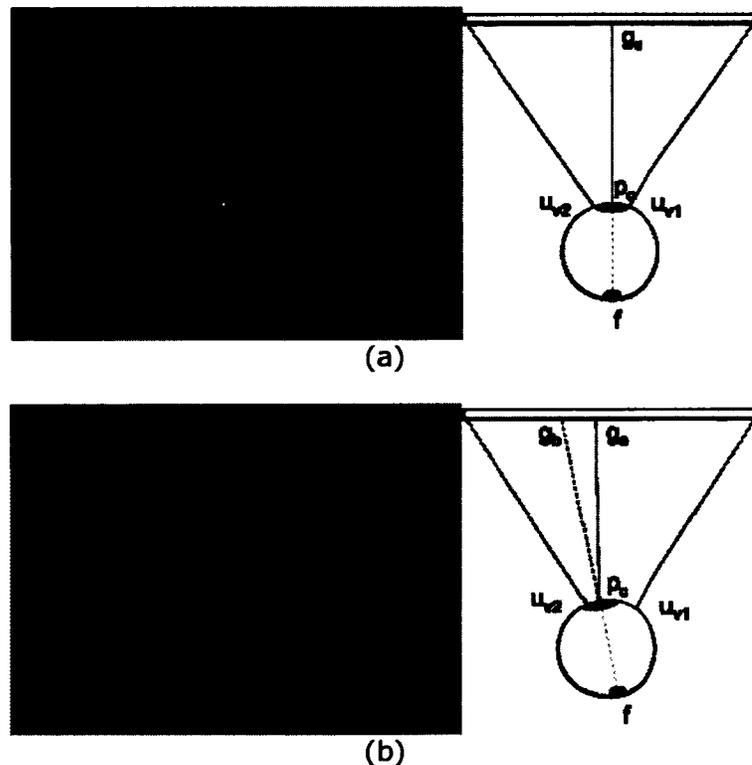


Figure 6-2: Picture of pupil with four Purkinje reflections. The participants' head and the light sources for the reflections are fixed in position and only the eye is allowed to move. In (a), the participant's gaze is on a target directly ahead on the monitor screen, while in (b) the participant is looking at a target to the side of the original target. In both figures: f represents the fovea, p_c represents the pupil center, u_{v1} and u_{v2} are two of the purkinje reflections, g_e is the estimated gaze point determined by cross ratios, and g_b is the true gaze point.

6.3 Bias Compensation Method

To eliminate the bias without running a calibration routine, we hypothesize that we can determine the amount of bias correction to add to an eye gaze estimate based on the angle of rotation and elevation of the eye to compensate for the

location of the fovea. By examining the angle of rotation in the x-axis and elevation in the y-axis we can estimate an error associated with the eye gaze estimate.

There are a few assumptions that need to be made in order to estimate the error of the eye gaze estimation. The first assumption is that the system can determine the position of the eye relative to the monitor. With an eye tracking or head tracking method to control the PTZ camera and with the position of the monitor and camera being fixed, it could be possible to estimate the relative position of the eye by determining the pan and tilt angles of the camera. Furthermore, the distance of the eye from the monitor could be estimated from features such as eye size. However, this method estimates the eye position based on the fixed four light sources producing the Purkinje reflections. It takes the centroid of the four light sources to represent the center of the monitor. This centroid is then used as a reference point with the pupil center to determine the difference in the x-axis and y-axis to produce rotation and elevation values. The distance of the eye from the monitor was fixed during data collection. The second assumption made is that the eye is a sphere with a known diameter. If we refer to Figure 6-1, we can see that the eye is not completely spherical, but instead a sphere with another rounded section on the surface making up the cornea. We set the value of the sphere to be 24mm, as defined by [6] as being the average diameter of the human eye. This is to avoid measuring each individual user's eye prior to using the system as this would be time consuming and not practical.

Examining Figure 6-3 of the eye, we can determine the rotation, θ , and elevation, φ , as the following:

$$\tan\theta = \frac{dx}{r} \quad (9)$$

$$\tan\varphi = \frac{dy}{r} \quad (10)$$

where: d_x is the distance between P_x and $\left|\frac{u^4_x+u^1_x}{2}\right|$

d_y is the distance between P_y and $\left|\frac{u^4_y+u^3_y}{2}\right|$

r is the pupil radius

When examining the eye in relation to the monitor, we can see the following:

$$\tan\theta = \frac{b_x}{D+r} \quad (11)$$

$$\tan\varphi = \frac{b_y}{D+r} \quad (12)$$

where: D is the distance between the eye and monitor screen

b_x is the bias needed to be added along the x-axis

b_y is the bias needed to be added along the y-axis

Rearranging and substituting (9) and (10) into (11) and (12) respectively, we get the following:

$$b_x = \frac{D+r}{d_x/r} \quad (13)$$

$$b_y = \frac{D+r}{d_y/r} \quad (14)$$

When we substitute (13) and (14) into the equations for the method proposed by [6] and described in Chapter 2, we obtain the equations (16) and (18) for the eye gaze estimates:

$$CR_{image}^x = \frac{\hat{x}_g + b_x}{w - \hat{x}_g} \quad (15)$$

$$\hat{x}_g = \frac{CR_{image}^x(w - b_x) - b_x}{1 + CR_{image}^x} \quad (16)$$

$$CR_{image}^y = \frac{\hat{y}_g + b_y}{h - \hat{y}_g} \quad (17)$$

$$\hat{y}_g = \frac{CR_{image}^y(h - b_y) - b_y}{1 + CR_{image}^y} \quad (18)$$

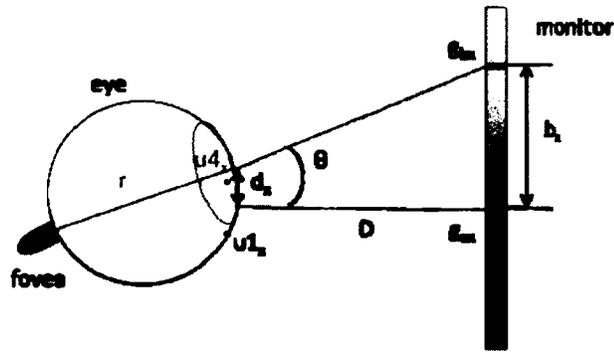


Figure 6-3: Eye gaze diagram used to determine the rotation of the eye where: r is the estimated radius of the pupil, d_x is the distance along the x-axis between the x-coordinate of the centroid created by the four Purkinje reflections and the x-coordinate of the pupil center, D is the distance between the user and the monitor screen, g_{ex} is the x-coordinate of the estimated gaze point, g_{bx} is the x-coordinate of the estimated gaze point with a bias correction added, and b_x is the biased distance needed to be added to estimated gaze point. The eye diagram used to determine the elevation is similar to the above with d_x , g_{ex} , g_{bx} and b_x replaced with d_y , g_{ey} , g_{by} and b_y respectively to represent the y-coordinate of the points.

6.4 Results

Using the data obtained in Chapter 4, a plot of the estimated eye gaze points with and without a bias compensation for each participant relative to the targets on the monitor screen is presented in Figure 6-4.

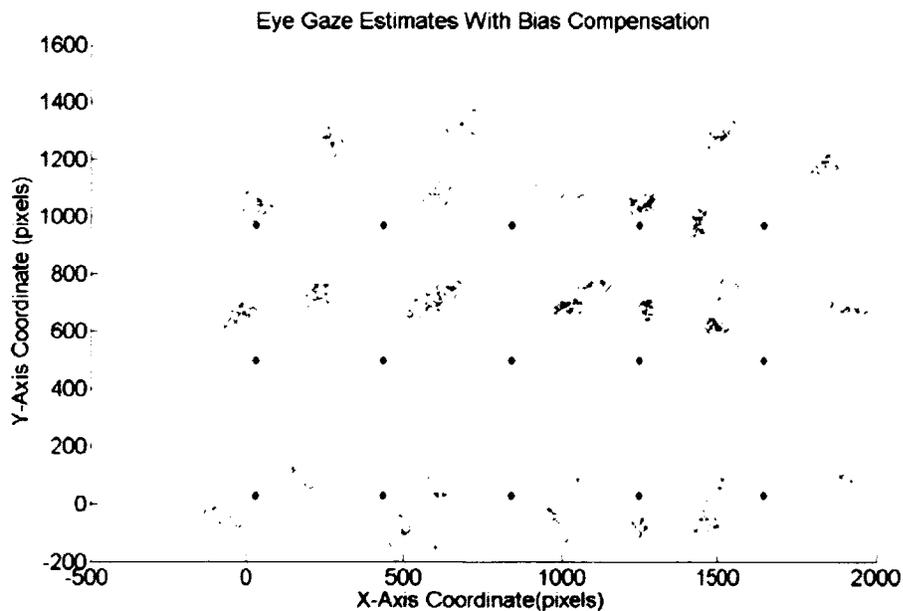


Figure 6-4: Eye gaze estimates for P1 using 640x480 pixel eye images. Magenta 'x' represent the original eye gaze estimates. Green 'x' represent the corrected eye gaze estimates with bias compensation.

Table 6-1 presents the average reduction in RMSe for the eye gaze estimates of each participant for each target using 640x480 pixel resolution eye images.

Table 6-1: Table displaying the average reduction in pixels of the RMSe along the x-axis, y-axis, and radial distance of participants P1 through P5 using the proposed bias compensation method using 640x480 pixel eye image resolution. A negative value expresses an increase in error.

Target	X-Axis (px)	X-Axis (mm)	Y-Axis (px)	Y-Axis (mm)	Radial Distance (px)	Radial Distance (mm)
1	179.8	46.55	170.4	43.81	255.5	93.23
2	79.3	20.53	200.7	51.60	198.0	72.25
3	95.3	24.67	184.4	47.41	204.5	74.62
4	182.1	47.15	155.8	40.06	242.1	88.34
5	-41.4	-10.72	104.0	26.74	26.6	9.71
6	140.8	36.45	74.7	19.21	156.2	57.00
7	92.7	24.00	83.6	21.49	130.1	47.47
8	94.6	24.49	77.0	19.80	126.7	46.23
9	198.0	51.26	78.7	20.23	182.9	66.74
10	32.8	8.49	56.1	14.42	58.4	21.31
11	86.1	22.29	-14.1	-3.63	39.5	14.41
12	81.9	21.20	-41.7	-10.72	22.7	8.28
13	98.7	25.55	-43.4	-11.16	37.9	13.83
14	182.0	47.12	-83.0	-21.34	46.2	16.86
15	12.3	3.18	-78.7	-20.23	-40.3	-14.71

Figure 6-5 is a box plot of the average RMSe of the radial distance of the estimated gaze point to the targets with and without bias compensation using different eye image resolutions. The data used for the box plot contain the average eye gaze estimate for all 15 targets for all participants. Each box represents the data using each eye image resolution.

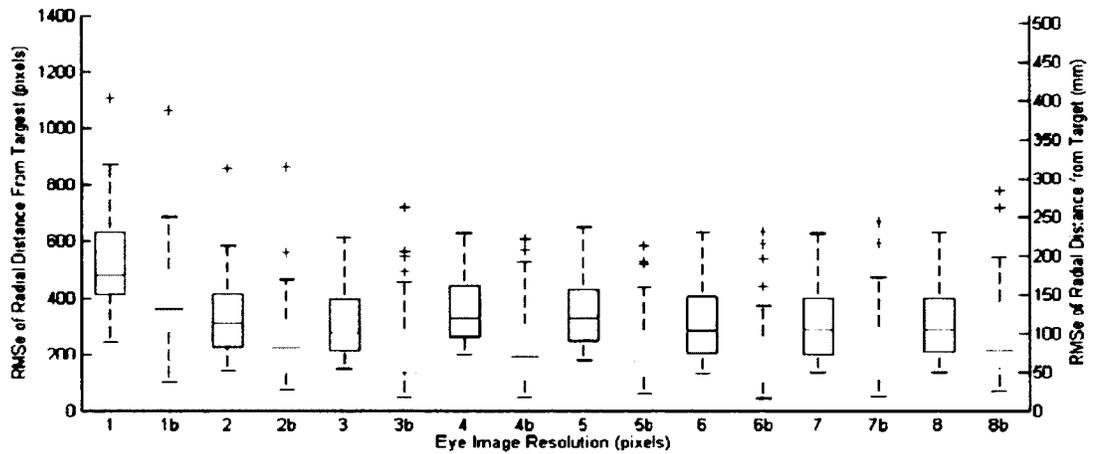


Figure 6-5: Box plots displaying the trend of the average RMSe of the radial distance from the targets as eye image resolution increases. Data for each box plot consists of the average radial RMSe for each target of all participants. Each box represents a different eye image resolution used with (yellow box) or without (blue box) bias compensation where 1 = 200x150, 2 = 240x180, 3 = 320x240, 4 = 400x300, 5 = 500x375, 6 = 640x480, 7 = 800x600, and 8 = 1024x768 pixel resolution. Eye image resolution box plots for eye gaze estimates using bias compensation are denoted with n (ie. 1b = bias compensation using 200x150 pixel resolution). The red line in the box denotes the median of the data. The top and bottom edges of the boxes represent the 25th and 75th percentiles respectively. The black whiskers represent the maximum and minimum data range, while the red crosses represent outliers.

Taking the average RMSe in the x-axis and y-axis for all the 15 targets for all participants and plotting box plots similar to that previously presented yields Figure 6-6.

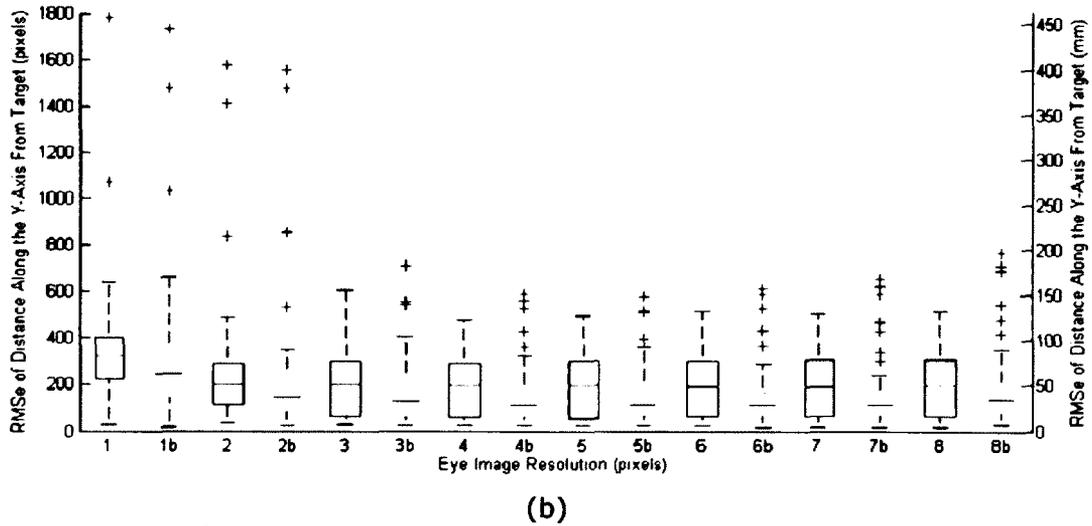
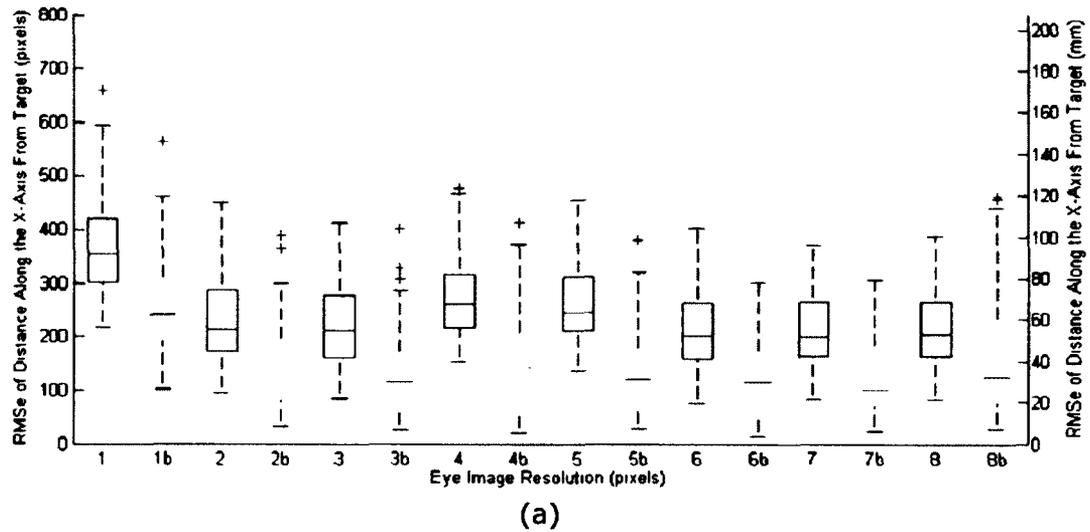


Figure 6-6: Box plots displaying the trend of the average RMSe in (a) x-axis and (b) y-axis as eye image resolution increases. Data for each box plot consists of the average RMSe along the x-axis or y-axis for each target of all participants. Each box represents a different eye image resolution used with (yellow box) or without (blue box) bias compensation where 1 = 200x150, 2 = 240x180, 3 = 320x240, 4 = 400x300, 5 = 500x375, 6 = 640x480, 7 = 800x600, and 8 = 1024x768 pixel resolution. Eye image resolution box plots for eye gaze estimates using bias compensation are denoted with n (ie. 1b = bias compensation using 200x150 pixel resolution). The red line in the box denotes the median of the data. The top and bottom edges of the box represent the 25th and 75th percentiles respectively. The black whiskers represent the maximum and minimum data range, while the red crosses represent outliers.

Figure 6-7 shows the average pixel reduction in RMSe of the eye gaze estimates for all the participants of all targets using different eye image resolutions.

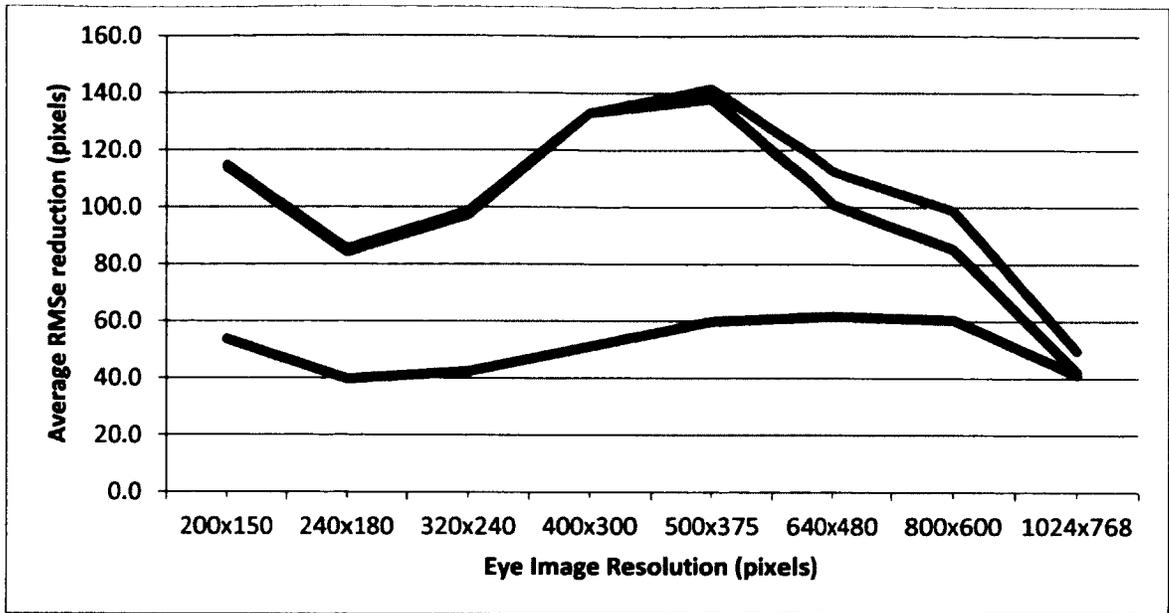


Figure 6-7: The average pixel reduction of RMSe along the x-axis (blue), y-axis (red) and radial distance (green). A negative number denotes that there was an increase in RMSe.

Figure 6-8 shows the average variance of error for original eye gaze estimates and bias compensated eye gaze estimates of all 15 targets for all participants using different eye image resolutions.

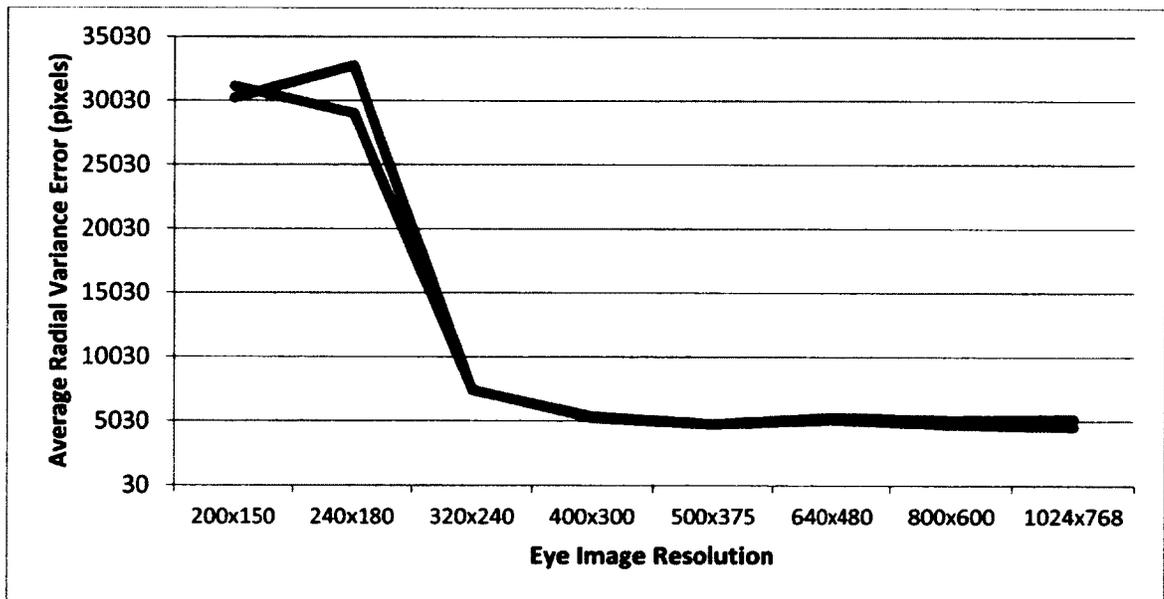


Figure 6-8: Graph displaying the average radial variance of error of all targets for all participants P1 through P5 for different eye image resolutions. Blue line represents the original eye gaze estimates while the red line represents the bias compensated eye gaze estimates.

Using the data obtained in Chapter 5 with participants in different head positions, plots of the average estimated gaze points relative to target points with and without bias compensation for participants P1 through P5 are presented in Figure 6-9.

Figure 6-10 presents box plots of the RMSe along the x-axis and y-axis of the estimated gaze points of each target with and without bias compensation in each head position. The data used for the box plot contain the average eye gaze estimate of each target for all participants.

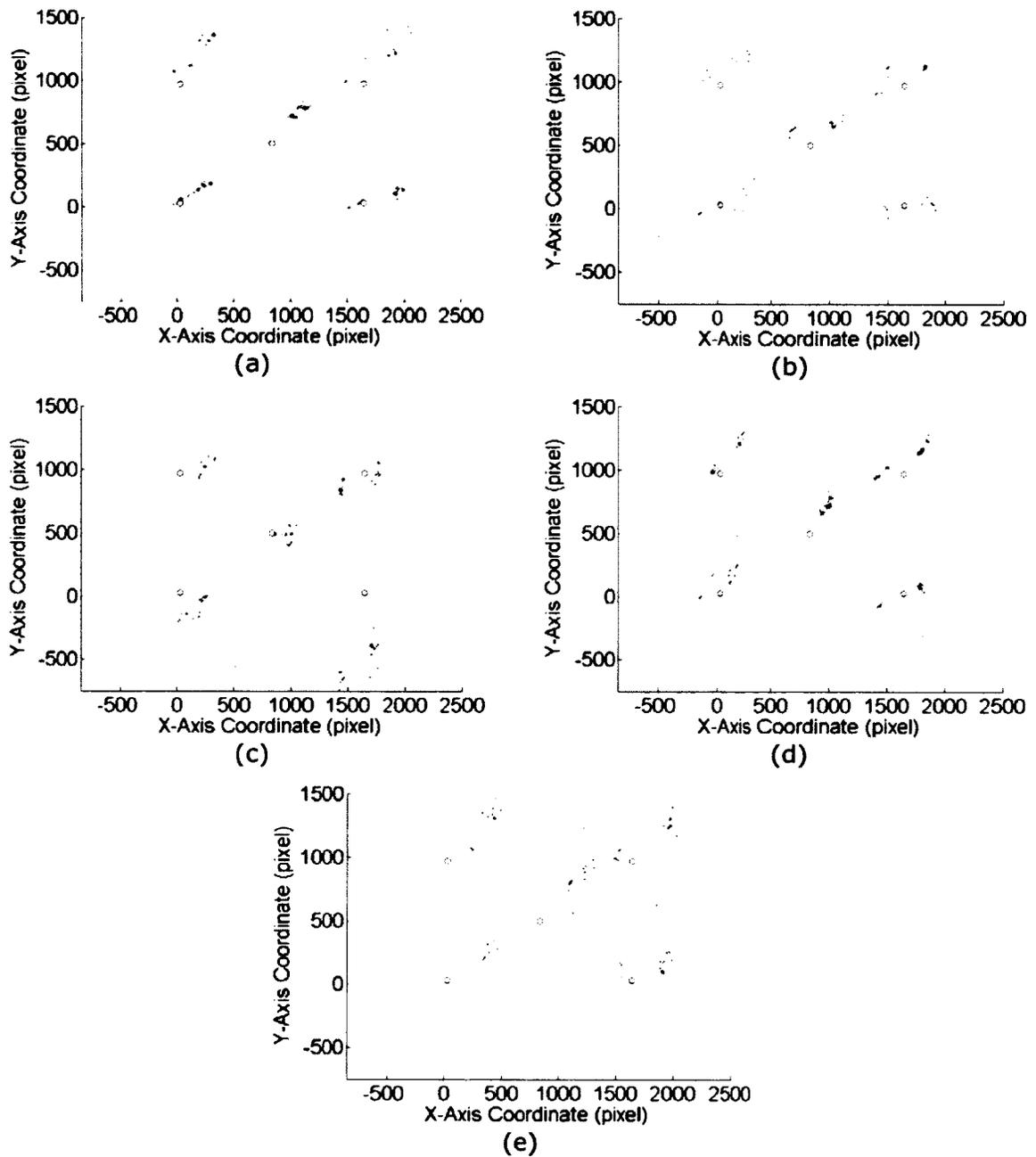
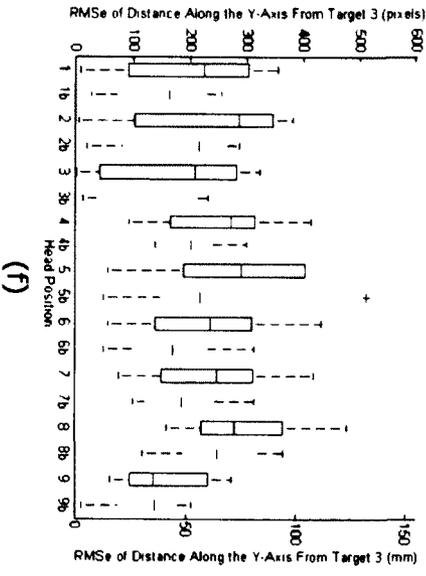
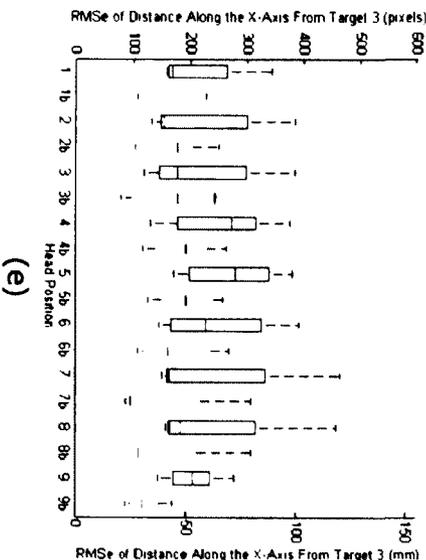
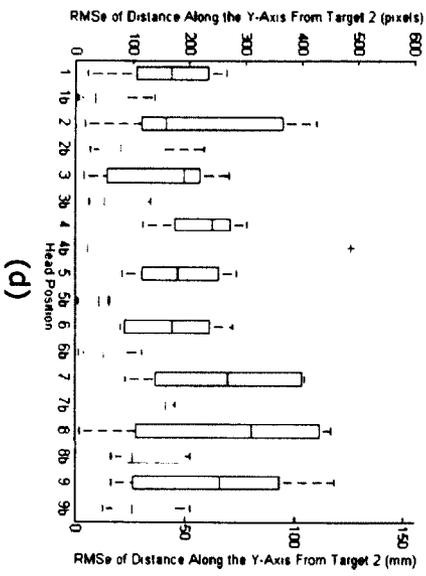
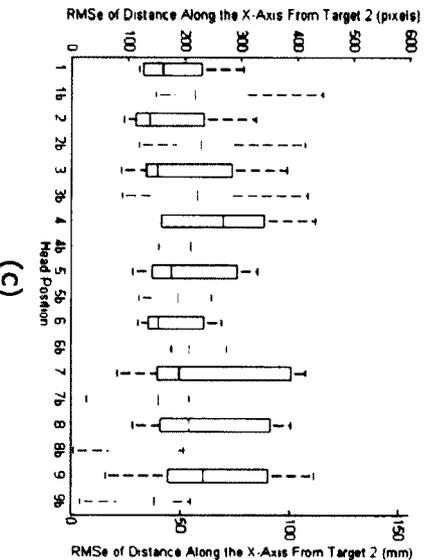
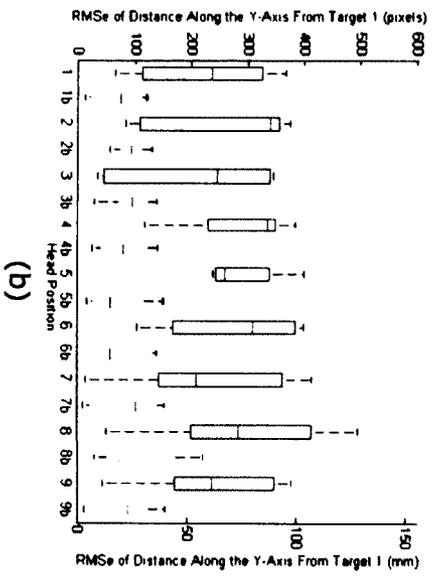
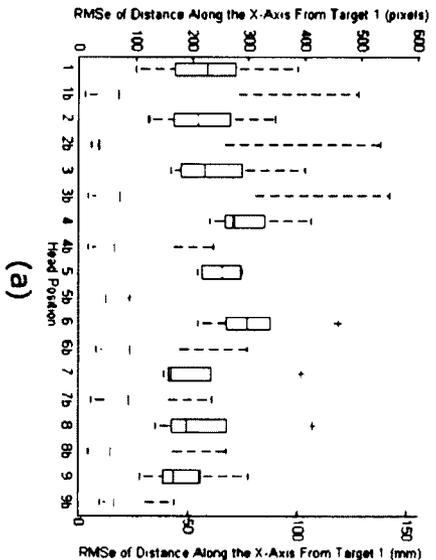
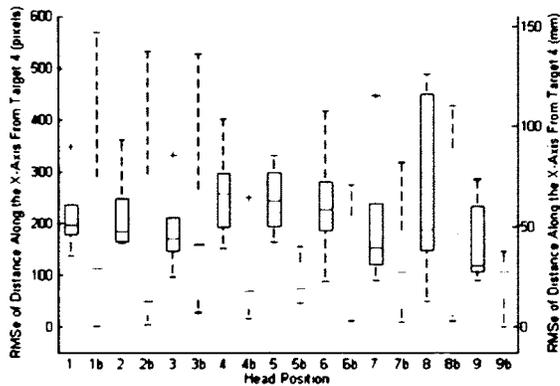
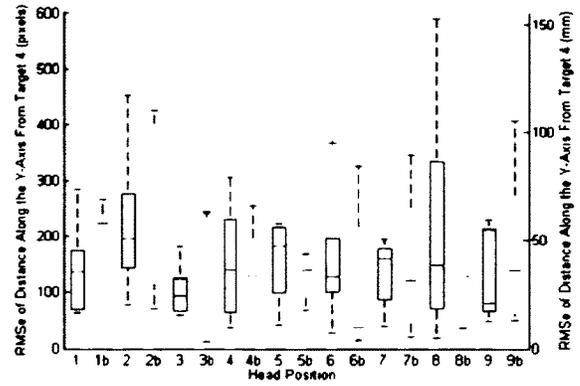


Figure 6-9: Plots (a) through (e) show the mean eye gaze estimates for each target for participants P1 through P5 respectively for all trials 1 through 9. Target positions are represented by the black circles. Original eye gaze estimates for trials 1 through 9 are represented as a blue 'x' while eye gaze estimates with bias compensation are represented by a red 'x'.

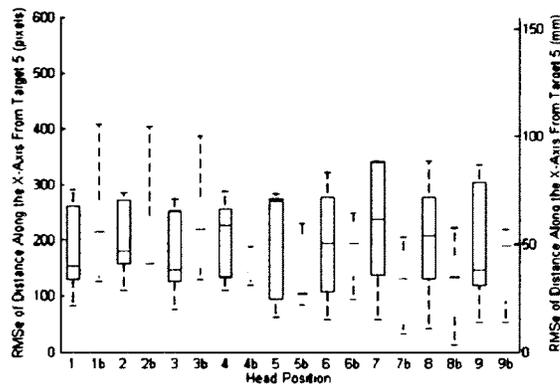




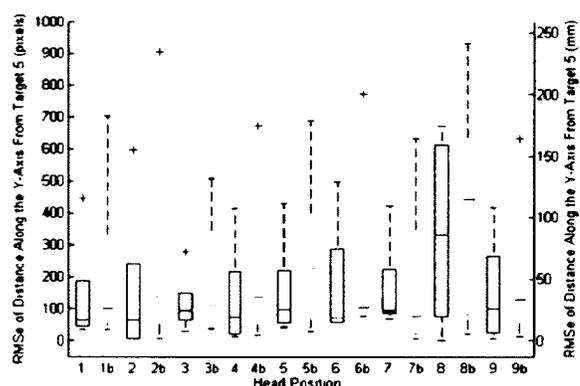
(g)



(h)



(i)



(j)

Figure 6-10: Box plots displaying the average RMSe in the x-axis and y-axis for different head positions with (yellow box) and without (blue box) bias compensation. Data for each box plot consists of the average RMSe along the x-axis or y-axis of all participants for each target. The red line in the box denotes the median of the data. The top and bottom edges of the boxes represent the 25th and 75th percentiles respectively. The black whiskers represent the maximum and minimum data range, while the red crosses represent outliers.

Table 6-2 displays the average pixel reduction in RMSe for the eye gaze estimates of each participant for the different head positions.

Table 6-2: Table showing the average percentage in reduction of the RMSe along the x-axis, y-axis, and radial distance of participants P1 through P5 using the proposed bias compensation method for different head positions. A negative value expresses an increase in error.

Head Position	X-Axis (px)	Y-Axis (px)	Radial Distance (px)
1	9.4	44.0	12.8
2	23.4	40.9	13.0
3	8.4	27.2	8.1
4	86.7	61.3	87.4
5	96.4	70.2	98.9
6	71.3	72.1	84.0
7	101.0	58.7	97.3
8	99.5	57.3	80.6
9	84.3	46.1	75.2
Avg ± Std	64.5 +/- 39.3	53.1 +/- 14.7	61.9 +/- 38.7

6.5 Discussions

Plotting the uncompensated eye gaze estimates formed a cluster that was shifted to the right and up from the targets. The shifts were expected, as there was horizontal error from the participants being right eye dominant and vertical error from the camera being positioned below the monitor. Analyzing Figure 6-4, we can observe that the bias compensation method shifted the cluster of eye gaze estimates to the left and down from its original position when there was no bias compensation. This brought the average eye gaze estimate closer to the target location for most targets, but not all. For all participants, the average estimated x-coordinate was closer to the x-coordinate of the target except for targets 5, 10, and 15 (far right column), which were the three furthest targets on the right. Instead, for these targets, the eye gaze estimates were shifted further past the target point to the left. This could be characteristic of right eye dominant users if the right eye did not fully rotate so that the target was in their gaze line. Participants may have been relying more on peripheral vision, which would have allowed them to focus outside the center of gaze.

The average y-axis coordinates varied between participants. In most cases, the eye gaze estimates were improved; however, there were occasions where the error had increased due to the bias compensation method. The largest increase in error of the y-axis coordinates occurred along the bottom row (targets 11 through 15). This could be a result of the eye image not being truly coplanar with the monitor. The system setup has the camera lens angled upwards to capture eye images since it is positioned below the monitor. As the participant's eye moves to focus on the bottom row targets, the offset that is calculated becomes greater than needed and as a result, overcompensates the eye gaze estimate.

From Figure 6-5 and Figure 6-6, we can see that the radial, x-axis, and y-axis RMSe decreased for different eye image resolutions when using the bias compensation method. We can observe that the error trend had shifted down, showing that the error had decreased from the bias; however, there were still errors associated with using different eye image resolutions. Accuracy improvements also varied with different eye image resolutions. The bias compensation method improved the x-coordinate of the eye gaze estimate by an average of 99.3 +/- 30.6 pixels (25.71 +/- 7.92mm) while also improving the y-coordinate of the eye gaze estimate by an average of 51.2 +/- 9.2 pixels (13.25 +/- 2.36 mm) for the eye image resolutions tested. The highest improvement radially occurred when using 500x375 pixel eye images (reduced RMSe by 133.05 pixels / 48.55mm), while lowest improvements occurred when using 1024x768 pixel eye images (reduced RMSe by 49.5 pixels / 18.06 mm).

Analyzing Figure 6-8, we can see that the variance of the eye gaze estimates using different eye image resolutions with and without bias compensation are approximately the same. This shows that with the bias compensation, the accuracy of the eye gaze estimates can be improved; however, the precision does not increase.

From Figure 6-9 we can see the cluster of average eye gaze estimates with bias compensation with the head in different positions had shifted to the left and down from the cluster of average eye gaze estimates without bias compensation. This demonstrates that the bias compensation method can be used with the user's head in different positions along the x-axis and y-axis. Analyzing

Figure 6-10 (a) through (j), we can observe the following:

- X-coordinate eye gaze estimates for targets 1, 3, and 4 improved with the bias compensation method. These targets were representative of the top left corner, center, and bottom left corner of the monitor screen respectively. For targets 2 and 5, the x-coordinate RMSe had increased. These targets represented the top right and bottom right corners of the monitor screen.
- Y-coordinate eye gaze estimates for targets 1, 2, and 3 improved with the bias compensation method, while there were increases with RMSe for targets 4 and 5.

These results are similar to what was observed in Chapter 5, where the estimation error increased as the distance between the target and dominant eye increased. Based on all head positions, the average reduction in RMSe of the x-coordinate was 64.5 ± 39.3 ($16.7 \pm 10.2\text{mm}$) pixels and of the y-coordinate was 53.1 ± 14.7 pixels ($13.7 \pm 3.8\text{mm}$).

The bias compensation method allowed the eye gaze system to reduce some of the error without having users calibrate the system before use. However, not all the error was eliminated, which can be attributed to the assumptions that were made earlier in Section 6.3. The first assumption was: knowing the exact position of the eye. Since this was estimated based on reflections of the eye, the method only holds true if the eye is perpendicular to the center of the monitor. When the eye is located

elsewhere, degrees of rotation and elevation will change slightly since the center of the monitor is used as a reference and is no longer perpendicular to the eye. Nonetheless, this method can still provide a reasonable estimate to determine rotation and elevation values. The second assumption was: the eye is perfectly spherical with a known diameter. In reality, the surface is not perfectly spherical, but rather spherical-like with another rounded section on top that forms the cornea. Therefore, the reflection from the LED light sources, L1 through L4, that lands off the cornea may not be the best representation of the corners of the LCD monitor, which would skew rotation and elevation angle estimates. It would also lead to errors when determining estimates using the cross-ratios method. An average diameter of the eye was also used to avoid measuring each individual's eye prior to use. This diameter will affect rotation and elevation values directly as smaller diameter eyes will have increased angles while the opposite would hold true for larger diameter eyes. Another error that may have been a factor with the bias compensation method is the index of refraction that occurs when light enters the eye. As light enters the cornea, it will slightly bend since it will be travelling from air, into the aqueous humour behind the cornea, through the lens, and once again through the aqueous humour behind the lens.

6.6 Conclusions

This chapter described a bias compensation method used to reduce the eye gaze estimation errors in the implemented system. This method involved determining the rotation and elevation of the user's dominant eye to correct for the fact that a user's gaze originates from the fovea and not at the surface of the pupil. Based on a sample of $n=5$, the method was able to reduce the average RMSE of the estimated x-coordinate by 99.3 ± 30.6 pixels (25.71 ± 7.92 mm) and the estimated y-coordinate by 51.2 ± 9.2 pixels (13.25 ± 2.36 mm) using the eight

different eye image resolutions tested. The method was also tested with data collected where user's heads were in different positions relative to the center of the monitor, giving favourable results where there was a reduction of average RMSE of the estimated x-coordinate by 64.5 ± 39.3 pixels (16.7 ± 10.2 mm) and of the estimated y-coordinate by 53.1 ± 14.7 pixels (13.7 ± 3.8 mm).

7 Thesis Summary and Future Recommendations

Our main objective of this thesis was the implementation of a simple eye gaze tracking system that would be cost effective and versatile in non-cooperative environments. The main focus was to explore the effects of eye image resolutions and head movements on the system and also determine a method to eliminate the use of a personal calibration process before operating the system. Section 7.1 summarizes the contributions drawn from this work and Section 7.2 presents recommendations for areas of future work.

7.1 Summary of conclusions

In this thesis, we implemented a simple low cost eye gaze tracking system using a single camera and five IR LED light sources. Eye gaze was estimated based on the cross ratios method. This system was implemented for approximately \$2400 USD, but could have been implemented for as low as \$100 to \$150 CAD, where both system costs are much more cost effective compared to commercially available systems on the market that range from \$5,000 to \$16,000 USD with the possibility of also being implemented for much less by choosing a different camera and eliminating the need for a frame grabber. With the system, we observed that there was some evidence to suggest that increasing eye image resolution above 640x480 pixels may not provide large gains in accuracy. By increasing eye image resolutions from 240x180 pixels to 500x375 pixels, the average radial RMSe between the eye image resolutions was found to decrease only by 5 pixels (1.82mm). This provides some evidence that with a higher resolution camera tracking the user's head, the system can extract the eye from the image which would lower the resolution of the eye but would still consist of an image resolution suitable for eye gaze estimation. We also observed that when the top bridge of the user's nose was located 65cm

away from the monitor, there was minimal affect on the accuracy of the system when there was head movement anywhere from 10cm to the right, 17cm to the left, 5cm up, and 5cm down. This further reinforces the system can be invariant to head movement. The original implementation of the system by Yoo and Chung [1] required user's to perform a calibration routine prior to use to eliminate any biases contained in the system. For this thesis, a bias compensation method was implemented to eliminate the calibration routine. This method showed promising results when tested as the average RMSe of the x-coordinate was reduced by 64.5 +/- 39.3 pixels (16.7 +/- 10.2mm) while the y-coordinate was reduced by 53.1 +/- 14.7pixels (13.7 +/- 3.8mm). These results can be compared with some of the previous systems mentioned in Chapter 2 in Appendix C.

7.2 Recommendations for future work

The focus of this research was to implement a low cost eye gaze system for non-cooperative environments. The effects of eye image resolution and head movement were also investigated and reported. During the research as well as the analysis process, several shortcomings became apparent that would be worthwhile to investigate further.

Z-Axis Head Movement

This research analyzed the effects of head movement with the eye gaze system; however, head movement was restricted to movement only in the x-plane and y-plane. In practice, the system will also need to accommodate movement along the z-plane. An investigation would need to determine the limitation of the system when the head would move forwards and backwards from the monitor screen. This could be implemented in a number of ways. One method could be to use a higher resolution camera and keep the camera with a farther field of view.

Based on features in the image, the system could estimate the distance the head is from the camera. Eye features could then be extracted to estimate eye gaze. Another method could have the camera track the size of the user's head and automatically zoom in and out so that the eye always fills the boundary box in the image captured. Depending on the zoom information from the camera, this could be calibrated to relate to distance from the monitor.

High Resolution Camera

The eye images were synthetically generated from the camera's max image resolution of 640x480pixels to observe the effects of higher resolution eye images. Even though there was some evidence based on downsampling eye images to suggest that increasing the eye image resolution would provide minimal increase in accuracy, we cannot conclusively say that it will not. Therefore, tests with a higher resolution camera would need to be investigated.

Real Time Implementation

This research focused on an eye gaze system setup that was designed to capture eye images and process them offline. For practical applications, the system would need to be implemented in a fashion that allowed for real time processing. The application program was developed within a Matlab environment that does not lend itself well to real time applications despite being very convenient for our investigative purposes. Eye extraction speeds and efficiency from captured images need to be further investigated to determine if the single camera approach can pan, tilt, and zoom quickly enough to allow for smooth usage.

Eye Model Improvements

For the bias compensation method, we implemented a simple eye model based on a simple shape and population averages for the eye. An investigation looking into developing a method for estimating eye size, possibly based on the size of the iris, could be conducted. The eye could also be modeled as intersecting spheres to be more representative of the true shape of the eye. This could compensate any errors that may be related to skewing of Purkinje reflections in eye images when the eye is rotated.

Larger Sample Size

This study had a sample size of $n=5$. Data from a larger sample of the population would need to be gathered to determine if the eye gaze system would behave similarly for different users based on gender, age, eye sizes, dominant eye, and known vision disorders. An investigation would also need to be conducted to determine if the system is applicable to user's wearing corrective eye glasses or contact lenses.

Appendix A: The letter of information and informed consent

Appendix A includes the letter of information and informed consent that was presented and signed by each participant prior to data collection.

Eye Gaze Tracking Research Letter of Information



Dr. Adrian D.C. Chan
Assistant Professor
Dept Systems & Computer Engineering
1125 Colonel By Drive
Ottawa ON K1S 5B6

Principal Investigator

Dr. Adrian Chan
Carleton University
Department of Systems and Computer Engineering
1125 Colonel By Drive
Ottawa ON K1S 5B6
Tel: 613-520-2600 ext. 1535
Fax: 613-520-5727
E-mail: adcchan@sce.carleton.ca

Feedback

Participants may contact one of the investigators for consultation of any concerns they have about the research. In addition, participants can also request any publications and information about the final results or conclusions from the study.

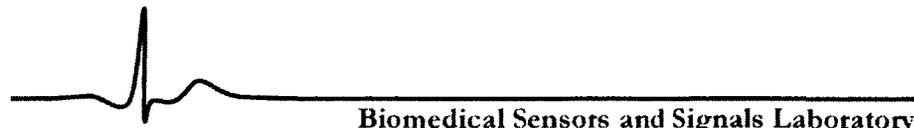
Purpose

This research will involve eye gaze tracking and applications involving eye gaze tracking.

Eye gaze can be used as an alternative interface modality to typical input devices, such as keyboards, mice, and voice. Eye gaze technology could be beneficial for persons with high level physical disabilities. Eye gaze technology could also be used by everyday users to enhance or augment existing interfaces.

Procedure

The exact procedure will depend on the specific study being performed. Participants will be fully informed of the complete and exact procedures of the study they are participating in, prior to obtaining consent.



The experimental component of this research consists of two stages: data collection and data processing. Only the data collection stage will require the presence of the subjects, as the data processing stage will be performed offline.

During the data collection stage, eye gaze tracking will be performed using infrared light and a video camera. The infrared light enables one to detect the pupil through the red eye effect and corneal reflections that provide a reference for eye position. The video camera is used to record the red eye effect and corneal reflections, and through image processing methods can determine eye gaze.

Participants will be asked to perform particular tasks. The nature of these tasks will be similar to one of the following:

1. remain with relatively still head position (a chin rest may be used)
2. gaze at different parts of a computer screen (this may be following a target or the subject may be free to selective gaze from a number of targets).

Participants may be asked some questions to ascertain some additional data, which may include:

1. age
2. sex
3. race/skin tone (the skin color can have a significant effect on some algorithms and therefore may need to be documented)
4. whether they have any known vision disorders
5. whether they have any known physical disabilities (if persons with high level physical disabilities are used in the research, information on the disability is required)

A data collection session will typically be around half an hour; however, may extend up to two hours, depending on the exact study being performed. Participants will be informed of the typical length of their particular study prior to the data collection session.

Some subjects may be asked to return for additional sessions but are in no way obligated to participate in these sessions.

Measurement Equipment and Risks

Infrared light

During the data collection stage, eye gaze tracking will be performed using infrared light and a video camera. The infrared light enables one to detect the pupil through the red eye effect and corneal reflections that provide a reference for eye position.



Risks: Whenever light is being used to illuminate the eye, there is a risk of damaging eye sight. With infrared light, the light is beyond a person's visible spectrum and cannot be detected.

There is an associated risk with the use of light to illuminate the eye. The lighting equipment has been evaluated to ensure that it conforms to Class 1 requirements defined in IEC 60825-1 and ANSI- Z-136, which are laser safety standards. While infrared light will be provided using LEDs, there does not currently exist a commonly accepted safety standard for their usage. The laser safety standards are considered more rigorous than what would be required for LEDs, since lasers provide coherent that is strongly directional. A Class 1 laser is considered safe under all conditions of normal use (i.e. maximum permissible exposure cannot be exceeded).

Photographs/Video camera

A camera is used to record the red eye effect and corneal reflections, and through image processing methods can determine eye gaze. This may be in the form of photographs and/or videos.

Risks: While there is minimal risk in the use of a video camera, its use does compromise anonymity of subjects.

Potential Benefits

There are no direct benefits or remuneration for participants in this study.

Photographs/Videos

Photographs/video (conventional/digital) may be required to perform eye gaze tracking and to document portions of the study. Usage of these photographs/videos will be restricted to scientific reports, presentations, or publications. Note the participant's anonymity will be maintained as best as possible; however, the usage of photographs/videos may compromise a participant's anonymity. The participant should be aware of this if consenting to the usage of photographs/videos.

Confidentiality

Information on the participant's sex, age, race (differences in race, including skin tone may be important in this research), and the presence of known vision disorders or physical disabilities may be used. Data will be kept electronically and participant consent forms and information will be kept as a hardcopy. Access to the data will be restricted to the researcher investigators specified by the ethics committee. Data may be shared with research associates at other academic institutions (excluding photographs and video, unless the permission of the participant is given). Data may be kept for an indefinite period of time.



Anonymity

Identity will be kept strictly confidential unless otherwise discussed with the researcher. Any scientific report, presentation, or publication of the data will refer to participants using a subject number; however, usage of photographs/video in publications will compromise anonymity.

Ethical review

This study has been reviewed and received ethics clearance through the Carleton University Research Ethics Committee in accordance to the Tri-Council Policy Statement for Ethical Conduct for Research Involving Humans. Participants with concerns or questions about their involvement in this study may contact the ethics committee chair.

Withdrawal

Participation in this study is strictly voluntary. Participants are free to withdraw from the experiment at any time and without any consequences.

Acknowledgments

The researchers would like to acknowledge the funding support of the National Sciences and Engineering Research Council of Canada.



Eye Gaze Tracking Research Informed Consent



Dr. Adrian D.C. Chan
Assistant Professor
Dept Systems & Computer Engineering
1125 Colonel By Drive
Ottawa ON K1S 5B6

I, _____ have been invited by Dr. Adrian Chan, and his research associates, of the Department of Systems and Computer Engineering at Carleton University, Ottawa, ON to participate in a study on eye gaze tracking.

This research will involve eye gaze tracking and applications involving eye gaze tracking. Eye gaze can be used as an alternative interface modality to typical input devices, such as keyboards, mice, and voice. Eye gaze technology could be beneficial for persons with high level physical disabilities. Eye gaze technology could also be used by everyday users to enhance or augment existing interfaces.

Research undertaken as part of a graduate/undergraduate course requirement:
YES NO

If YES specify course name and course number

Course Name	
Course Number	

Research undertaken as part of a graduate or undergraduate thesis research:
YES NO

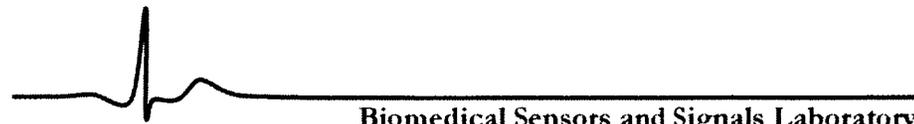
Measurement Equipment and Risks

Infrared light

During the data collection stage, eye gaze tracking will be performed using infrared light and a video camera. The infrared light enables one to detect the pupil through the red eye effect and corneal reflections that provide a reference for eye position.

Risks: Whenever light is being used to illuminate the eye, there is a risk of damaging eye sight. With infrared light, the light is beyond a person's visible spectrum and cannot be detected.

There is an associated risk with the use of light to illuminate the eye. The lighting equipment has been evaluated to ensure that it conforms to Class 1 requirements



defined in IEC 60825-1 and ANSI- Z-136, which are laser safety standards. While infrared light will be provided using LEDs, there does not currently exist a commonly accepted safety standard for their usage. The laser safety standards are considered more rigorous than what would be required for LEDs, since lasers provide coherent that is strongly directional. A Class 1 laser is considered safe under all conditions of normal use (i.e. maximum permissible exposure cannot be exceeded).

Infrared light will be used in this study: YES NO

If YES, I acknowledge this requires the use of infrared light

Photographs/Videos

A camera is used to record the red eye effect and corneal reflections, and through image processing methods can determine eye gaze. This may be in the form of photographs and/or videos.

Risks: While there is minimal risk in the use of a video camera, its use does compromise anonymity of subjects. Refer to the section on Anonymity for additional details.

Photographs will be used in this study: YES NO

If YES, I acknowledge this requires the use of photographs

Video will be used in this study: YES NO

If YES, I acknowledge this requires the use of video

Procedures

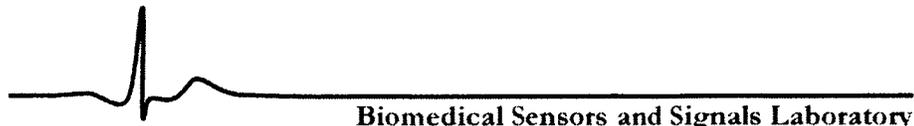
I may be asked some questions to ascertain some additional data, which may include:

1. age
2. sex
3. race/skin tone (the skin color can have a significant effect on some algorithms and therefore may need to be documented)
4. whether I have any known vision disorders
5. whether I have any known physical disabilities

I have been fully informed of the study that I am participating in: YES

I have been informed in advance that the study session for the data collections will be:

- approximately half an hour
- approximately one hour
- approximately two hours



I may be asked to return for additional sessions, and will be informed of this at the end of the study session. I am in no way obligated to participate in these additional sessions.

Should the experimenter note any unusual readings during the course of the experiment the study will be stopped immediately. The experimenter is not a physician and cannot make a medical diagnosis. I will be asked to contact my family physician. The researcher will contact my physician in writing explaining why the experiment was stopped. I may not return to the study or undertake any further experiments without the written consent of my physician.

Benefits

There are no direct benefits or remuneration for my participation in this study.

Confidentiality

Data will be kept electronically and my consent forms and information will be kept as a hardcopy. Access to the data will be restricted to the researcher investigators. Data may be shared with research associates at other academic institutions (excluding photographs and video, unless the permission of the participant is given). Data may be kept for an indefinite period of time. Information on my gender, age, race (differences in race, including skin tone may be important in this research), and the presence of known vision disorders or physical disabilities may be used. I understand that the data from this research will be used in scientific reports, presentations, and publications and my identity will remain confidential.

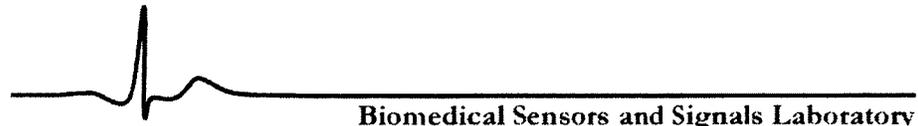
Anonymity

My identity will be kept strictly confidential unless otherwise discussed with the researcher. Any scientific report, presentation, or publication of the data will refer to me using a subject number; however, usage of photographs/video in publications will compromise anonymity.

Ethical review

This study has been reviewed and received ethics clearance through the Carleton University Research Ethics Committee in accordance to the Tri-Council Policy Statement for Ethical Conduct for Research Involving Humans. Should I have any concerns or questions about my involvement in this study I may contact the committee chair:

Professor Antonio Gualtieri, Chair
Research Ethics Committee
613-520-2517
ethics@carleton.ca



Withdrawal

Participation in this study is strictly voluntary. I am free to withdraw from the experiment at any time and without any consequences. I will also declare if the researcher can/cannot use the data I have provided should I decide to withdraw from the study.

Consent for Participation in the Study

I hereby agree to participate in this study and consent to the use of this research data in scientific reports, presentations, and publications with the understanding that my identity will remain confidential. I have read and understand the above explanation of the research procedure and all my questions have been answered to my satisfaction. I understand that I am free to withdraw from this research at any time and without any consequence.

Participant		Signature	
Experimenter		Signature	
Date			

Consent for Photographs/Video

I understand that photographs/video (conventional/digital) are required to perform eye gaze tracking and/or to document portions of the study. Usage of these photographs/videos will be restricted to scientific reports, presentations, or publications. I understand that my anonymity will be maintained as best as possible; however, the usage of photographs/video will compromise my anonymity.

Photographs will be used in this study: YES NO

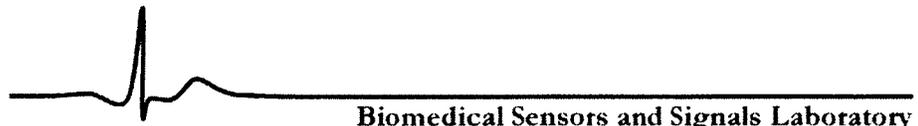
If YES photographs will be: CONVENTIONAL DIGITAL

Video will be used in this study: YES NO

If YES videos will be: CONVENTIONAL DIGITAL

I give consent for photographs/video in this study to be used in scientific reports, presentations, or publications (I am aware that this will compromise my anonymity as hiding identity in such cases will be difficult): YES NO

Participant		Signature	
Experimenter		Signature	
Date			



Appendix B: Porta Test Protocol

Appendix B includes the Porta test protocol that was used to determine the dominant eye of each participant prior to data collection.

The following steps are used to determine each participant's dominant eye.

1. The participant is asked to sit 200cm away from monitor screen of the eye gaze system and extend their arm in front of them.
2. Now the participant is asked to align their thumb or index finger with a target that is shown on the screen while keeping both eyes open
3. Now the participant is asked to alternate closing and opening their left and right eye.
4. When the other eye is closed, whichever eye does not cause the thumb or index finger to appear off target is considered the dominant eye.

Appendix C: System Comparison

Appendix C includes a table that compares the implemented system with some of the previous existing systems.

Table C-1: Table comparing the properties of the implemented system with those that have been previously been implemented in literature.

System	# Cameras	# Lights	Head Movement	Calibration	Accuracy (degrees)			Distance From Monitor (mm)	Sample Size
					X-Axis	Y-Axis	Radial		
Implemented	1	4	+10/-17cm x-axis +/- 5cm y-axis	0	5.06	4.63	6.80	650	5
Implemented w/ Bias Compensation	1	4	+10/-17cm x-axis +/- 5cm y-axis	0	3.60	3.43	4.90	650	5
[1]	2	2	200mm volume in X,Y, and Z direction	7	1.47	1.87	-	350	7
[6]	2	4	“significant head motion”	4	1.2	1.03	-	400-600	N/A
[15]	1	0	None	2	-	-	0.5	500	3
[16]	2	2	N/A	0	-	-	1.3	750	4
[23]	1	5	N/A	9	-	-	0.91	600	5

References

- [1] Z. Zhu, Q. Ji. "Novel Eye Gaze Tracking Techniques Under Natural Head Movement," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 12, pp. 2246-2260, Dec 2007.
- [2] S.T. Moore, I.S. Curthoys, T. Haslwanter. "Potential clinical applications of video-based eye position measurement," in *Engineering in Medicine and Biology Society, 1995., IEEE 17th Annual Conference*, vol.2, pp. 1627-1628 vol. 2, 20-23 Sept 1995.
- [3] L. Wang, D. Yu, F. Qiu, J. Shen. "A Digital Diagnosis Instrument of Hess Screen for Paralytic Strabismus," *Bioinformatics and Biomedical Engineering, 2007. ICBBE 2007. The 1st International Conference on*, pp1234-1237, 6-8 Jul 2007.
- [4] M. S. Atkins, A. Moise, R. Rohling. "An Application of Eyegaze Tracking for Designing Radiologists' Workstations: Insights for Comparative Visual Search Tasks," *Journal ACM Transactions on Applied Perception (TAP)*, vol.3, no. 2, pp. 136-151, Apr 2006.
- [5] S.M. Ali, L.A. Reisner, B. King, A. Cao, G. Auner, M. Klein, A.K. Pandya. "Eye gaze tracking for endoscopic camera positioning: an application of hardware/software interface developed to automate Aesop," *Medicine Meets Virtual Reality 16*, Los Pr Inc, 2008. pp. 4-7
- [6] D.H. Yoo, M. J. Chung. "A novel non-intrusive eye gaze estimation using cross-ratio under large head motion," *Computer Vision and Image Understanding*, vol. 98, no. 1, pp. 25-51, 2005
- [7] T.E. Doyle, Z. Kucerovsky, W.D. Greason. "Design of an Electroocular Computing Interface," *Electrical and Computer Engineering, 2006, CCECE '06. Canadian Conference on*, pp. 1458-1461, May 2006
- [8] R.G. Lupu, R.G.Bozomitu. V. Cehan, D.A. Cehan. " A New Computer-Based Technology for Communicating with People with Major Neuro- Locomotor Disability Using Ocular Electromyogram," *Electronics Technology (ISSE), 2011 34th Internation Spring Seminar on*, pp. 442-446, 11-15 May 2011.
- [9] T. Yagi, Y.Kuno, K. Koga, T. Mukai. "Drifting and Blinking Compensation in Electro-oculography (EOG) Eye-gaze Interface," *Systems, Man and Cybernetics, 2006. SMC '06. IEEE International Conference on*, vol. 4, pp. 3222-3226, 8-11 Oct. 2006
- [10] S. Handa, Y. Ebisawa. "Development of head-mounted display with eye-gaze detection for the severely disabled," *Virtual Environments, Human-Computer Interfaces and Measurement Systems, 2008. VECIMS 2008. IEEE Conference on*, pp. 140-144, 14-16 Jul 2008.

- [11] L. Yu, and M. Eizenman. "A new methodology for determining point-of-gaze in head mounted eye tracking systems," *Biomedical Engineering, IEEE Transactions on*, vol.51, no.10, pp. 1765-1773, Oct. 2004
- [12] V. Rantanen, T. Vanhala, O. Tuisku, P.H. Niemenlehto, J. Verho, V. Surakka, M. Juhola, J. Leikkala. "A wearable, wireless gaze tracker with integrated selection command for human-computer interaction," *Information Technology in Biomedicine, IEEE Transactions on*, vol.15, no.5, pp.795-801, Sept. 2011
- [13] D. Robinson. "A method of measuring eye movement using a scleral search coil in a magnetic field," *Bio-medical Electronics, IEEE Transactions on*, vol.10, no.4, pp.137-145, Oct. 1963.
- [14] D. W. Hansel, Q. Ji. "In the Eye of the Beholder: A Survey of Models for Eyes and Gaze," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol.32, no.3, pp.478-500, Mar. 2010.
- [15] A. Villanueva, R. Cabeza, S. Porta. "Eye tracking: Pupil orientation geometrical modeling," *Image and Vision Computing*, vol.24, no.7, pp.663-679, Jul. 2006.
- [16] D. Model, E. Guestrin. "An Automatic Calibration Procedure for Remote Eye-Gaze Tracking Systems," *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*, pp.4751-4754, 3-6 Sept. 2009.
- [17] Z. Zhu, Q. Ji. "Eye and gaze tracking for interactive graphic display," *Machine Vision and Applications*, pp.129-148.
- [18] E. Guestrin, and M. Eizenman, "General Theory of Remote Gaze Estimation Using the Pupil Center and Corneal Reflections," *Biomedical Engineering, IEEE Transactions on*, vol.53, no.6, pp.1124-1133, Jun 2006.
- [19] T. Ohno, N. Mukawa, and A. Yoshikawa, "FreeGaze: A Gaze Tracking System for Everyday Gaze Interaction," *Proceeding ETRA '02 Proceedings of the 2002 symposium on Eye tracking research & applications*, pp. 125-132, 2002
- [20] S. Shih, and Y.T. Wu, "A Calibration-Free Gaze Tracking Technique," *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, vol.4, pp.201-204, 2000.
- [21] A. Meyer, M. Bohme, T. Martinetz, and E. Barth, "A Single-Camera Remote Eye Tracker," *Proceeding PIT '06 Proceedings of the 2006 international tutorial and research conference on Perception and Interactive Technologies*, pp.208-211, 2006
- [22] C. Hennessey, B. Nouredin, and P. Lawrence, "A Single Camera Eye-Gaze Tracking System with Free Head Motion," *Computer Vision and Image Understanding*, vol.98, pp.52-82, 2005

- [23] F. Coutinho, and C. H. Morimoto, "Free head motion eye gaze tracking using a single camera and multiple light sources," *Computer Graphics and Image Processing, 2006. SIBGRAPI '06. 19th Brazilian Symposium on*, pp.171-178, Oct. 2006.
- [24] D. Beymer, and M. Flickner, "Eye Gaze Tracking Using an Active Stereo Head," *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, vol.2, pp. II-451 – 8 vol.2, 18-20 Jun. 2003.
- [25] K. Zhang, X. Zhao, Z. Ma, and Y. Man, "A simplified 3D gaze tracking technology with stereo vision," *Optoelectronics and Image Processing (ICOIP), 2010 International Conference on*, vol.1, pp.131-134, 11-12 Nov. 2010
- [26] "Mirametrix." <http://mirametrix.com/products/eye-tracker/>, 2011
- [27] "LC Technologies, Inc" <http://www.eyegaze.com/>, 2012
- [28] "tobii Eye Tracking Research." <http://www.tobii.com/>, 2011
- [29] "DynaVox." <http://ca.dynavoxtech.com/> , 2011
- [30] "H.K. EyeCan Ltd." <http://www.eyecan.ca/>, 2010
- [31] "EyeTech Digital Systems." <http://www.eyetechds.com/>, 2010
- [32] "Arrington Research." <http://www.arringtonresearch.com/>, 2012
- [33] E. Guestrin, and M. Eizenman, "Remote Point-of-Gaze Estimation with Free Head Movements Requiring a Single-Point Calibration," *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, pp.4556-4560, 22-26 Aug. 2007
- [34] A. Villanueva, and R. Cabeza, "A Novel Gaze Estimation System With One Calibration Point," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol.38, no.4, pp.1123-1138, Aug 2008
- [35] D. Model, and M. Eizenman, "An Automatic Personal Calibration Procedure for Advanced Gaze Estimation Systems," *Biomedical Engineering, IEEE Transactions on*, vol.57, no.5, pp.1031-1039, May 2010
- [36] S. Shih, Y. Wu, and J. Liu, "A Calibration-Free Gaze Tracking Technique," *Pattern Recognition, 2000. Proceedings, 15th International Conference on*, vol.4, pp.201-204, 2000
- [37] B. Nouredin, 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, pp.52-82, 2005

- [38] K.R. Park, "A Real-Time Gaze Position Estimation Method Based on a 3-D Eye Model," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol37, no.1, pp.199-212, Feb. 2007.
- [39] MATLAB version 7.10.0.499. Natick, Massachusetts: The MathWorks Inc., 2010
- [40] A. Goneid, S. El-Gindi, A. Sewisy. "A method for the hough transform detection of circles and ellipses using a 1-dimensional array," *Systems, Man, and Cybernetics, 1997. Computational Cybernetics and Simulation, 1997 IEEE International Conference on*, vol 4, pp.3154-3157, 12-15 Oct 1997.
- [41] S.Z. Li, *Encyclopedia of Biometrics*, Springer, 2009
- [42] "EM 1110-1-4014 Identification and Handling of Outliers", 31 Jan 2008, [Internet] Available: http://140.194.76.129/publications/eng-manuals/EM_1110-1-4014_sec/EM_1110-1-4014_Sections/a-i.pdf [Accessed: Sept 2011].
- [43] N.A. Dodgson, "Variation and extrema of human interpupillary distance," *Stereoscopic Displays and Virtual Reality Systems*, vol.5291, pp.36-46, Jan 2004.
- [44] "X-Plain Macular Degeneration Reference Summary", 25 April 2007, [Internet] Available: <http://www.nlm.nih.gov/medlineplus/tutorials/maculardegeneration/ot059103.pdf> [Accessed: Apr 2011]
- [45] Rhcastihos, "Schematic diagram of the human eye in english" *Wikipedia, The Free Encyclopedia*. 25 Jan 2007.