Point-of-Gaze Estimation in Three Dimensions

by

Craig Hennessey

B.A.Sc., Simon Fraser University, 2001
M.A.Sc., The University of British Columbia, 2005

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in

The Faculty of Graduate Studies

(Electrical and Computer Engineering)

The University Of British Columbia

July, 2008

Abstract

Binocular eye-gaze tracking can be used to estimate the point-of-gaze (POG) of a subject in real-world three-dimensional (3D) space using the vergence of the eyes. In this thesis, a novel non-contact, model-based technique for 3D POG estimation is presented. The non-contact system allows people to select real-world objects in 3D physical space using their eyes, without the need for head-mounted equipment. Using a model-based POG estimation algorithm allows for free head motion and a single stage of calibration. The users were free to naturally move and reorient their heads while operating the system, within an allowable headspace of 3.2 x 9.2 x 14 cm. A relatively high precision, as measured by the standard deviation of the 3D POG estimates, was measured to be 0.26 cm and was achieved with the use of high speed sampling and digital filtering techniques. When observing points in a 3D volume, large head and eye rotations are far more common than when observing a 2D screen. A novel corneal reflection pattern matching algorithm is presented for increasing image feature tracking reliability in the presence of large eye rotations. It is shown that an average accuracy of 3.93 cm was achieved over seven different subjects and a workspace volume of 30 x 23 x 25 cm (width x height x depth). An example application is presented illustrating the use of the 3D POG as a human computer interface in a 3D game of Tic-Tac-Toe on a 3 x 3 x 3 volumetric display.

Table of Contents

List of Tables

List of Figures

Acknowledgments

I would foremost like to thank Dr. Peter Lawrence for his support and guidance over the course of my graduate studies at UBC. Peter has always provided his time and effort to support my research goals, while still allowing me the freedom to pursue my own potential solutions (and failures), which was greatly appreciated. Peter made possible many of the opportunities I took advantage of throughout my graduate studies, including the teaching assistantships in PIP, TA awards, IEEE project fair judging, scholarships and conferences, for which I owe a debt of thanks.

I would like to thank the thesis committee members for their participation in the thesis process. I appreciate the time and effort provided for making this thesis the best it can be. I would also like to acknowledge the support and feedback from my fellow graduate students in the Robotics and Control Lab at UBC.

Finally I would like to thank my family and friends for their support and friendship over the years spent working on this thesis. In particular I would like to thank Julie for always encouraging me and for her understanding and patience.

Dedication

Dedicated to my family, my friends, and to Julie.

Statement of Co-Authorship

This thesis is based on several manuscripts, resulting from collaboration between multiple researchers.

A version of Chapter 2 appeared in the Proceedings of the 2006 Symposium on Eye Tracking Research & Applications. This paper was co-authored with Bourna Noureddin and Peter Lawrence.

A version of Chapter 3 appeared in 2008 in IEEE Transactions on System Man and Cybernetics, Part-B, and was co-authored with Bourna Noureddin and Peter Lawrence.

A version of Chapter 4 was submitted to IEEE Transactions on Biomedical Engineering. This paper was co-authored with Peter Lawrence.

A version of Chapter 5 has been accepted by IEEE Transactions on Biomedical Engineering and is currently in revisions. This paper was co-authored with Peter Lawrence.

The author's specific contributions in these collaborations are detailed below.

- Identification and design of research program: The research program was designed jointly based on ongoing discussion between the author and Peter Lawrence.

  Identification and design in specific research topics also depended on input from collaborators as follows:

  - The work performed by Bourna Noureddin during his M.A.Sc. program was useful in the development of the methods and system described in Chapter 2.
  - An algorithm for recording images at the full frame rate in the high speed system was developed and implemented by Bourna Noureddin for Chapter 3. This algorithm was used in debugging the high speed eye-gaze tracking system.

- Performing the research: All major research, including detailed problem specification, model design, performance of analysis and identifi-

cation of results was performed by the author, with assistance from Peter Lawrence.

- Data analyses: All numerical examples, simulation, and data analysis were performed by the author.
- Manuscript preparation: The author prepared the majority of all manuscripts, with the exception of the following:
  - Bourna Noureddin assisted with editing and suggestions in Chapters 2 and 3.
  - Peter Lawrence assisted with editing and various suggestions and improvements throughout this thesis.

Chapter 1

Introduction

The point of conscious attention of an individual can be used to provide insight into cognitive processes, information that otherwise may be difficult to obtain [1]. The point-of-gaze (POG) of a subject can be determined automatically by the use of eye-gaze tracking devices. With the real-time capabilities of modern eye-gaze tracking systems the use of eye-gaze has expanded from a purely diagnostic tool to applications in which the POG is used for control as well [2].

Using eye-gaze information as a control tool offers a number of potential advantages over alternative methods used for human computer interaction. Operation of the eye is intuitive as the link between the control of the visual system and the resulting retinal images is well established in the brain [3]. Eye movements are distinctly faster than hand-held pointing, as users typically look at the destination to which they wish to go before initiating the movement command [4]. Eye-gaze may also be the only form of communication possible for the severely disabled such as those with cerebral palsy, ALS and high level spinal injuries [5] [6] [7]. In addition to communication via on-screen keyboards in a computing environment, eye-gaze has also been investigated as a means for interaction with real-world objects, allowing a greater range of independence for the disabled [8]. The Attention Responsive Technology (ART) proposed by Shi *et al* uses a scene camera mounted on the subjects head, and eye-gaze tracking with dwell time selection to toggle on and off appliances such as a lamp, television or fan. The scene view is processed to identify and track valid controllable appliances using the Scale Invariant Feature Transformation (SIFT) technique [9].

Eye-gaze tracking has historically been used in the fields of psychology and physiology to link the natural movements of the eye to perceptual and cognitive processes such as learning, memory, workload, and deployment of attention [10]. As more user-friendly eye-gaze tracking systems were developed, their use expanded to commercial applications such as the analysis of driver awareness [11], advertising effectiveness [12; 13], website layout design [14], gaze contingent displays [15], enhanced mouse pointing [16], and assistive devices for the disabled [17] [18] [19].

Eye-gaze tracking is most commonly performed on a two dimensional (2D) surface such as a computer display. For 2D POG estimation, tracking a single eye is sufficient, as both eyes generally point to the same position [20]. If the position and orientation of both eyes are tracked however, the POG in 3D space can be determined from the intersection of the converging lines-of-sight of the left and right eyes [21]. If the 3D POG is known, it can be used in novel human machine interfaces such as interaction with 3D displays and as a means for the disabled to interact with 3D real-world environments [22].

The value of tracking the 3D POG will become increasingly important as 3D displays become more widely available [23] [24]. Human machine interfaces for interaction with 3D environments currently require multistage sequences of operations with the standard 2D computer mouse [25], or some form of 3D input device such as a stylus held by the user which is tracked optically or electromagnetically to determine the desired 3D input [26]. When tracking a stylus in 3D however, it must be held against gravity, eventually leading to user fatigue with extended use [27]. The 3D POG as an interface mechanism requires no physical effort greater than simply directing the gaze to the point of interest. The 3D POG additionally avoids the visual disconnect when the tracked tool cannot be physically located within the environment in which it is supposed to be acting [28].

Research into 3D POG estimation has been limited so far however, as a number of limitations inherent in 2D POG estimation remain. These 2D limitations are further exacerbated when extending POG estimation from 2D to 3D. Current areas of research include improving the accuracy and precision of POG estimation, decreasing the response time, or real-time capabilities, minimizing the time and effort required for user calibration and enhancing system reliability to handle various lighting conditions and differences between human users [29]. Developing eye-gaze tracking systems that are non-intrusive while still allowing for natural, unrestricted head motion is also a considerable area of focus.

1.1 Thesis Objectives

In this thesis the theoretical and current technical limitations for non-contact eye-gaze tracking are identified and novel means for improving upon these limitations investigated. The objectives of the thesis include:

- Improved 2D eye-gaze tracking: Existing limitations prevent the successful development of 3D POG estimation. Overcoming these limita-

tions improves both 2D POG estimation and enables remote 3D POG estimation.

- Remote 3D eye-gaze tracking: Based on the refinements developed for 2D POG estimation, techniques for remote 3D POG estimation were developed.

The motivation and requirements for the 2D refinements are: 1) improved usability of the system with non-contact free head motion eye-gaze tracking, 2) improved precision, latency and reduced signal aliasing with high speed POG sampling and filtering, and 3) improved reliability of image feature tracking with binocular eye tracking, fast face tracking, and multiple redundant corneal reflection tracking. The system developed for 3D POG estimation has the same requirements as the 2D POG estimation, including non-contact, free head motion, high speed sampling for improved precision, and reliable image feature tracking.

In the course of achieving the objectives of this thesis, the following contributions were made:

- Model-based POG estimation: A novel monocular 2D POG estimation method based on a simplified eye model was developed which allowed for free head motion, without requiring contact with the user's face or eyes.

- High speed sampling: High speed image processing techniques using software and hardware regions-of-interest (ROI's) were developed for significantly increasing the update rate of monocular POG estimates. The high speed sampling was shown to improve response times, reduce aliasing of the POG estimates and improve precision with high speed filtering.

- Binocular tracking enhancements: A high speed face tracking method was developed for differentiating the left and right eyes when only a single eye is visible to the system. A novel technique for tracking multiple reflections off the surface of the cornea was developed for enhancing the reliability of image feature tracking with large eye-rotations. As well, the Pupil-Corneal Reflection vector method for POG estimation was enhanced and contrasted with the performance of the model-based method.

- 3D POG estimation: A high-speed, binocular, model-based method was developed for real-world 3D POG estimation requiring only a single stage of calibration.

- 3D POG application: A demonstration application (3D Tic-Tac-Toe) was developed using the 3D POG for interaction with a point-based 3D volumetric display.

In the remainder of this chapter an overview of the literature in eye-gaze tracking will be presented, providing background material and motivation for the eye-gaze tracking research undertaken. The basic types of eye movements will be presented along with their potential impact on eye-gaze tracking. An overview of historical and contemporary eye-gaze tracking methods will then be presented, covering contact-based and remote methods. The current state of the art in 3D displays and interface methods will then be presented. Finally an overview of the remainder of this manuscript based thesis will be presented describing the contents of the following chapters, and their relation to the overall thesis goals outlined above.

1.2 Eye Movements

The movements of the eye have been extensively studied and a number of distinct patterns have been identified [2]. In the context of interactive eye-gaze tracking the eye motions of interest are fixations during which the sensory system collects information for cognitive processing, and saccades during which the eye is reoriented to observe new objects of interest. When observing points on a computer screen the eye accommodates to focus on the surface of the screen, with both eyes converging on the point of interest. Specialty eye movements such as smooth pursuit and nystagmus are not often found in the normal interaction between a user and a desktop monitor [2].

Our perception of the surrounding world during a fixation, lasting from 200 to 600 ms, appears stable, however, the images formed on the retinas of the eyes are constantly changing due to natural head and eye motions. The size of the fovea, or high resolution portion of the retina, is approximately 1° of visual angle which roughly corresponds to the size of variations of the eye during a fixation [30]. The eye exhibits a slow drift as well as small translations due to head motions which are corrected with fast shifts in eye orientation called microsaccades. The microsaccades keep the point of interest located within the foveal region of the retina. Microscaades have a typical amplitude of less than 0.1° of visual angle and a frequency of oscillation of 2 to 5 Hz. Superimposed on this motion is a tremor with a typical amplitude of less than 0.008° of visual angle with frequency components from 30-100 Hz and at times up to 150 Hz [31]. These small eye motions

during a fixation are thought to be required to continuously refresh the sensors in the eye, as an artificially stabilized image will fade from view [32]. The small eye motions result in fluctuating POG estimates which appears contrary to the stability in the POG expected by the user.

Saccades are the large motions of the eye which are used to reorient the fovea to another area of interest. Saccades most frequently travel from 1 to 40° of visual angle and last 30 to 120 ms. Between saccades there is typically a minimum of a 100 to 200 ms delay [30]. The sensitivity of the eye to visual input is reduced during a saccade [10] and as such, the POG estimates computed while the eye is in motion during a saccade do not correspond to conscious POG positions.

Both eyes do not always move in unison; depending on the depth of the object of interest the eyes will converge or diverge to center the images on the fovea of the eyes. The converging or diverging of the eyes (known as vergence) positions the images on the fovea of the left and right eyes to create binocular fusion [33]. In addition to re-orienting the eye to a point of interest, the image must also be focused upon the retina. Accommodation of the eye is the means by which the ciliary muscles compress or expand the flexible lens in the eye to change the focal depth of the eye [34]. When observing a standard computer monitor there is little change in depth and the focal length of the eyes remains relatively constant. To observe points in 3D space at different depths however, the eyes must both accommodate as well as converge or diverge. While accommodation occurs within the eye and is not externally visible, the vergence of the eyes can be tracked externally to determining the subject's intended POG in 3D space.

1.3 Eye-gaze Tracking Systems and Methods

1.3.1 Contact-Based Methods

Eye-gaze tracking has been a tool used in physiological and psychological studies for over a century. Quantitative methods were developed in the late 1800's in which plaster of Paris rings were attached directly to the cornea and mechanically coupled to pens [35]. In the early 1900's non-invasive methods were developed using light reflected from the eye and recorded on a falling photographic plate, capturing only horizontal movements [36]. Motion picture photography was later used to track the motion of the eye in both horizontal and vertical directions [37].

The development of electronics enabled methods such as electrooculography (EOG) and the scleral search coil method. EOG systems use electrodes

attached to the face around the eye to measure small DC potentials that vary with eye movement [21]. For the scleral search coil method, the motion of the eye is determined by applying a coil of wire, embedded in a contact lens, to the subject's eye. Measurements are taken as the eye moves the coil through an externally applied magnetic field and the position of the contact lens and consequently of the eye is determined. Both methods are electrical in nature, which allows for very high sampling rates using analog-to-digital conversion integrated circuits. The methods however are considerably intrusive as they require contact with the subjects face or eye and therefore EOG and the sceleral search coil methods are not typically used outside of laboratory environments. For a more detailed survey of contact based methods see Young and Sheena [38].

1.3.2 Video-Based Methods

Pupil Corneal-Reflection POG Estimation

Optical methods have been developed to remotely image the eye. Early systems however required a fixed head-to-camera distance which was difficult to achieve [2]. Head mounted systems are susceptible to slippage which can require frequent recalibration, as well, the weight of the system may result in fatigue if used for an extended period of time. The early remote optical systems avoided placing system elements on the subject's head, however, to constrain the position of the head, bite bars and chin rests were needed, resulting in a restrictive and intrusive system to use.

To determine the subject's POG based on images of the eyes, the position of the center of the pupil was tracked in the recorded images which, after a calibration procedure, was used to estimate the POG on a planar surface [21]. The pupil center only method requires a strictly rigid eye to camera displacement. An improved method for POG estimation that allowed for a small degree of head movement was developed based on the relative displacements of the center of the pupil and an infrared (IR) reflection formed off the surface of the corneal, known as the Pupil-Corneal Reflection (P-CR) method [17]. The corneal reflection, generated by external lighting, provides a reference point for determining the relative motion of the pupil. A simple first or second order polynomial mapping is used to relate the 2D image vector formed from the center of the corneal reflection to the center of the pupil to the 2D POG screen vector. After calibration, average accuracies for this method are typically 0.5 to 1° of visual angle [29]. Infrared light is used to generate the corneal reflections as IR is outside of the visible

spectrum and avoids disturbing the system user. Additionally, using system controlled IR light avoids the potential problems encountered with variable ambient lighting conditions.

The simplicity of the P-CR vector method and its ability to handle minor head motions led to its widespread adoption within the eye-gaze tracking community. Unfortunately the accuracy of the P-CR method decreases considerably as the head is displaced from the calibration position [29] [39]. The degradation in accuracy has led further research into techniques for POG estimation that allow for free head motion [40] [41] [42].

Model-Based POG estimation

Algorithms based on 3D models have been developed to overcome the decrease in accuracy that the P-CR method exhibits with larger head movements. The model-based methods use models of the camera, eye and system to compute the position of the eye in 3D space, the position of the center of the pupil and consequently the optical axis (vector between these two points) of the eye. Population averages are typically used for the parameters of the eye model, while calibration is required to compensate for the offset between the optical axis and visual axis. The optical and visual axis offset is due to the position of the fovea on the retina which varies between different subjects. The intersection of the visual axis with an object upon which the user is looking in real-space then results in the POG. This object is typically the planar surface of the computer screen. The model-based method determines the location of the eyes in 3D space and therefore is able to estimate the POG regardless of the position of the head. The model-based method will be described in greater detail in the following chapters. A summary of a few model-based contemporary systems is described here.

Shih and Liu [42] developed a novel model-based method for estimating eye-gaze. The system they designed uses two RS-170 based cameras and frame grabbers to record images with a resolution of 640 x 240 pixels at a frame rate of 30 Hz. Average accuracy was shown to be better than 1° of visual angle. Unfortunately their system design required the cameras to be quite close to the subjects' eyes in order to acquire high spatial resolution images, restricting the freedom of head motion due to the limited field of view of the camera.

To overcome the limitation of a narrow field of view, Ohno and Mukawa [43] developed a model-based system with a camera mounted on a pan / tilt mechanism with a narrow angle (NA) lens, and two fixed cameras with wide angle (WA) lenses. The fixed cameras used stereo imaging to determine the

location of the head within the scene and directed the pan / tilt mechanism to orient the NA camera towards the eye. The WA cameras recorded images with a resolution of 320 x 120 pixels while the NA camera recorded images with a resolution of 640 x 480 pixels, all at frame rates of 30 Hz. System accuracy was reported to be better than 1.0° of visual angle. The pan / tilt mechanism allowed the NA camera to track the motion of the eye with a larger effective field of view. However, the speed at which the mechanism could move was not sufficient to keep up with the faster motion of the head and eye, resulting in loss of tracking and slow re-acquisition.

Beymer and Flickner [41] used high speed galvonometers for their model-based system in an attempt to overcome the limitations of the slow pan / tilt systems. A pair of fixed WA cameras used stereo imaging to direct the orientation of two NA cameras by controlling the pan and tilt of rotating lightweight mirrors mounted on galvonometers. The focus of each camera was controlled with a lens mounted on a bellows and driven by another motor. The NA cameras recorded NTSC images (with a typical resolution of 640 x 480 pixels) at a frame rate of 30 Hz. Due to the significant processing involved in the multiple video-stream system, a POG sampling rate of only 10 Hz was achieved. The accuracy reported for this system was 0.6° of visual angle for a single subject tested. While their system was capable of tracking the eye in the presence of natural high speed head motion, considerable calibration was required, and the overall complexity of the system may have contributed to the low POG sampling rate.

1.3.3 3D POG estimation

The remote eye-gaze tracking systems based on mechanical tracking of the eye typically only track a single eye to reduce the complexity of the tracking mechanism. As both eyes generally point to the same position, tracking a single eye is sufficient for 2D POG estimation [20]. With binocular eye-gaze tracking however, it becomes possible to track the position and orientation of both eyes, and therefore determine the 3D POG based on the vergence angle between the left and right eyes.

While a remote 3D POG estimation system had not previously been developed, two researchers have investigated 3D POG estimation using binocular head mounted eye-gaze tracking systems. With head mounted systems, two cameras can be mounted on the head, one for each eye. The system by Duchowski *et al* [44] used a commercial binocular head mounted eye-tracker (ISCAN RK-726), combined with binocular head mounted displays (HMD) for their 3D virtual reality display. The left and right POG were individually

estimated in 2D on each of the HMD screens using the P-CR POG estimation method. A magnetic position tracker (Flock of Birds by Ascension Technologies), also worn by the subject, was used to determine the position and orientation of the head. The head pose information was combined with the disparity found between the left and right eyes to develop a geometric method for estimating the POG in virtual 3D space. Two stages of user calibration were required, one for the commercial head mounted eye-tracker and one for the geometric method for POG estimation.

Another head mounted system was developed by Essig *et al* [45], which also used a commercial head mounted eye-gaze tracker. The virtual 3D environment was created using anaglyph images displayed on a 20" desktop display. The anaglyph images were formed by rendering one image composed of a red scene for the left eye and a second image composed of a blue scene for the right eye. To separate the images a pair of eye-glasses with red and blue filters were worn by the user. The P-CR method was used for 2D POG estimations for both the left and right eyes on the surface of the desktop monitor. Rather than use the geometric method for 3D POG estimation, the authors used the 2D POG estimates tracked on the remote 20" desktop monitor as input to a neural network which then estimated the 3D POG. An integrated head tracker provided some degree of head motion compensation. Two stages of user calibration were required, the first to calibrate the eye-gaze tracker on the desktop display and the second to train the neural network.

In both systems the virtual 3D environment was presented to the user on 2D displays. The head mounted eye-gaze trackers were used to generate the 2D POG estimates which were then used as inputs to a geometric algorithm [44] or a neural network algorithm [45] for computation of the 3D POG. Both systems required multiple stages of calibration, both for the 2D eye-gaze trackers as well as for calibration or training of the 3D POG estimation methods. Both systems also used 2D displays (two HMD screens in [44] and a 2D monitor in [45]) to create a stereo presentation of a virtual scene in which the virtual 3D POG was estimated.

1.4 3D Display and User Interface Technologies

The two 3D POG estimation systems described above utilize stereoscopic virtual 3D displays. Stereoscopic displays present different images to the left and right eyes with slightly different perspectives creating the illusion of depth [33]. These displays require the user to accommodate (or focus)

their eyes at a fixed distance while changes in the vergence of the eyes create the feeling of depth. If there is any discrepancy between the visual cues for vergence and accommodation the user may feel nausea, dizziness or headaches as a result [46]. The stereoscopic 3D displays also require contact with the subject, either through head mounted displays or colour filter glasses worn by the user.

A number of techniques for creating autostereoscopic displays in which 3D information is presented without the user having to wear any equipment are currently under development [47]. For a literature survey of the state of the art in 3D displays see Favalor [23], Dodgson [24] or Benzie [48]. Volumetric 3D displays show considerable promise in that they present virtual objects in a true 3D volume, and are therefore viewable from any angle by any number of users simultaneously [49]. The problems associated with vergence and accommodation discrepancies in stereoscopic displays are avoided with volumetric displays [50].

As 3D displays become more prevalent, the demand for user-friendly, interactive tools that operating in 3D environments will grow [51]. The current 2D mouse is insufficient for naturally interacting in a 3D environment as it lacks sufficient degrees of freedom [52]. The most common methods for interaction in 3D are optical or electro-magnetic based 3D position tracking [26]. Optical methods use stereo imaging to track reflective markers affixed to a stylus such as the OptoTrak system by Northern Digital Inc [53]. Electro-magnetic based trackers use pulsed magnetics fields in orthogonal transmitter coils and matching coils located in a remote sensor block such as the Flock of Birds system by Ascension Technologies [54]. With both the optical and electro-magnetic based trackers the tracked stylus must be held against gravity, eventually resulting in user fatigue with extended use [27]. There is also a visual disconnect when the tracked tool cannot be physically located within the environment in which it is supposed to be acting [28].

1.5 Chapter Summary

The unifying theme of the research presented here is the goal of developing methods for real-world 3D POG estimation using remote, non-contact eye-gaze tracking. The thesis presented here is written in manuscript style, as permitted by the Faculty of Graduate Studies at the University of British Columbia. In the manuscript style thesis, each chapter represents an individual research effort, culminating in a peer reviewed submission or publication. Each chapter can be read individually with an overview of the motivation of

the research and a review of relevant literature presented for each chapter. The references are summarized in the bibliography found at the end of each chapter as per the requirements of the Faculty of Graduate Studies.

In Chapter 2 a model-based method for monocular POG estimation is presented [55]. The use of image-based tracking provides a means of following the motion of the head without mechanical tracking. The simplified eye model used allows for tracking the position of the center of the cornea, modeled as a sphere, and the center of the pupil in 3D real-world space. With the center of the cornea and pupil, along with calibration, the visual axis along which the eye is looking can be determined. The intersection of the visual axis with the 3D model of the screen results in the 2D POG. The techniques developed operate with a single fixed remote camera, require no contact with the user and allow for free head motion.

Given that the accuracy and head motion compensation performance of the model-based method developed for monocular POG estimation was shown to match that of leading contemporary systems, a set of techniques are then presented in Chapter 3 for improving the precision of the POG estimates during fixations [56]. The image processing algorithms were refined for high speed operation using a combination of software and hardware regions-of-interest for reducing the quantity of image information to process. Fixation detection provides for fast response times while high speed filtering is shown to considerably reduce the effects of the naturally jittery motions of the eyes.

In Chapter 4 the monocular POG estimation methods are extended to high speed binocular tracking [57]. A simple face tracking technique is presented for differentiating the left and right eyes when one eye is lost due to head motion, enlarging the effective field of view of the system. A multiple corneal reflection pattern tracking algorithm is presented for compensating for head and eye motions which result in the loss or distortion of corneal reflections. The face tracking and multiple corneal reflection pattern matching algorithms are designed to operate at high speed to maintain rapid POG estimation.

The model-based method for high speed binocular POG estimation is then extended to estimation of the 3D POG in Chapter 5 [58]. An intersection method is presented for determining the closest point of approach of the binocular left and right eye visual axis vectors. The vergence intersection magnifies the natural jittery motion of the eyes, the effects of which are reduced using low pass filtering on the high speed 3D POG estimates. An evaluation of the accuracy and precision throughout the workspace volume of the 3D POG estimation techniques is also presented.

In Chapter 6 the results of the collected works are related to one another in the context of the overall thesis goal of 3D POG estimation. An illustrative application will be presented in which the 3D POG is used as an interface tool with a simple 3D volumetric display [59]. The strengths and weaknesses of the research is then presented, along with future directions for research.

References

[1] E. Kowler, *Eye Movements and their Role in Visual and Cognitive Processes.* Elsevier Science, 1990, vol. 4, ch. The role of visual and cognitive processes in the control of eye movement., pp. 1–70.

[2] R. Jacob and K. Karn, *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research.* Amsterdam: Elsevier Science, 2003, ch. Eye Tracking in Human-Computer Interaction and Usability Research: Ready to Deliver the Promises (Section Commentary), pp. 573–605.

[3] D. M. Stampe and E. M. Reingold, *Eye Movement Research: Mechanisms, Processes and Applications.* Elsevier Science, 1995, ch. Selection by looking: A novel computer interface and its application to psychological research, pp. 467–478.

[4] L. E. Sibert and R. J. K. Jacob, "Evaluation of eye gaze interaction," in *Proceedings of the SIGCHI conference on Human factors in computing systems.* New York, NY, USA: ACM Press, 2000, pp. 281–288.

[5] J. P. Hansen, K. Tørning, A. S. Johansen, K. Itoh, and H. Aoki, "Gaze typing compared with input by head and hand," in *Proceedings of the 2004 symposium on Eye tracking research & applications.* New York, NY, USA: ACM Press, 2004, pp. 131–138.

[6] P. Pellegrino, D. Bonino, and F. Corno, "Domotic house gateway," in *Proceedings of the 2006 ACM symposium on Applied computing.* New York, NY, USA: ACM, 2006, pp. 1915–1920.

[7] H. Istance, "Communication through eye-gaze: where we have been, where we are now and where we can go from here," in *Proceedings of the 2006 symposium on Eye tracking research & applications.* New York, NY, USA: ACM, 2006, pp. 9–9.

[8] F. Shi, A. Gale, and K. Purdy, "Helping people with ict device control by eye gaze," in *Lecture Notes in Computer Science.* Springer Berlin / Heidelberg, 2006, vol. 4061, pp. 480–487.

[9] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[10] K. Rayner, "Eye movements in reading and information processing: 20 years of research." *Psychol Bull*, vol. 124, no. 3, pp. 372–422, Nov 1998.

[11] Y. Matsumoto and A. Zelinsky, "An algorithm for real-time stereo vision implementation of head pose and gaze direction measurement," in *Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, 28-30 March 2000, pp. 499–504.

[12] G. L. Lohse, "Consumer eye movement patterns on yellow pages advertising," *Journal of Advertising*, vol. 26, no. 1, pp. 61–73, 1997.

[13] R. Radach, S. Lemmer, C. Vorstius, D. Heller, and K. Radach, *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research.* Amsterdam: Elsevier Science, 2003, ch. Eye movements in the processing of print advertisements, p. 609632.

[14] D. Beymer and D. M. Russell, "Webgazeanalyzer: a system for capturing and analyzing web reading behavior using eye gaze," in *CHI '05 extended abstracts on Human factors in computing systems.* New York, NY, USA: ACM Press, 2005, pp. 1913–1916.

[15] L. C. Loschky and G. W. McConkie, "User performance with gaze contingent multiresolutional displays," in *Proceedings of the 2000 symposium on Eye tracking research & applications.* New York, NY, USA: ACM Press, 2000, pp. 97–103.

[16] S. Zhai, C. Morimoto, and S. Ihde, "Manual and gaze input cascaded (magic) pointing," in *CHI '99: Proceedings of the SIGCHI conference on Human factors in computing systems.* New York, NY, USA: ACM Press, 1999, pp. 246–253.

[17] T. Hutchinson, J. White, W. Martin, K. Reichert, and L. Frey, "Human-computer interaction using eye-gaze input," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 19, no. 6, pp. 1527–1534, 1989.

[18] L. Frey, K. White, and T. Hutchison, "Eye-gaze word processing," *IEEE Transactions on Systems, Man and Cybernetics, Part B*, vol. 20, no. 4, pp. 944–950, July-Aug. 1990.

[19] D. J. Ward and D. J. C. MacKay, "Fast hands-free writing by gaze direction," *Nature*, vol. 418, no. 6900, p. 838, 2002.

[20] R. J. K. Jacob, *Eye Movement-Based Human-Computer Interaction Techniques: Toward Non-Command Interfaces.* Norwood, N.J.: Ablex Publishing Co., 1993, vol. 4, pp. 151–190.

[21] A. T. Duchowski, *Eye Tracking Methodology: Theory and Practice.* Springer-Verlag, 2003.

[22] R. Bates, M. Donegan, H. O. Istance, J. P. Hansen, and K.-J. Raiha, "Introducing cogain: communication by gaze interaction," *Univers. Access Inf. Soc.*, vol. 6, no. 2, pp. 159–166, 2007.

[23] G. E. Favalora, "Volumetric 3d displays and application infrastructure," *Computer*, vol. 38, no. 8, pp. 37–44, Aug. 2005.

[24] N. Dodgson, "Autostereoscopic 3d displays," *Computer*, vol. 38, no. 8, pp. 31 – 36, Aug. 2005.

[25] M. Chen, S. J. Mountford, and A. Sellen, "A study in interactive 3-d rotation using 2-d control devices," *SIGGRAPH Comput. Graph.*, vol. 22, no. 4, pp. 121–129, 1988.

[26] K. Meyer, H. L. Applewhite, and F. A. Biocca, "A survey of position trackers," *Presence: Teleoper. Virtual Environ.*, vol. 1, no. 2, pp. 173–200, 1992.

[27] S. Zhai, "User performance in relation to 3d input device design," *SIGGRAPH Comput. Graph.*, vol. 32, no. 4, pp. 50–54, 1998.

[28] C. Ware, "Using hand position for virtual object placement," *Vis. Comput.*, vol. 6, no. 5, pp. 245–253, 1990.

[29] C. H. Morimoto and M. R. M. Mimica, "Eye gaze tracking techniques for interactive applications," *Comput. Vis. Image Underst.*, vol. 98, no. 1, pp. 4–24, 2005.

[30] R. Jacob, *Virtual Environments and Advanced Interface Design.* New York, NY, USA: Oxford University Press, 1995, ch. Eye tracking in advanced interface design, pp. 258–288.

[31] A. Spauschus, J. Marsden, D. Halliday, J. Rosenberg, and P. Brown, "The origin of ocular microtremor in man," *Experimental Brain Research*, vol. 126, no. 4, pp. 556–562, June 1999.

[32] U. Tulunay-Keesey, "Fading of stabilized retinal images." *J Opt Soc Am*, vol. 72, no. 4, pp. 440–447, Apr 1982.

[33] Z. Wartell, L. F. Hodges, and W. Ribarsky, "Balancing fusion, image depth and distortion in stereoscopic head-tracked displays," in *Proceedings of the 26th annual conference on Computer graphics and interactive techniques.* New York, NY, USA: ACM Press/Addison-Wesley Publishing Co., 1999, pp. 351–358.

[34] D. A. Goss and R. W. West, *Introduction to the Optics of the Eye.* Butterworth Heinemann, 2001.

[35] E. Javal, "Essai sur la physiologie de la lecture," *Annales d'Oculistique*, vol. 79, pp. 97–117, 155–167, 240–274, 1878.

[36] Dodge and Cline, "The angle velocity of eye movements," *Psychological Review*, vol. 8, pp. 145–157, 1901.

[37] C. Judd, C. McAllister, , and W. Steel, "General introduction to a series of studies of eye movements by means of kinetoscopic photographs," *Psychological Review, Monograph Supplements*, vol. 7, pp. 1–16, 1905.

[38] L. Young and D. Sheena, "Methods & designs: survey of eye movement recording methods," *Behav. Res. Methods Instrum.*, vol. 5, pp. 397–429, 1975.

[39] J. J. Cerrolaza, A. Villanueva, and R. Cabeza, "Taxonomic study of polynomial regressions applied to the calibration of video-oculographic systems," in *Proceedings of the 2008 symposium on Eye tracking research & applications.* New York, NY, USA: ACM, 2008, pp. 259–266.

[40] T. Ohno, N. Mukawa, and A. Yoshikawa, "Freegaze: a gaze tracking system for everyday gaze interaction," in *Proceedings of the 2002 symposium on Eye tracking research & applications.* New York, NY, USA: ACM Press, 2002, pp. 125–132.

[41] D. Beymer and M. Flickner, "Eye gaze tracking using an active stereo head," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, 18-20 June 2003, pp. II–451–II–458.

[42] S.-W. Shih and J. Liu, "A novel approach to 3-d gaze tracking using stereo cameras," *IEEE Transactions on Systems, Man and Cybernetics, Part B*, vol. 34, no. 1, pp. 234–245, Feb. 2004.

[43] T. Ohno and N. Mukawa, "A free-head, simple calibration, gaze tracking system that enables gaze-based interaction," in *Proceedings of the*

2004 symposium on Eye tracking research & applications. New York, NY, USA: ACM Press, 2004, pp. 115–122.

[44] A. T. Duchowski, V. Shivashankaraiah, T. Rawls, A. K. Gramopadhye, B. J. Melloy, and B. Kanki, "Binocular eye tracking in virtual reality for inspection training," in *Proceedings of the 2000 symposium on Eye tracking research & applications.* New York, NY, USA: ACM Press, 2000, pp. 89–96.

[45] K. Essig, M. Pomplun, and H. Ritter, "A neural network for 3d gaze recording with binocular eyetrackers," *International Journal of Parallel, Emergent and Distributed Systems*, vol. 21, no. 2, pp. 79–95, April 2006.

[46] O. Bimber and R. Raskar, "Modern approaches to augmented reality," in *ACM SIGGRAPH 2006 Courses.* New York, NY, USA: ACM, 2006, p. 1.

[47] M. Halle, "Autostereoscopic displays and computer graphics," *SIGGRAPH Comput. Graph.*, vol. 31, no. 2, pp. 58–62, 1997.

[48] P. Benzie, J. Watson, P. Surman, I. Rakkolainen, K. Hopf, H. Urey, V. Sainov, and C. von Kopylow, "A survey of 3dtv displays: Techniques and technologies," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 11, pp. 1647–1658, Nov. 2007.

[49] A. Jones, I. McDowall, H. Yamada, M. Bolas, and P. Debevec, "Rendering for an interactive 360° light field display," in *ACM SIGGRAPH.* New York, NY, USA: ACM, 2007, p. 40.

[50] T. Grossman and R. Balakrishnan, "The design and evaluation of selection techniques for 3d volumetric displays," in *Proceedings of the 19th annual ACM symposium on User interface software and technology.* New York, NY, USA: ACM Press, 2006, pp. 3–12.

[51] D. A. Bowman, "Interaction techniques for common tasks in immersive virtual environments: Design, evaluation, and application," Ph.D. dissertation, Georgia Institute of Technology, 1999.

[52] K. Hinckley, R. Pausch, J. C. Goble, and N. F. Kassell, "A survey of design issues in spatial input," in *Proceedings of the 7th annual ACM symposium on User interface software and technology.* New York, NY, USA: ACM Press, 1994, pp. 213–222.

[53] Optotrak, Northern Digital Inc., Waterloo Ontario, Canada 2008.

[54] Flock of Birds, Ascension Technology Corporation, Burlington Virgina, USA 2008.

[55] C. Hennessey, B. Noureddin, and P. Lawrence, "A single camera eye-gaze tracking system with free head motion," in *Proceedings of the 2006 symposium on Eye tracking research & applications.* New York, NY, USA: ACM Press, 2006, pp. 87–94.

[56] ——, "Fixation precision in high-speed noncontact eye-gaze tracking," *IEEE Transactions on Systems, Man and Cybernetics, Part B*, vol. 38, no. 2, pp. 289–298, April 2008.

[57] C. Hennessey and P. Lawrence, "Improving the accuracy and reliability of remote system-calibration-free eye-gaze tracking," *IEEE Transactions on Biomedical Engineering*, in submission.

[58] ——, "Non-contact binocular eye-gaze tracking for point-of-gaze estimation in three dimensions," *IEEE Transactions on Biomedical Engineering*, in submission.

[59] ——, "3d point-of-gaze estimation on a volumetric display," in *Proceedings of the 2008 symposium on Eye tracking research & applications.* New York, NY, USA: ACM, 2008, pp. 59–59.

Chapter 2

A Single Camera Eye-Gaze Tracking System with Free Head Motion [1]

2.1 Introduction

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

Eye-gaze information has proven useful in a diverse number of applications such as psychological studies [60], usability studies in driving and aviation [61; 62], and analysis of layout effectiveness in advertising [63]. In particular it is well suited to human computer interfaces for mouse augmentation and control [64] and eye typing for the physically disabled [65].

Recent advances in electronics and computing technology have made possible non-contact and real-time video based eye-gaze tracking systems. These systems are replacing the traditional methods used for eye-gaze tracking in many applications due to their increased ease of use, reliability, accuracy and comfort for the subject. To be acceptable to the general population, eye-gaze tracking systems should be non-contact, non-restrictive, sufficiently accurate for the user's range of tasks, easy to set up and simple to use.

The system described in this chapter meets these key requirements as follows. A single high resolution camera with a fixed field of view is used which does not make any contact with the user. A model-based method

[1] A version of this chapter has been published. Hennessey, C., Noureddin, B., and Lawrence, P. 2006. A single camera eye-gaze tracking system with free head motion. In Proceedings of the 2006 Symposium on Eye Tracking Research & Applications (San Diego, California, March 27 - 29, 2006), 87-94.

based on multiple reflections off the surface of the cornea (also known as glints) is used to allow free head motion within the field of view of the camera. The high resolution images permit a larger field of view while still possessing accurate image features, resulting in accurate eye gaze estimation. Using a single camera with no moving parts simplifies the system geometry and calibration and leads to short reacquisition times. These advantages make the system easy to set up and simple to use.

The motivation for this chapter is to present a preliminary evaluation of the system and how the design choices affect overall eye-gaze system accuracy. In particular the effect of processing power, camera resolution and frame rate on eye-gaze accuracy for free head movement are assessed. To the best of our knowledge this is the first reported implementation of a single camera, multiple glint, eye-gaze tracking system that permits free head motion.

2.2 Related Works

There have been many methods developed for tracking the eye-gaze of a subject including sensors attached to the face and eye, restrictive video systems requiring a fixed head location, head mounted video systems, and non-contact and non-restrictive video based systems. We feel that the non-contact, non-restrictive video based methods hold the greatest promise for a widely acceptable eye-gaze tracking interface and as such will focus on research in this area. For an overview of alternative methods for eye-gaze tracking see the review by Young and Sheena [66], and more recently Morimoto and Mimica [67].

Video based systems require high resolution images of the eye to accurately estimate the point of gaze (POG). Ohno *et al.* developed a single camera system which achieved accuracies of under 1° of visual angle [68]. The system verified the ability of their methods to determine the POG, however it had a relatively small field of view of 4 x 4 cm at 60 cm distance. Morimoto *et al.* proposed a single camera method for estimating the POG which achieved an average accuracy of 2.5° in simulations [69]. To date there is no reported system implementation based on this proposal. Shih and Liu proposed a method which used only a single camera [70]. The system they implemented utilized two stereo cameras however, which were required to provide additional constraints for their algorithms. The fixed field of view was restricted to 4 x 4 cm. Their system operated at 30 Hz with an accuracy of approximately 1° of visual angle.

The main difficulty with the above fixed single camera systems is the limited field of view required to capture sufficiently high resolution images. To allow for free head motion a large field of view is required. Many systems utilize multiple cameras to achieve these goals, with wide angle (WA) lens cameras used to direct a movable narrow angle (NA) lens camera. Yoo and Chung developed a free head system which utilizes a WA camera to direct a NA camera mounted on a pan-tilt mechanism [71]. Their system operates at 15 Hz and achieves an accuracy of 0.98° of visual angle in the horizontal direction and 0.82° in the vertical direction. Noureddin *et al.* developed a two camera system where the fixed WA camera uses a rotating mirror to direct the orientation of the NA camera [72]. The rotating mirror can achieve faster slew rates when compared with pan-tilt mechanisms. Their system operates at 9 Hz with an accuracy of 2.9°. The latest reported system by Ohno and Mukawa utilizes 3 cameras, two fixed stereo WA cameras and a NA camera mounted on a pan-tilt mechanism [73]. Their system uses two computers and achieves an accuracy of about 1° of visual angle while operating at 30 Hz. Beymer and Flickner developed a 4 camera system which uses 2 stereo WA cameras and 2 stereo NA cameras [74]. The WA cameras direct galvanometer motors to orient the NA cameras. The calibration task is considerable due to the multiple stereo cameras and the variable focal lengths of the NA cameras. Their system operates at 10 Hz and has a reported accuracy of 0.6°.

Whenever the eye moves outside the NA field of view, these multi camera systems mechanically reorient the NA camera towards the new eye position. The time required to reacquire the eye in this way can be long, resulting in high reacquisition times when the head moves. Considerable system calibration is required for larger numbers of cameras, as well as increased processing power for the increased number of video streams.

The system we have developed requires only a single camera and has no moving parts, resulting in short reacquisition times, while maintaining comparably accurate POG estimation and a larger field of view than other single camera systems. Other differences include the use of ray tracing rather than depth from focus, the method for dealing with refraction at the surface of the eye, the calibration method, the pupil image contour refinement techniques and an implementation to validate the design. The Tobii system developed by Tobii Technologies is a proprietary single camera, multiple glint system that may have similarities to ours, however, no information in the open literature is available on its complete design, implementation, or testing methodologies.

2.3 Methods

The methods we have developed for estimating the POG are based on 3D models of the camera, system and eye. The camera is modeled using the pinhole camera model and the eye is modeled using a simplified version of the Gullstrand schematic eye [75]. Population averages compiled by Gullstrand are used for the model parameters of interest. Subject deviations from the population averages are compensated by a one-time per user calibration.

Shown in Figure 2.1 is an example of the simplified eye model with the parameters of interest, r, r_d, and n, and the points P_c and C, which are required to compute the optical axis vector $\boldsymbol{L}$. The optical axis is defined as the vector from the center of the cornea C to the center of the pupil P_c. The optical axis is different from the visual axis which is the vector that traces from the fovea (high acuity portion of the retina) through the center of the pupil and ultimately to the real POG. The location of the fovea varies from person to person, and can be located up to 5° from the optical axis [75]. The offset between the estimated POG and the real POG due to the difference between the optical axis and the visual axis is fixed for each user and is compensated for by the calibration technique described in Section 2.3.4.

The following outline provides an overview of the steps required to determine the POG using the intersection of the optical axis vector $\boldsymbol{L}$ with the monitor plane (where $\boldsymbol{L} = P_c - C$) to determine the POG P:

1. To determine the cornea center C, the eye model is used along with the image locations of two glints off the surface of the cornea. Using multiple glints provides a method for triangulating the 3D cornea center.

2. The pupil center P_c is determined by using the eye model, the cornea center C and the perimeter points of the pupil image.

3. The estimated POG is corrected for possible errors by a one-time per user calibration.

4. The image locations of the glints and pupil contour used in steps 2 and 3 above are extracted from images of the eye using image processing techniques.

The following sections describe each of these steps in more detail.

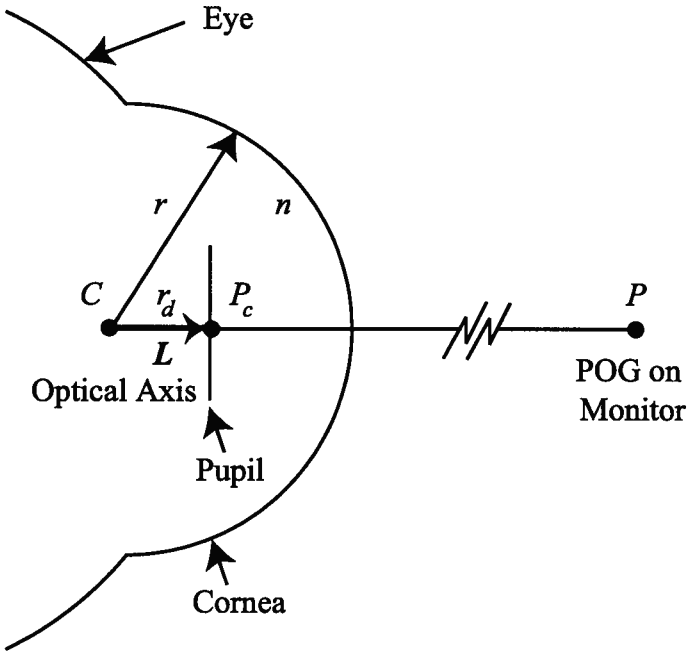


Figure 2.1: Eye model used to calculate the POG. The parameters of interest taken from population averages are: radius of the cornea, r, distance from the center of the cornea to the center of the pupil r_d and the index of refraction of the aqueous humor, n. The center of the cornea is located at point C and the center of the pupil is located at point P_c. The optical axis $\boldsymbol{L}$ is the vector formed from C to P_c, and the POG P is the intersection of the optical axis with the monitor plane.

2.3.1 POG Estimation

The POG P is the intersection of the optical axis vector $\boldsymbol{L}$ with the surface of the computer monitor. The monitor surface is modeled with a plane equation given the measured locations of three of the screen corners. The 3D parametric equation of a line defined by (2.1) is used to determine the POG.

$$P = C + t \cdot \boldsymbol{L} \tag{2.1}$$

This 3D vector equation has 4 unknowns $P = (p_x, p_y, p_z)$ and t. Adding the constraint that the POG must lie on the plane defined by the monitor provides the additional constraint required to solve for the POG explicitly, assuming that C and P_c are known.

The methods for determining the location of the cornea center C and the pupil center P_c required to compute the optical axis are given in the following sections.

2.3.2 Cornea Center Estimation

We have implemented an extension of Shih and Liu's proposed single camera method for estimating the location of the cornea center in 3D space [70].

A ray can be traced from each glint light source Q_i through the points G_i, O, and I_i as shown in Figure 2.2, where i is the index of the two or more point light sources generating the rays.

Shih and Liu noted that the set of points (Q_i, G_i, C, O, I_i) are co-planar. An auxiliary coordinate system can be defined for each glint light source such that all these points lie in a plane defined by two axes of the coordinate system, thus reducing the solution space from three degrees of freedom to two. A rotation matrix R_i and its inverse can be formulated for each glint to transform points between the auxiliary coordinate systems and the world coordinate system.

Using the geometry illustrated in Figure 2.3 it is possible to define the center of the cornea $\hat{C}_i$ in the auxiliary coordinate system as a function of a single unknown parameter $\hat{g}_{ix}$ for each glint as follows:

$$\hat{C}_i = \begin{bmatrix} \hat{c}_{ix} \\ \hat{c}_{iy} \\ \hat{c}_{iz} \end{bmatrix} = \begin{bmatrix} \hat{g}_{ix} - r \cdot \sin\left(\frac{\hat{\alpha}_i - \hat{\beta}_i}{2}\right) \\ 0 \\ \hat{g}_{ix} \cdot \tan\left(\hat{\alpha}_i\right) + r \cdot \cos\left(\frac{\hat{\alpha}_i - \hat{\beta}_i}{2}\right) \end{bmatrix} \tag{2.2}$$

where

$$\hat{\alpha}_i = \cos^{-1}\left(\frac{-\hat{I}_i \cdot \hat{Q}_i}{\left\|-\hat{I}_i\right\| \cdot \left\|\hat{Q}_i\right\|}\right) \tag{2.3}$$

$$\hat{\beta}_i = \tan^{-1}\left(\frac{\hat{g}_{ix} \cdot \tan(\hat{\alpha}_i)}{\hat{l}_i - \hat{g}_{ix}}\right) \tag{2.4}$$

When the auxiliary cornea center $\hat{C}_i$ is transformed back to the world coordinate system using

$$C_i = R_i^{-1}\hat{C}_i \tag{2.5}$$

the result is a set of 3 equations with 4 unknowns (c_{ix}, c_{iy}, c_{iz}, $\hat{g}_{ix}$). Using two glints provides a total of 6 equations with 8 unknowns. The constraint that the cornea center defined for each glint must be coincident in the world coordinate system results in another set of 3 equations as follows

$$C_1 = C_2 \tag{2.6}$$

The over defined set of equations then consist of 9 equations with 8 unknowns which are solved numerically for C using a gradient descent algorithm.

2.3.3 Pupil Center Estimation

The second point required for the optical axis is the center of the pupil P_c. The optical axis requires the center of the real pupil and not its refracted image recorded by the camera. The center of the real pupil can be found by computing the average of at least two opposing points on the real pupil perimeter, although in practice we found using six perimeter points provided a more robust estimate.

To determine a real pupil perimeter point, a ray defined by a 3D parametric equation of a line

$$U_i = K_i + s_i \cdot \boldsymbol{K}_i \tag{2.7}$$

is traced from the pupil perimeter point K_i on the surface of the camera sensor to the surface of the cornea through the focal point of the pin-hole camera, as illustrated in Figure 2.4.

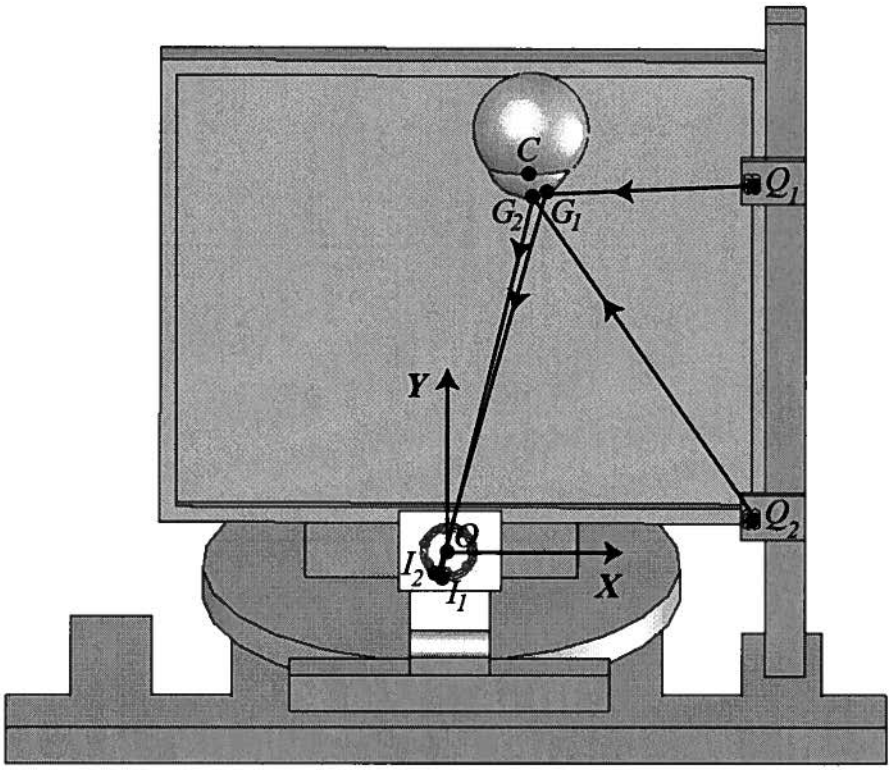


Figure 2.2: Rays traced from multiple glint light sources to the surface of the camera sensor. The glint light source is located at point Q_i. The glints on the surface of the spherical cornea (center C, radius r), are located at point G_i. The focal point of the pin-hole camera model is located at point O and the image of the glint on the surface of the CCD sensor is located at point I_i. The index i is 1 for points along rays from glint source 1 and 2 for points along rays from glint source 2.

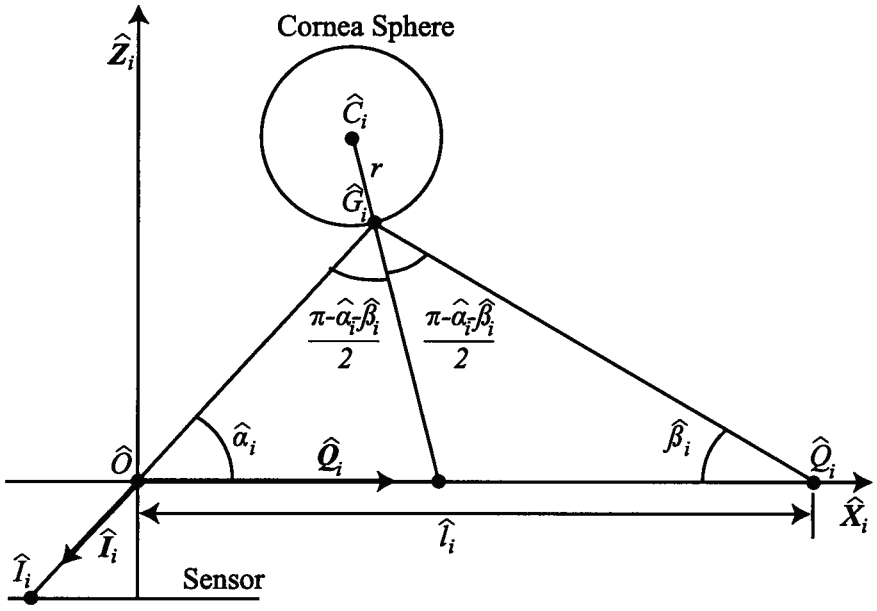


Figure 2.3: Auxiliary coordinate system geometry. Each auxiliary coordinate system is defined with the origin at point $\hat{O}$, where $\hat{O} = O$. The $\hat{X}_i$-axis is defined along $\hat{Q}_i$ and the $\hat{Z}_i$-axis such that the vector from $\hat{I}_i$ lies in the $\hat{X}_i$-$\hat{Z}_i$ plane. Finally the $\hat{Y}_i$-axis is defined orthonormal to the $\hat{X}_i$ and $\hat{Z}_i$ axes. The vectors $\hat{I}_i$ and $\hat{Q}_i$ are the vectors from points $\hat{O}$ to $\hat{I}_i$ and from $\hat{O}$ to $\hat{Q}_i$ respectively. The scalar $\hat{l}_i$ is the distance from points $\hat{O}$ to $\hat{Q}_i$.

Adding the constraint that the point U_i must lie on the surface of the spherical cornea with center C and radius r

$$(u_{ix} - c_x)^2 + (u_{iy} - c_y)^2 + (u_{iz} - c_z)^2 = r \tag{2.8}$$

provides a set of 4 equations with 4 unknowns which can then be solved explicitly for U_i. The vector $\boldsymbol{K_i}$ is then refracted into the eye using Snell's law of refraction, the indices of refraction of both air and the aqueous humor, and an equivalent angle rotation. The refracted vector $\hat{\boldsymbol{K}}_i$ is then traced to the real pupil perimeter point using another parametric equation of a line

$$\hat{U}_i = U_i + w_i \cdot \hat{\boldsymbol{K}}_i \tag{2.9}$$

Again we have 3 equations with 4 unknowns ($\hat{u}_{ix}$, $\hat{u}_{iy}$, $\hat{u}_{iz}$, w_i) which can be solved explicitly by adding a constraint on the distance between the pupil perimeter point and the cornea center:

$$\left\| \hat{U}_i - C \right\| = r_{ps} \tag{2.10}$$

where r_{ps} is defined as

$$r_{ps} = \sqrt{r_d^2 + r_p^2} \tag{2.11}$$

r_d is given by the population averages by Gullstrand and r_p is estimated by using the pinhole camera model and the major axis of the pupil image contour ellipse equation.

The pupil center P_c is computed by averaging the pupil perimeter values $\hat{U}_i$. The optical axis can thus be computed with the estimated pupil and cornea centers, and ultimately used to estimate the POG as per (2.1).

2.3.4 Calibration Method

There are a number of simplifications employed in the models above which may result in POG inaccuracies. Such simplifications include the pin-hole camera model used to approximate the real camera and lens, the simplified eye model and the use of population averages for the parameters of the eye. A one-time calibration is performed on a per-user basis to correct for all of the possible sources of errors. The calibration procedure is automated, in that the system detects when to switch to the next calibration point, and can be performed in under five seconds. Figure 2.5 illustrates an example of the parameters used in performing the calibration and correction of the

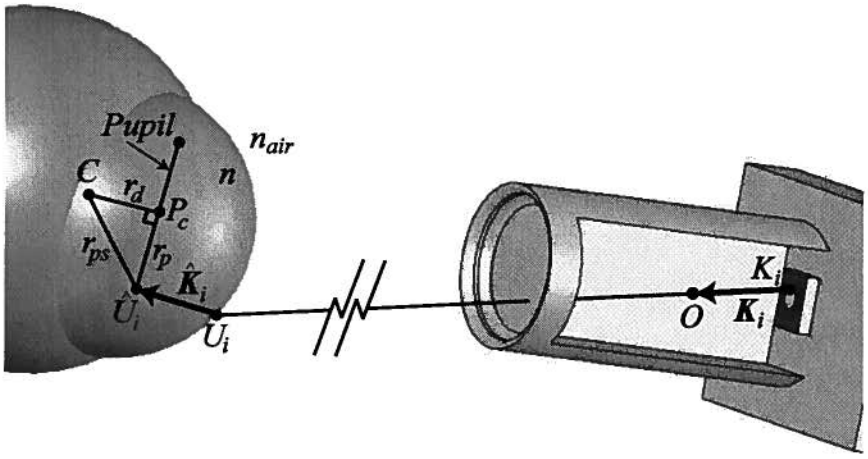


Figure 2.4: Estimating the pupil center through ray tracing. The pupil perimeter image point on the surface of the camera sensor is denoted by K_i. The ray $\boldsymbol{K_i}$ is traced from the camera sensor to a point U_i on the surface of the cornea. The refracted vector $\hat{\boldsymbol{K}}_i$ points from U_i to the real pupil perimeter point $\hat{U}_i$. The distance from the center of the cornea to the center of the pupil is given by r_d and to the perimeter of the pupil by r_{ps}. The radius of the pupil is given by r_p, the index of refraction of air is given by n_{air} and the index of refraction of the aqueous humor is given by n. The index i denotes the pupil perimeter point (from 1 to 6).

computed POG. The calibration consists of computing the error in the estimated POG when the user is looking at each of the four corners of the monitor

$$E_i = (M_i - N_i) \tag{2.12}$$

Future POG estimates are adjusted by applying the four correction factors E_i, each weighted inversely proportional to the distance the computed POG is from each of the original calibration POGs as shown in (2.13) through (2.15).

$$d_i = \|P_{computed} - N_i\| \tag{2.13}$$

$$w_i = \frac{1}{d_i \cdot \sum_{k=1..4} \frac{1}{d_k}} \tag{2.14}$$

$$P_{corrected} = P_{computed} + \left(\sum_{i=1..4} w_i \cdot E_i \right) \tag{2.15}$$

In the event that any d_i is 0, w_i is set to 1 in (2.14) and the remaining weights set to zero.

2.3.5 Eye and Feature Tracking

The points I_i used in Section 2.3.2 and K_i used in Section 2.3.3 are located on the surface of the camera CCD sensor. These locations are determined from information extracted from the recorded images using video processing techniques. The image processing required to extract these features is the most processor intensive operation of the system. To reduce the required processing time, a series of regions-of-interest (ROI) calculations are employed to reduce the quantity of image information. Initially the full image (Figure 2.6(a)) must be processed to detect the location of the eye. The ROI is then applied, sized to contain only the image of the eye (Figure 2.6(b)). When the pupil in the eye is detected roughly, the size of the ROI reduces further to contain just the cornea and pupil for final processing (Figure 2.6(c)). When the image processing has completed, the ROI is increased in size to encompass the eye and re-centered on the estimated center of the pupil image contour for the next processing loop. Re-centering the ROI on the pupil allows the ROI of Figure 2.6(b) to effectively track the eye without having to reprocess the entire image. If the eye is lost (due to a blink) or

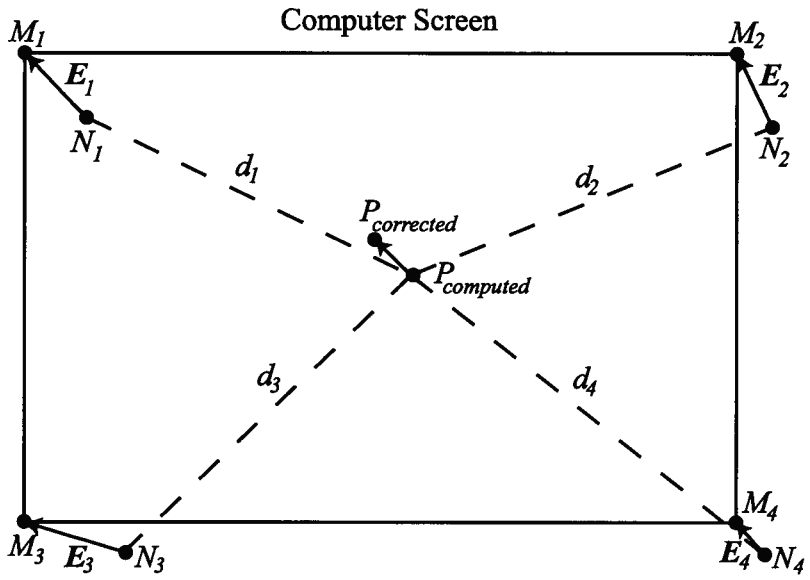


Figure 2.5: Example POG calibration and correction. The system is initially calibrated by recording the computed POG locations N_i while the user is looking at known screen locations M_i. The error E_i is used to convert future POG estimates $P_{computed}$ to $P_{corrected}$. The distance d_i from the point $P_{computed}$ and each of the calibration locations N_i is used to weight the correction factors. The index i denotes the calibration position for each of the four monitor corners.

moves outside of the ROI within one frame the entire image is reprocessed and the ROI then reapplied.

To compute the pupil center, points along the perimeter of the pupil contour image are required. The image differencing technique [76] is used to aid in identifying the pupil contour. Images are recorded with alternating light sources, one in which the pupil is brightly illuminated from lighting close to the optical axis of the camera and one in which the scene is illuminated by off axis light sources. The off-axis lighting illuminates the face to the equivalent intensity of the bright pupil image but does not cause the pupil to reflect as brightly. A ring of LED's located around the optical axis of the camera are used to generate the bright pupil image. Two lights located beside the computer screen generate the dark pupil image, which will also then contain the required dual glints. Subtracting the dark pupil image from the bright pupil image enhances the pupil contour, making it easier to detect in the scene.

The pupil contour is detected in two stages to improve system accuracy and performance. The pupil is first identified quickly and roughly in the scene using the difference image. A finer pupil detection algorithm is then used to extract the pupil contour from just the bright pupil image. Using just the bright pupil image avoids differencing artifacts due to motion between image frames. The fine pupil detection method also compensates for possible artifacts which may corrupt the pupil perimeter, such as glints or eyelashes. An example of the identified pupil contour is shown in Figure 2.7(a).

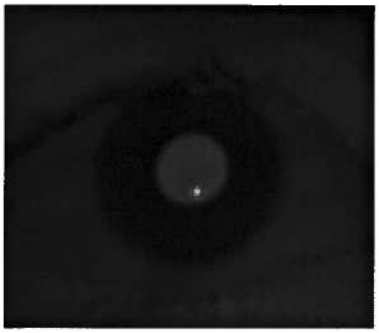
The locations of the centers of the dual glints in the recorded images are required for computing the center of the cornea. The glints off the surface of the cornea result in the brightest pixels in the image and are easily detected. Possible artifacts are rejected using the expected displacements between the two glint centers. Examples of the identified dual glints are shown in Figure 2.7(b).

2.4 Evaluation

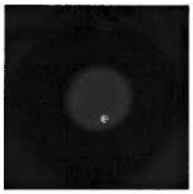
2.4.1 Implementation

The physical implementation of the eye-gaze tracking system is shown in Figure 2.8. The system was tested on a moderately powerful AMD 1.4 GHz computer and a higher end Pentium IV 2.8 GHz computer. Using different computers provides some insight into how well the system will perform with respect to the available processing power.

(a) Full sized scene image

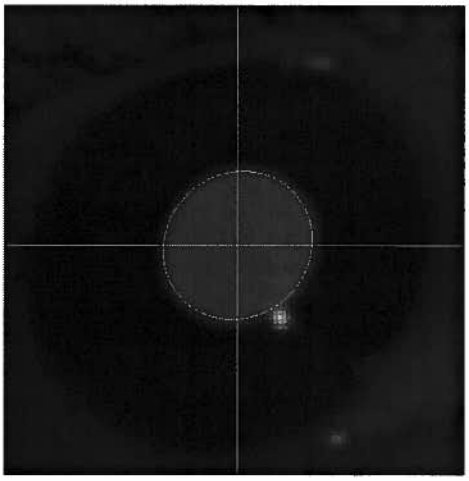


(b) Eye ROI

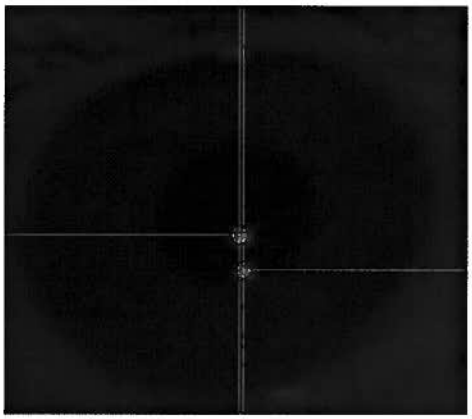


(c) Pupil ROI

Figure 2.6: Regions of interest used to decrease processing time.



(a) Bright Pupil Image



(b) Dark Pupil Image

Figure 2.7: Identified pupil (a) and dual glints (b). An ellipse equation is fitted to the perimeter of each identified contour. The pupil perimeter ellipse is used to estimate the real pupil center location, while the centers of the dual glint ellipses are used to estimate the center of the cornea.

The system was also tested with two different cameras, one with a resolution of 1024 x 768 pixels and a frame rate of 15 Hz, and another camera with a resolution of 640 x 480 pixels and a frame rate of 30 Hz. Both cameras are versions of the digital Firewire based Dragonfly from Point Grey Research. Using different cameras also provides insight into how the system may perform with respect to available frame rates and image resolutions. The higher resolution camera had an allowable range of motion of approximately 14 x 12 x 20 cm (width x height x depth) while for the faster but lower resolution camera the allowable range of motion reduced to approximately 7.5 x 5.5 x 19 cm. The width and height are specified at approximately the midpoint of the field of view volume. The focal length of the lens for both cameras was 32 mm.

Agilent HSDL-4220 880 nm diodes were used for scene illumination. An optical low pass filter was used on the camera to filter out ambient visible light and pass only the system generated lighting.

2.4.2 Free Head Motion

The accuracy of the system was measured over the full range of allowable head positions. The AMD system was used with the 15 Hz camera which had the larger field of view.

Calibration was performed by the system user at position 1. Accuracy was then measured by recording the POG error when looking at points on a 4 x 4 grid on the screen. Average accuracy was determined for a total of 7 different head locations within the allowable field of view. An electromagnetic position tracker was worn by the user during this test to verify that the full field of view was spanned. The range of X, Y and Z positions reported by the position tracker across the field of view volume was 14.2 cm, 12.3 cm, and 20.6 cm respectively.

The average accuracies (difference between POG estimate and reference point) at each head location are listed in Table 2.1 in units of screen pixels. The screen had dimensions of 35 cm and 28 cm in width and height respectively, and a resolution of 1280 x 1024 pixels. Accuracy in pixels rather than degrees of visual angle is reported here because the distance from the eye to the screen required for computing degrees of visual angle was not readily available. For ease of comparison, accuracy in the subsequent test is reported in pixels as well. An average value for the distance from eye to screen is used to convert from pixels to degrees of visual angle in the Discussion in Section 2.5.

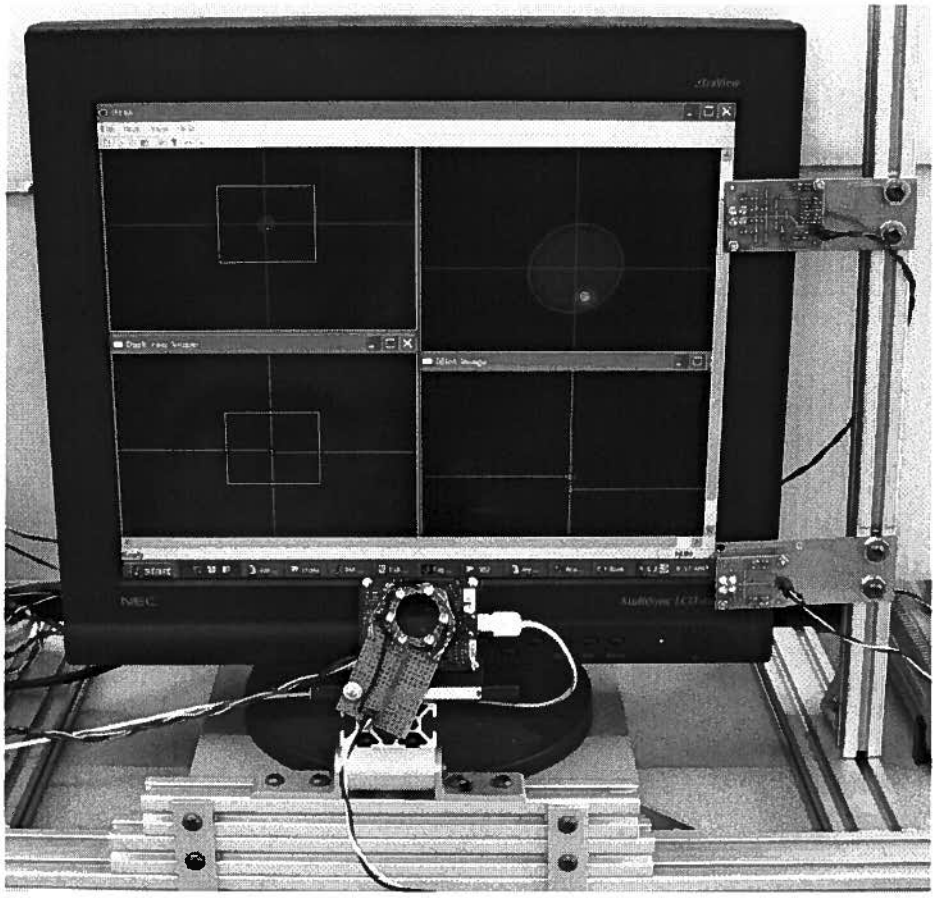


Figure 2.8: The physical system implementation. The digital Firewire camera is located below the screen and oriented towards the users face. The on-axis lighting is provided by the ring of LEDs surrounding the camera lens, while the dual glint off-axis light sources are located to the right of the monitor. The entire assembly is mounted on extruded aluminum rails to fix the relative displacements of the LEDs, camera and screen.

Table 2.1: Average POG accuracy measured across a 4 x 4 grid for each different head position.

	Average Accuracy (Pixels)	
Position	X	Y
1	17.3	15.3
2	41.6	21.4
3	25.0	27.3
4	20.9	18.7
5	33.8	25.5
6	35.6	23.6
7	46.5	21.0

2.4.3 Multiple Hardware Configurations and Subjects

Two subjects were tested on different hardware configurations to test the ability of the system to handle several different subjects and operating conditions. In addition the subjects were evaluated at a calibrated position (Trial 1) and away from the calibrated position (Trial 2) to evaluate the range of accuracies over the free head motion.

The test procedure was to perform a calibration and then record a dataset on the 4 x 4 grid (Trial 1). The user was asked to move away from the system, then to return and sit down in front of the computer again, resulting in a different head position. A second 4 x 4 grid dataset was recorded away from the calibrated position (Trial 2). The average accuracies for these tests are shown in Table 2.2.

The time required to process one video image for each system configuration was recorded both when the ROI was locked on the eye and when the eye was lost. When the ROI is locked on the eye only a small portion of the image is processed; when the ROI is lost the full image must be processed to reacquire the eye. These processing times are shown in Table 2.3.

2.5 Discussion

Across the span of possible head positions (see Table 2.1), the best average pixel errors for the uncalibrated positions in X and Y are [20.9, 18.7] pixels and at the worst are [46.5, 21.0] pixels. Across various hardware configurations and different subjects (see Table 2.2), when the eye was not at the

Table 2.2: Average POG accuracy measured across a 4 x 4 grid for multiple trials, subjects and system configurations.

	Subject 1 Ave Accuracy (Pixels)				Subject 2 Ave Accuracy (Pixels)			
	Trial 1		Trial 2		Trial 1		Trial 2	
	X	Y	X	Y	X	Y	X	Y
AMD 30 Hz	21.5	20.3	33.5	18.1	18.9	17.3	15.1	19.7
AMD 15 Hz	33.6	29.7	29.1	34.2	20.9	22.4	22.8	22.7
P4 30 Hz	31.2	27.0	26.0	27.6	19.1	19.4	32.2	22.9
P4 15 Hz	31.4	23.7	24.8	27.8	23.1	21.9	14.8	21.7

Table 2.3: Processing times per system update when the ROI is locked on the eye and when the eye is lost. Each of the four combinations of system configurations were tested.

	Processing Time (ms)	
	ROI Lock	Eye Lost
AMD - 30 Hz 640 x 480	27	35
AMD - 15 Hz 1024 x 768	28	110
P4 - 30 Hz 640 x 480	10	32
P4 - 15 Hz 1024 x 768	10	40

calibration location (Trial 2), the best average errors in X and Y are [14.8, 21.7] pixels and the worst are [29.1, 34.2] pixels. At an average distance of 75 cm from the eye to the screen for Trial 2 the best average accuracy in degrees of visual angle is 0.46° and the worst is 0.90°.

The system was able to estimate the POG over the full range of allowable head positions and with variations in processing power, camera resolution and camera frame rate. When the ROI was locked on to the eye, there was little difference in processing time required between the higher and lower resolution cameras, due to the equivalent ROI size for both cameras. These times indicate that the AMD system could achieve a maximum update rate of 35Hz while the P4 system could achieve a maximum update rate of 100Hz. The maximum update rates however were limited due to the lower frame rates of the cameras to 15 Hz for the higher resolution camera and 30 Hz for the lower resolution camera. The system update rate matches the

camera frame rates even though alternating bright and dark pupil images are recorded (required for the image differencing technique). The equivalent rates are achieved by estimating the POG using the latest captured image and the previously captured image (either bright then dark or dark then bright pupil images). When the ROI lock was lost, the processing time increased in all cases, thus reducing the effective system update rate. The processing time increased more for the higher resolution camera than for the lower resolution camera when the ROI lock was lost, as expected.

Increasing to a higher resolution camera increases the allowable range of head locations. Increased resolution of the ROI eye images may also be expected to improve the accuracy of the feature detection and consequently of the estimated POG. Increasing the frame rate permits a faster update rate and even faster reacquisition times. We found that the processing time required for the higher resolution images was not much greater than that required for the lower resolution images, provided the ROI was locked onto the eye. For our system, we expect similar results for even higher resolutions using the same size ROI.

2.6 Conclusions

This chapter describes the design, implementation and evaluation of an eye-gaze tracking system that meets key requirements as described in the introduction. As quantified below, the single camera, multiple glint system achieves the accuracy claimed in the presence of free head motion, within the field of view of the camera. Over various combinations of hardware configurations and subjects the best accuracy achieved with the eye away from the calibration position (Trial 2), averaged over the 4 x 4 screen grid, was 0.46° and the worst was 0.90° of visual angle, which is comparable to that of other reported systems. System accuracy is highest at the calibrated position and degrades slightly as the head is moved away.

The system developed has an allowable range of head positions of approximately 14 x 12 x 20 cm for a 1024 x 768 pixel resolution camera. As expected, although higher camera resolution increases the allowable range of head positions, for equivalent spatial resolution it does not necessarily improve eye gaze accuracy. There are no moving parts, resulting in fast re-acquisition times. For the P4 system, re-acquisition of the eye after loss of lock can be achieved in 67 ms for a 15 Hz camera and 33 ms for a 30 Hz camera. Employing a single camera with no moving parts also allows the use of a one-time per user calibration procedure that takes less than 5

seconds.

The system is capable of operating on platforms of varying processing power and with cameras of various resolutions and frame rates to provide a performance (as described in the Discussion) that is scalable to the task.

Acknowledgment

The authors are grateful to the Natural Sciences and Engineering Research Council of Canada for the funding of this project under the IRIS NCE program, and Discovery Grant A9341.

References

[60] K. Rayner, "Eye movements in reading and information processing: 20 years of research." *Psychol Bull*, vol. 124, no. 3, pp. 372–422, Nov 1998.

[61] L. Petersson, L. Fletcher, N. Barnes, and A. Zelinsky, "An interactive driver assistance system monitoring the scene in and out of the vehicle," in *IEEE International Conference on Robotics and Automation*, vol. 4, Apr 26-May 1, 2004, pp. 3475–3481Vol.4.

[62] J. P. Hansen, K. Tørning, A. S. Johansen, K. Itoh, and H. Aoki, "Gaze typing compared with input by head and hand," in *Proceedings of the 2004 symposium on Eye tracking research & applications.* New York, NY, USA: ACM Press, 2004, pp. 131–138.

[63] G. L. Lohse, "Consumer eye movement patterns on yellow pages advertising," *Journal of Advertising*, vol. 26, no. 1, pp. 61–73, 1997.

[64] S. Zhai, C. Morimoto, and S. Ihde, "Manual and gaze input cascaded (magic) pointing," in *CHI '99: Proceedings of the SIGCHI conference on Human factors in computing systems.* New York, NY, USA: ACM Press, 1999, pp. 246–253.

[65] P. Majaranta and K.-J. Räihä, "Twenty years of eye typing: systems and design issues," in *Proceedings of the 2002 symposium on Eye tracking research & applications.* New York, NY, USA: ACM Press, 2002, pp. 15–22.

[66] L. Young and D. Sheena, "Methods & designs: survey of eye movement recording methods," *Behav. Res. Methods Instrum.*, vol. 5, pp. 397–429, 1975.

[67] C. H. Morimoto and M. R. M. Mimica, "Eye gaze tracking techniques for interactive applications," *Comput. Vis. Image Underst.*, vol. 98, no. 1, pp. 4–24, 2005.

[68] T. Ohno, N. Mukawa, and A. Yoshikawa, "Freegaze: a gaze tracking system for everyday gaze interaction," in *Proceedings of the 2002 symposium on Eye tracking research & applications.* New York, NY, USA: ACM Press, 2002, pp. 125–132.

[69] C. H. Morimoto, A. Amir, and M. Flickner, "Free head motion eye gaze tracking without calibration," in *CHI '02 extended abstracts on Human factors in computing systems.* New York, NY, USA: ACM Press, 2002, pp. 586–587.

[70] S.-W. Shih and J. Liu, "A novel approach to 3-d gaze tracking using stereo cameras," *IEEE Transactions on Systems, Man and Cybernetics, Part B*, vol. 34, no. 1, pp. 234–245, Feb. 2004.

[71] D. H. Yoo and M. J. Chung, "A novel non-intrusive eye gaze estimation using cross-ratio under large head motion," *Comput. Vis. Image Underst.*, vol. 98, no. 1, pp. 25–51, 2005.

[72] B. Noureddin, P. D. Lawrence, and C. F. Man, "A non-contact device for tracking gaze in a human computer interface," *Comput. Vis. Image Underst.*, vol. 98, no. 1, pp. 52–82, 2005.

[73] T. Ohno and N. Mukawa, "A free-head, simple calibration, gaze tracking system that enables gaze-based interaction," in *Proceedings of the 2004 symposium on Eye tracking research & applications.* New York, NY, USA: ACM Press, 2004, pp. 115–122.

[74] D. Beymer and M. Flickner, "Eye gaze tracking using an active stereo head," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, 18-20 June 2003, pp. II–451–II–458.

[75] D. A. Goss and R. W. West, *Introduction to the Optics of the Eye.* Butterworth Heinemann, 2001.

[76] Y. Ebisawa, "Improved video-based eye-gaze detection method," *IEEE Transactions on Instrumentation and Measurement*, vol. 47, no. 4, pp. 948–955, Aug. 1998.

Chapter 3

Fixation Precision in High Speed Non-Contact Eye-Gaze Tracking [2]

3.1 Introduction

Eye-gaze tracking systems offer great promise as an interface between humans and machines. Eye-gaze can provide insight into the intention of a user, as a user typically looks at objects of interest before acting upon them [77]. Real-time eye-gaze tracking systems allow dynamic interaction between the user and system using the human visual system for both feedback and control [78]. Tracking the fixations of a user provides a means for using eye-gaze information as a pointing device [79]. The use of eye-gaze as an input modality has not had widespread appeal with the general population however, due in part to the shortcomings of current eye-gaze tracking technology. Some of the key issues which must be improved upon are accuracy, precision, latency, ease of use, comfort and cost [80] [81].

Recent advances in the development of non-contact video-based eye-gaze tracking systems have removed the need for contact with the user and have greatly improved the user's comfort [82]. Non-contact systems coupled with advanced Point-Of-Gaze (POG) estimation algorithms which compute the location of the eye in 3D space can now operate without significantly restricting the user's head motion. The increased freedom of motion greatly improves the ease of use of the system.

Eye-gaze tracking systems in general and non-contact video-based systems in particular suffer from low precision, or fluctuating fixation estimates. The low precision is caused not just by sensor and system noise but is also

[2] A version of this chapter has been published. Hennessey, C., Noureddin, B., and Lawrence, P. 2008. Fixation Precision in High Speed Non-Contact Eye-Gaze Tracking. IEEE Transactions on Systems, Man and Cybernetics, Part B, vol 38, no. 2, pp. 289-298, April 2008.

due in part to the natural motions of the unconstrained head and eye. Considerable research has focused on developing real-time applications which compensate for the low precision including the use of large pointing targets [83] [84], fisheye lenses [85], and enhanced pointing algorithms such as MAGIC pointing [79] and the Grab and Hold Algorithm [86].

In this chapter a definition for fixation precision in the context of eye-gaze tracking is provided. Techniques for improving the precision of non-contact, video-based eye-gaze tracking systems at very high sampling rates are described. The high speed sampling techniques developed are evaluated on the High Speed Pupil-Corneal Reflection vector method (HS P-CR) and a 3D model-based POG method allowing free head motion, at each of three different POG sampling rates. Given the achieved performance of each POG method it is shown how digital filtering can be used to improve fixation precision at each POG sampling rate for both methods.

3.2 Background

3.2.1 Eye Movements

Although the surrounding world appears stable, the head and eyes are continuously in motion and the images formed on the retinas are constantly changing. The stable view of the external world is only an artificially stabilized perception. Natural human vision is typically made up of short, relatively stable fixations connected by rapid reorientations of the eye (saccades). It is during fixations that the sensory system of the eye collects information for cognitive processing; during saccades, sensitivity of visual input is reduced [87].

Fixations typically remain within 1° of visual angle and last from 200 to 600 ms [77]. While fixating, the eye slowly drifts, with a typical amplitude of less than 0.1° of visual angle and a frequency of oscillation of 2 to 5 Hz. This drift is corrected by small fast shifts in eye orientation called microsaccades which have a similar amplitude to the drift. Superimposed on this motion is a tremor with a typical amplitude of less than 0.008° of visual angle with frequency components from 30-100 Hz and at times up to 150 Hz [88]. These small eye motions during a fixation are thought to be required to continuously refresh the sensors in the eye, as an artificially stabilized image will fade from view [89].

Saccades most frequently travel from 1 to 40° of visual angle and last 30 to 120 ms. Between saccades there is typically a 100 to 200 ms delay [77]. A number of other task specific eye motions exist, such as smooth pursuit,

nystagmus and vergence which are not often found in the normal interaction between a user and a desktop monitor. The focus of this chapter is on the POG during fixations which are located between saccadic reorientations of the eye.

3.2.2 Fixation Detection and Filtering

A clear identification of the beginning and end of a fixation within the raw eye data stream is important as filtering should only be performed on POG data located within a single fixation. Poor identification of the beginning or end of a fixation may result in a degradation in fixation precision by incorporating POG data from saccades or neighboring fixations. There have been a number of methods developed for identifying the start and end of fixations in raw eye data streams using position, velocity and acceleration thresholding based on *a priori* knowledge of the behavior of eye-gaze movements [90] [91]. The fixation identification method used in this chapter is based on position variance of eye data as described by Jacob [77].

Due to the natural motions of the eye, fixation precision in eye-gaze tracking systems may be low, limiting the range of potential applications. However, as noted by Jacob [77], this low precision can be improved by low pass filtering the estimated POG data to reduce noise, at the expense of increased latency. The desired degree of filtering within a fixation will depend on the particular application under consideration. For high precision a higher order filter may be used at the expense of a longer latency or lag between the start of a fixation and the desired filter response. Alternatively a lower order filter may be used to allow the POG fixation to drift slightly over time to follow the natural drift of the eye. Using digital finite-impulse-response (FIR) filtering techniques allows the filter order to be easily modified, as well, clearing the filter history (memory) provides a simple means for reseting the filter when a fixation termination is detected.

3.2.3 Eye-gaze Tracking Systems

The development of non-contact eye-gaze tracking systems is an important step in improving the acceptability of eye-gaze as a general form of human machine interface. One of the recent trends in eye-gaze tracking systems has been away from systems requiring contact with the subject's face and head and towards non-intrusive and non-restrictive systems.

Contact based methods such as electro-oculography (EOG), the scleral search coil and head mounted video-oculography (VOG) are seen as less

desirable due to the requirement for contact with the users head, face or eyes. The EOG and search coil methods do benefit however from the ability to record the subject's eye gaze electronically, rather than optically as in the case of video-based tracking. Electronic recording can be performed easily at high data rates (1000's of Hz) using modern analog to digital integrated circuits. The sampling rate of video-based systems is limited to at most the frame rate of the imaging cameras (typically 30 Hz) and is often even lower due to the image processing techniques used and the high computational power required to process large quantities of image data in real-time.

In the late 1980's Hutchinson *et al.* [92] developed a non-contact video-based system which used the P-CR vector method for computing the POG. The P-CR method greatly enhanced the usability of remote eye-gaze tracking systems by providing tolerance to minor head displacements. The system they developed was targeted to work with the severely disabled who had no other easily available means of communication. Images were recorded with a resolution of 512 x 480 pixels with a POG sampling rate of 30 Hz. After calibration, average accuracies for this method are typically 0.5 to 1° of visual angle.

Over the past two decades, the P-CR vector method has been the favored means for non-contact, video-based POG estimation. However, the P-CR method still required a relatively stable head position. The accuracy of the method degrades considerably as the head is displaced from the calibration position [82].

To allow for free head motion, Shih and Liu [93] developed a novel 3D model-based method for estimating eye-gaze. Using models of the system, camera and eye, their algorithm was designed to accurately estimate the POG regardless of head location. Their system used two RS-170 based cameras and frame grabbers to record images with a resolution of 640 x 280 pixels at 30 Hz. Average accuracy was shown to be better than 1° of visual angle. Unfortunately their system design required the cameras to be quite close to the subjects' eyes to acquire high spatial resolution images, restricting the freedom of head motion due to the limited camera field of view.

To overcome the limitation of a narrow field of view, Ohno and Mukawa [94] developed a 3D model-based system with a camera mounted on a pan / tilt mechanism with a Narrow Angle (NA) lens, and two fixed cameras with Wide Angle (WA) lenses. The fixed cameras used stereo imaging to determine the location of the head within the scene and directed the pan / tilt mechanism to orient the NA camera towards the eye. The WA cameras recorded images with a resolution of 320 x 120 pixels while the NA camera

recorded images with a resolution of 640 x 480 pixels, all at frame rates of 30 Hz. System accuracy was reported as better than 1.0° of visual angle. The pan / tilt mechanism allowed the NA camera to track the motion of the eye with a larger effective field of view; however, the speed at which the mechanism could move was not sufficient to keep up with the faster motion of the head and eye resulting in loss of tracking and slow re-acquisition.

Beymer and Flickner [95] used high speed galvonometers for their 3D model-based system in an attempt to overcome the limitations of the slow pan / tilt systems. A pair of fixed WA cameras used stereo imaging to direct the orientation of two NA cameras by controlling the pan and tilt of rotating lightweight mirrors mounted on galvonometers. The focus of each camera was controlled with a lens mounted on a bellows and driven by another motor. The NA cameras recorded NTSC images (with a typical resolution of 640 x 480 pixels) at a frame rate of 30 Hz. Due to the significant processing involved in the system a POG sampling rate of only 10 Hz was achieved. The accuracy reported for this system was 0.6° degrees of visual angle. While their system was capable of tracking the eye in the presence of natural high speed head motion, considerable calibration was required, and the overall complexity resulted in a low POG sampling rate.

The 3D model-based system by Hennessey *et al.* [96] was developed to minimize the physical system complexity while still allowing for fast head motion. The system was based on a single fixed camera with a high resolution sensor and no moving parts. The higher resolution sensor allowed a larger range of head motion with the eye remaining in the field of view of the camera, while still providing images with sufficient spatial resolution for the eye-tracking system to operate correctly. The system algorithms were designed to track the motion of the eye within the image and only operate on the portion of the image containing the eye. Processing only the portion of the image containing the eye allowed the POG to be computed rapidly, regardless of the overall image resolution. At the time of system development a camera with a resolution of 1024 x 768 pixels was available with a maximum frame rate of 15 Hz. Using this system, accuracies of better than 1° of visual angle were achieved.

The accuracies reported are the measure of conformity of the measured POG value with the true POG as determined by the system user. The precision during a fixation of the POG estimates can be defined as the degree of agreement among a series of individual POG measurements [97]. Fixation precision is typically quantified as the standard deviation of the recorded POG estimates. Fixation precision has not often been reported in evaluations of 3D model-based eye-gaze tracking systems as the focus tended

to be on the basic system functionality and accuracy of the novel POG algorithms. However, Yoo and Chung [98] did provide some insight into the fixation precision of their free head motion eye-gaze tracking system. Using a similar system design as Ohno and Mukawa [94] they reported an accuracy of 0.98° in horizontal error and 0.82° in vertical error when operating at 15 Hz. Precision in standard deviations was reported in millimeters which converted to 0.84° of visual angle. We believe that fixation precision is an important parameter in the evaluation of the performance of eye-gaze tracking systems and the goal of this chapter is to present methods for enhancing fixation precision.

3.3 Methods

3.3.1 Point-of-Gaze Estimation

There are currently two main types of methods for computing the POG from remote video images, the P-CR vector method and the 3D model-based method.

P-CR Method

The simplicity of the P-CR vector method and it's ability to handle minor head motions led to its widespread adoption. As the eye rotates to observe different points, the image of the reflection off the spherical corneal surface remains relatively fixed. The corneal reflection, generated by external lighting, provides a reference point for determining the relative motion of the pupil. A simple mapping is used to relate the 2D POG screen vector to the 2D image vector formed from the center of the corneal reflection to the center of the pupil as shown in Fig. 3.1.

Independent polynomial equations are determined to relate the 2D P-CR vector (g_x, g_y) to each of the 2D POG screen co-ordinates (p_x, p_y). The polynomial order varies between different system designs but is most often of first order as shown in (3.1). It has been shown that small increases in accuracy may be achieved by increasing the order of the polynomial, at the expense of a decrease in robustness to head motion and the need for an increasing number of calibration points [99].

$$\begin{aligned} p_x &= a_0 + a_1 g_x + a_2 g_y + a_3 g_x g_y \\ p_y &= b_0 + b_1 g_x + b_2 g_y + b_3 g_x g_y \end{aligned} \tag{3.1}$$

The parameters a_i and b_i are determined from a calibration procedure

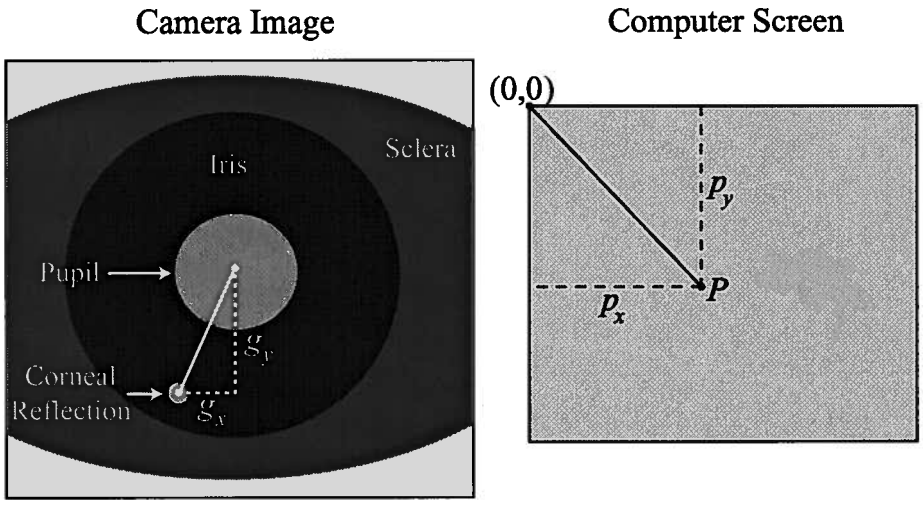


Figure 3.1: In this figure an example of a portion of a recorded bright pupil image is shown to illustrate the P-CR vector. In the P-CR method the vector (g_x, g_y) is determined from the center of the corneal reflection to the center of the pupil. A mapping is then defined to relate the P-CR vector to the POG screen coordinates (p_x, p_y).

in which the user fixates sequentially on a number of known screen locations while the P-CR vector is recorded. In the case of a first order polynomial fit, a minimum of 4 calibration points are required to solve for the 4 unknowns in each of the two equations in (3.1), typically using a least squares method.

3D Model-Based Method

Algorithms based on 3D models have been developed to overcome the degradation in accuracy that the P-CR method suffers with larger head movements. The 3D model-based methods compute the position of the eye in 3D space, which is then used in computing the POG regardless of the head and eye position. There are a large variety of 3D model-based algorithms, although each technique is typically based on a model of the physical system, camera and eye. The physical system is modeled geometrically through physical measurement or using optical methods as in [93]. The camera lens is modeled as a pin-hole with parameters identified through camera calibration [100] [101]. The models of the eye are most often based, in varying levels of sophistication, on the schematic eye developed by Gullstrand [102]. An example of a typical eye model with three parameters is shown in Fig. 3.2. Per-user calibration is required to fit the eye model parameters to individual users.

Feature information is extracted from recorded images and fit to the system models to solve for the location of the eye in 3D space, the Line-Of-Sight (LOS) and ultimately the POG as shown in Fig. 3.2. The location of the eye in 3D space is found by determining the center of the cornea C, when modeled as a spherical surface, using triangulation with images of multiple corneal reflections. With the 3D location of the cornea center and the image location of the center of the pupil, the 3D LOS vector can be computed. The LOS can be traced from C to intersect with any surface point P in the system by determining the parameter t in (3.2). The object of intersection is typically the surface of the computer screen which is parameterized as a plane in the system model.

$$P = C + t \cdot LOS \tag{3.2}$$

3.3.2 Image Processing

Both the P-CR vector and 3D model-based methods for estimating the POG require features extracted from the recorded images. The P-CR method requires the location of the pupil and the location of a single corneal reflection,

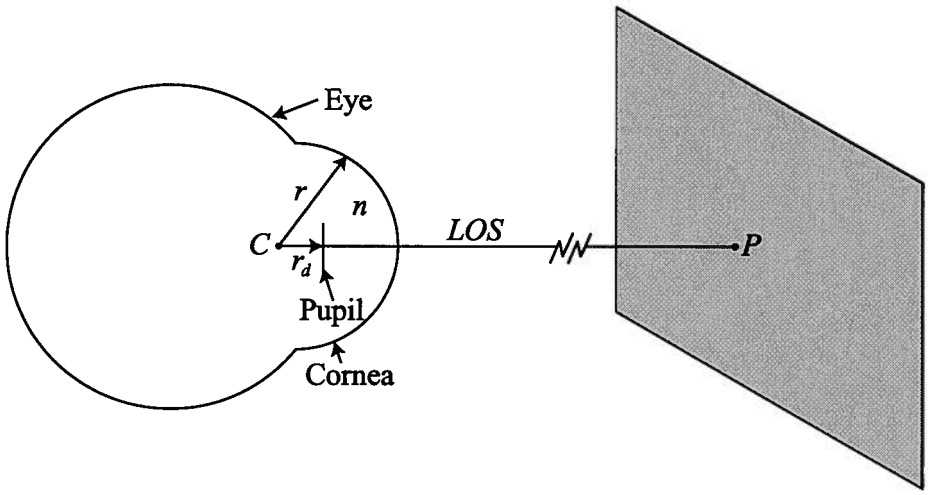


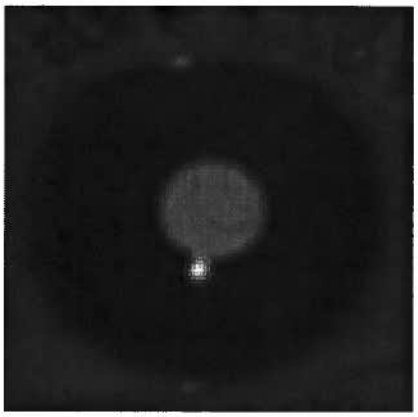
Figure 3.2: The 3D model-based method for computing the POG is based on determining the location of the center of the cornea and the line-of-sight vector. Using (3.2) the POG can be found by tracing the LOS vector from C to the surface of the screen P. The model of the eye is based on the schematic description by Gullstrand which in this case includes three parameters; the radius of the model of the corneal sphere r, the distance from the center of corneal sphere to the center of pupil r_d and the index of refraction of the aqueous humor fluid n.

while the 3D method requires the pupil and at least two corneal reflections for triangulation. The location of the pupil and corneal reflections are found by identifying the perimeter of their respective image contours. The pupil contour perimeter can be considerably difficult to segment due to the low contrast between the pupil and the surrounding iris. The corneal reflections can be difficult to segment due to their small size, often less than 3 x 3 pixels. Varying levels of ambient light can compound the feature extraction difficulty.

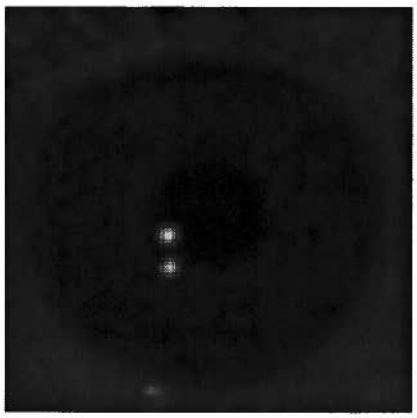
To improve the performance of the feature extraction task the bright pupil and image differencing techniques are used to create a high contrast image of the pupil [103] [104]. Computing a difference image from alternating bright pupil and dark pupil images removes most of the background features, ideally leaving only the high contrast pupil on a black background. An example of the bright pupil and image differencing techniques are shown in Fig. 3.3. Using a single on-axis light source generates a single corneal reflection which is used in the P-CR POG estimation method. By using two off-axis light sources for the dark pupil image, the two corneal reflections required for the 3D method are generated.

While the image differencing technique aids in the identification of the pupil contour within the image, it is also susceptible to significant artifacts which may corrupt the identified contour. When the difference image is computed, the corneal reflections formed by the off-axis lighting in the dark pupil image can result in removing a portion of the pupil as seen in the lower left side of Fig. 3.3(c). Also seen is the addition to the pupil contour of the corneal reflection from the on-axis lighting. The difference image is also susceptible to significant artifacts due to inter-frame motion. Inter-frame motion may distort the extracted pupil contour by misaligning the bright and dark pupil images which will significantly impact the accuracy of the POG estimation algorithms.

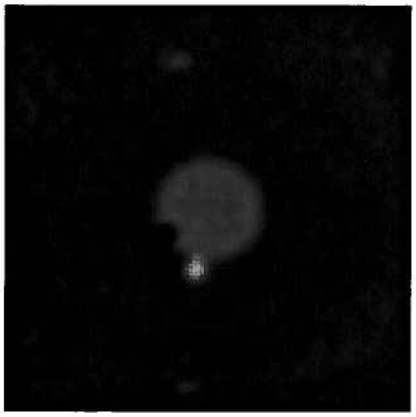
To avoid the inaccuracies resulting from inter-frame motion and the image differencing, a two stage approach to pupil detection was used. The first stage of pupil extraction determines the image difference pupil as described above. The corneal reflections are then identified in both the bright and dark pupil images based on their proximity to the roughly identified difference pupil (see Fig. 3.4(a)). In the second stage of pupil identification the pupil contour is segmented in only the bright pupil image using the previously detected difference pupil as a guide. Using only the bright pupil image avoids errors due to inter-frame motion and the accidental removal of pupil area by the subtraction of the dark pupil corneal reflections. The final step of the second stage is to mask off the portion of the pupil contour which may be



(a) Bright Pupil Image



(b) Dark Pupil Image



(c) Difference Image

Figure 3.3: Illustration of the bright pupil and image differencing techniques. The bright pupil in Fig. 3.3(a) is illuminated with on-axis lighting, while the dark pupil in Fig. 3.3(b) is illuminated with off-axis lighting. The background intensity of the two images is similar, which after differencing (3.3(a) - 3.3(b)) results in a bright pupil on an almost blank background as shown in Fig. 3.3(c).

due to the addition of the on-axis corneal reflection (see Fig. 3.4(b)). The resulting pupil perimeter retains its elliptical shape when compared with the initial roughly identified pupil perimeter. For a more detailed description of the methods used for pupil and corneal reflection segmentation see [105].

Before passing the identified pupil and glint locations to the POG estimation algorithms, the identified contour perimeters are further refined using an ellipse fitting algorithm which is both fast (computationally efficient) and robust to noise [106]. Subpixel accuracy in the identification of the contour centers may be achieved by using the center of the equation of an ellipse fit to the contour perimeters [107]. As well, using an ellipse fit to the available pupil perimeter points compensates for the loss of data when a gap appears as a result of the masking operation to remove the corneal reflection from the on-axis lighting.

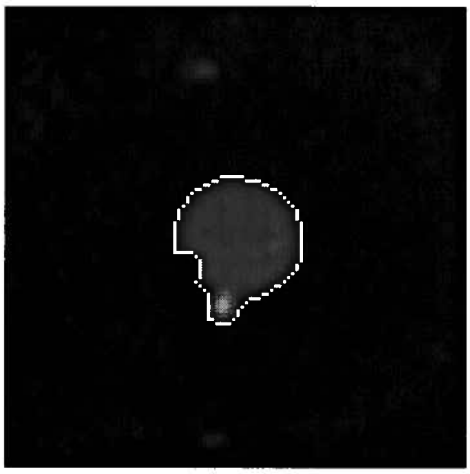
3.3.3 Point-of-Gaze Sampling Rate

The POG sampling rate in video-based eye-gaze tracking systems is at most equal to the frame rate of the camera, although it is often less due to image processing requirements and techniques such as image differencing. In order to achieve high speed eye-gaze tracking the POG sampling rate must be maximized.

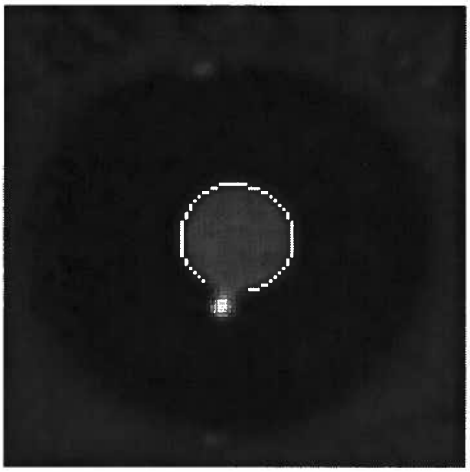
Software Region-Of-Interest

Image processing algorithms can be considerably time consuming due to the large quantity of information to process. To greatly reduce the processing load for our system, a software based Region-Of-Interest (ROI) was employed to constrain the processing to only the image area of interest. In the design of our system, rather than using mechanical tracking, the camera field of view encompasses a large area which allows the eye to move around within the scene. Accordingly, only a small portion of the overall scene contains information of interest as shown in Fig. 3.5(a).

The location of the ROI is continuously updated to track the location of the eye, which allows for head motion within the field of view of the camera. Initially, the first captured images are processed in their entirety to identify the location of the pupil within the overall scene. The ROI is then centered on the eye as each frame is processed and the center of the pupil identified. In this fashion only a small portion of the image will normally be processed. In the event that the pupil is lost due to blinking or rapid head or eye motion which relocates the eye outside of the ROI between image frames, the entire



(a) Difference Pupil Contour



(b) Bright Pupil Contour

Figure 3.4: An example of the results of the two stage pupil detection algorithm. In Fig. 3.4(a) the detected perimeter of the identified image difference pupil contour is shown overlying the difference image. Using the difference pupil contour as a guide, the pupil perimeter is detected in the bright pupil image as shown in Fig. 3.4(b). The gap in the pupil perimeter is a result of masking off the on-axis corneal reflection, which is subsequently compensated for by fitting an ellipse to the bright pupil contour perimeter.

image is reprocessed until the pupil location is re-identified or in the case of a blink, the eye reopens.

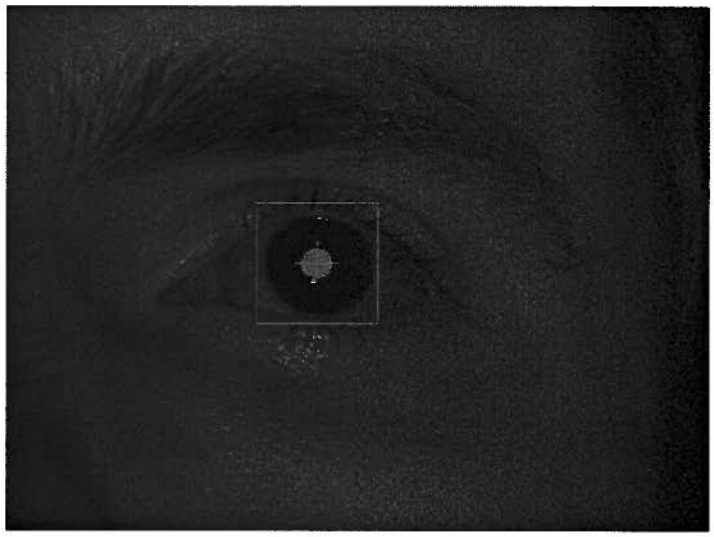
Hardware Region-Of-Interest

The basis of data reduction using a software ROI may also be applied to the reduction of data transmit from the camera to computer. Reducing the transmission of information per frame allows for an increase in the overall frame rate and consequently the maximum achievable POG sampling rate. The Firewire2 (IEEE-1394b) Digital Camera (DCAM) specification for data transmission defines the operation of hardware based ROIs (using Format 7), although some variation in behavior may be found between different camera manufacturers. Using commands in the Firewire2 protocol, the camera can be configured to apply a hardware ROI to an image before the imaging sensor is exposed and read.

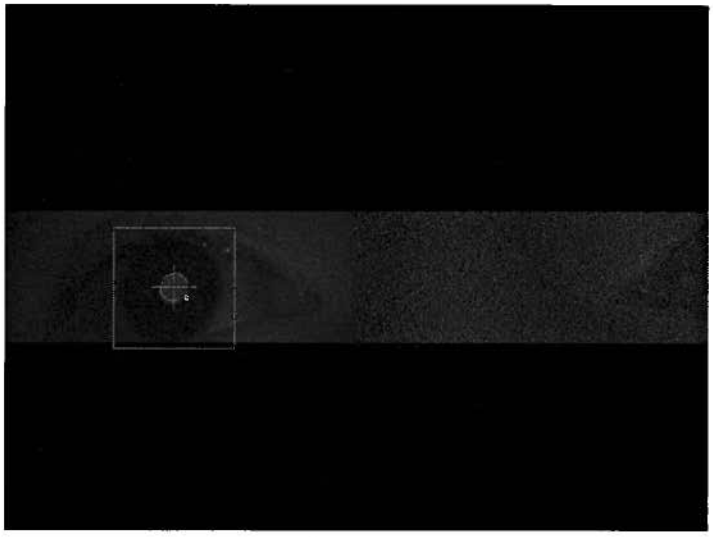
The frame rate for the camera used by our system (described in Section 3.3.4) only increased by skipping image rows, no frame rate improvement was achieved for skipping image columns. Using the software ROI in conjunction with the hardware ROI allowed the flexibility to maximize the frame rate while minimizing the required processing. Similar to the software ROI, the location of the hardware ROI was re-centered on the pupil each image frame to track the motion of the eye. Unfortunately, changing the location of the hardware ROI in real-time aborted the exposure of the current image, resulting in an underexposed image for one frame. To minimize the number of hardware ROI location changes, the size of the hardware ROI was chosen to be the full width of the original image and slightly larger than the height of the cornea, while the size of the software ROI was set to the width of the cornea and slightly smaller than the height of the hardware ROI as shown in Fig. 3.5(b). The software ROI then tracks all horizontal motion and most small vertical motions without requiring a change in the hardware ROI location. The hardware ROI is then only repositioned for larger vertical displacements in the position of the eye.

Image Sequencing

Recording alternating bright and dark pupil images for the image differencing technique aids in the detection of the pupil within the overall scene, however it also reduces the effective POG sampling rate. When a 1:1 ratio of alternating bright and dark pupil images are recorded, the P-CR method can only generate a unique POG (P_i) at half the camera frame rate as



(a) Full Image and Software ROI



(b) Hardware and Software ROI's

Figure 3.5: Regions-Of-Interest are used to reduce the quantity of image information to process as well as increase the camera frame rate. In Fig. 3.5(a) only the software ROI is applied to the original full sized bright pupil image (640 x 480 pixels). Only the portion of image within the rectangular box (110 x 110 pixels) surrounding the eye will be processed. In Fig. 3.5(b) the hardware ROI (640 x 120 pixels) has been applied in addition of the software ROI.

shown in Table 3.1, since all the information required to compute the POG is contained within the bright pupil image. Recall that for the P-CR POG estimation method the image features required are the pupil and a single corneal reflection, which are both found in the bright pupil image. The 3D method uses image information from both the bright and dark pupil images and as such can compute a unique POG (P_i) at the camera frame rate by using features from each current image (f_{i+1}) along with the image previously recorded (f_i).

In the HS P-CR method reported here, the system operation was enhanced by increasing the sampling rate of unique POG estimates through increasing the ratio of bright pupil images with respect to dark pupil images. As the HS P-CR method only requires the dark pupil image to roughly identify the location of the pupil in the scene, the ratio of bright to dark pupil images may be increased until inter-frame motion results in loss of tracking due to misaligned image differencing. To illustrate the improvement in POG sampling rate an example of a 3:1 bright to dark pupil ratio is also shown in Table 3.1 in which the sampling rate has increased from 50% of the camera frame rate to 75%.

Increasing the rate of unique POG estimates for the HS P-CR method by increasing the ratio of bright to dark pupil images is preferable to maintaining a 1:1 ratio and using a corneal reflection from the dark pupil image as is done in the 3D method. In the HS P-CR method, using image information for POG estimation from only a single bright pupil image (see Table 3.1) avoids the errors in POG estimation that may result from misaligned bright and dark pupil image features due to inter-frame motion.

Unfortunately a similar technique cannot be used for the 3D method to avoid inter-frame motion while increasing the POG update rate. The 3D method would require two additional corneal reflections in the bright pupil image to compute the POG with information contained solely in a single image. The extra reflections would have to be masked off of the pupil contour as described in section 3.3.2, potentially removing large portions of the pupil contour and consequently decreasing the accuracy of the pupil feature identification. The corneal reflection from the on-axis lighting in the bright pupil image cannot be used with the 3D method as the on-axis light source is located coaxially with the focal point of the camera, which results in a singularity in the 3D model algorithm, see Equation (4) in [96].

Table 3.1: POG sampling sequences for HS P-CR and 3D POG estimation methods with 1:1 and 3:1 bright to dark pupil ratios.

Frame Sequence	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	...
1:1 ratio									
Image Type	D	B	D	B	D	B	D	B	...
P-CR POG	-	P_1	-	P_2	-	P_3	-	P_4	...
3D POG	-	P_1	P_2	P_3	P_4	P_5	P_6	P_7	...
Frame Sequence	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	...
3:1 ratio									
Image Type	D	B	B	B	D	B	B	B	...
HS P-CR POG	-	P_1	P_2	P_3	-	P_4	P_5	P_6	...

Dark pupil image (D), Bright pupil image (B), No unique POG sample (-)

3.3.4 Hardware

The Dragonfly Express from Point Grey Research was the digital camera used for the system described in this chapter. The camera is capable of recording full sized images of 640 x 480 pixels at frame rates up to 200 Hz. To increase the frame rate further, a hardware ROI was used to reduce the size of the recorded images.

The camera uses the Firewire2 (IEEE-1394b) standard to transmit images from the camera to the computer. An electronic strobe signal generated by the camera at the start of each image frame was monitored by a custom microcontroller to synchronize the on-axis and off-axis lighting with the image exposure. The microcontroller also controlled the ratio of bright to dark pupil images as directed by the computer through the serial port.

The system evaluation was performed on a Pentium IV 3 GHz processor with 2 GB of RAM. A flat screen LCD monitor with a width of 35.8 cm and a height of 29.0 cm was set to a resolution of 1280 x 1024 pixels and located at a distance of approximately 75 cm from the users eye. The physical system is shown in Fig. 3.6.

3.4 Experimental Design and Results

The techniques to perform high-speed, non-contact eye-gaze tracking described above were evaluated with the HS P-CR and 3D model-based methods for estimating the POG. Both POG methods were tested at three different camera frame rates to determine the effect of sampling rate on fixation

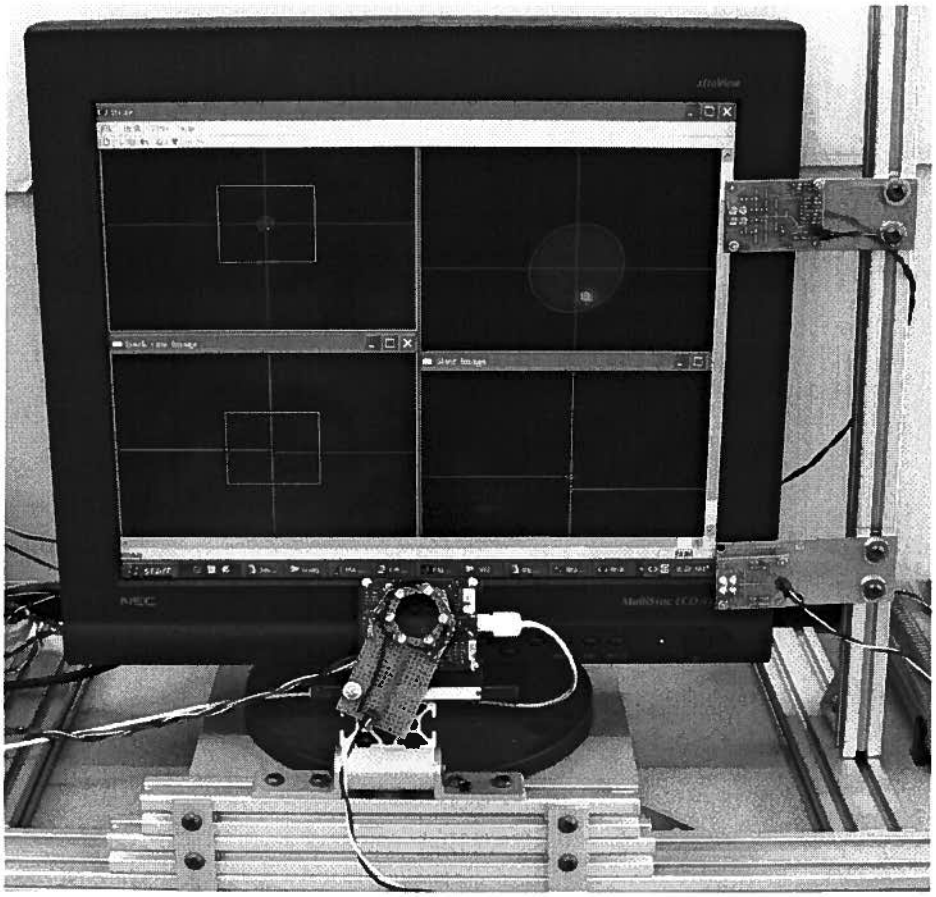


Figure 3.6: Physical system showing the camera located beneath the monitor, the on-axis lighting (ring of LEDs surrounding the lens), the two off-axis point light sources located to the right of the monitor and the monitor upon which the POG is estimated.

precision. Varying levels of digital filtering were applied to the recorded data for each POG method at each frame rate to show the resulting improvements in precision.

The sequences of POG estimates were collected on a total of four different subjects while performing a simple task, with a data set recorded for each combination of the two POG methods and three camera frame rates, resulting in a total of 6 data sets per subject and 24 data sets overall. The camera frame rates tested were 30 fps, 200 fps and 407 fps which allows for comparison between the equivalent of a 30 fps NTSC systems, 200 fps achievable when recording full sized images without a hardware ROI (640 x 480 pixels), and 407 fps achievable with the hardware ROI enabled (640 x 120 pixels).

The experimental procedure was comprised of a calibration phase, a familiarization phase and then the performance of a simple task during which the POG screen coordinates were recorded. The calibration consisted of having the subject observe the four corners of the screen for approximately one second each while the per-user parameters were estimated. After calibration, a short familiarization period was allowed in which the calibration was evaluated with the subject verifying that the computed POG across the screen was in fact the same (or at least very close to) their real POG. The subject was then asked to fixate on nine sequential points on a 3 x 3 grid which were displayed across the screen. Throughout the fixation task the screen coordinates of the POG were continuously recorded, along with a flag indicating the fixation status at each grid point. The fixation status flag was set to indicate the beginning of a fixation when the relative stability of a fixation was detected, and the flag was cleared when the larger motion of a saccade was detected, as per the position variance algorithm described by Jacob [77]. At least two seconds of fixation data was acquired before moving to the next point. An example of the fixation data collected on the 3x3 grid for a single subject is shown in Fig. 3.7 while a subset of 10 POG estimates from a single fixation point are shown in Fig. 3.8.

As discussed previously, the POG sampling rate for the HS P-CR POG estimation method was enhanced by increasing the ratio of bright to dark pupil images for the 200 fps and 407 fps camera frame rates. At 30 fps the ratio had to remain at 1:1 bright to dark pupil images as higher ratios resulted in frequent loss of tracking due to inter-frame motion and misaligned image difference pupil contours. At the higher camera frame rates, higher ratios were possible while still maintaining tracking as the magnitude of the motion between each image frame was less. Since loss of tracking rarely occurred at the 1:1 ratio and 30 fps rate, a similar period between dark

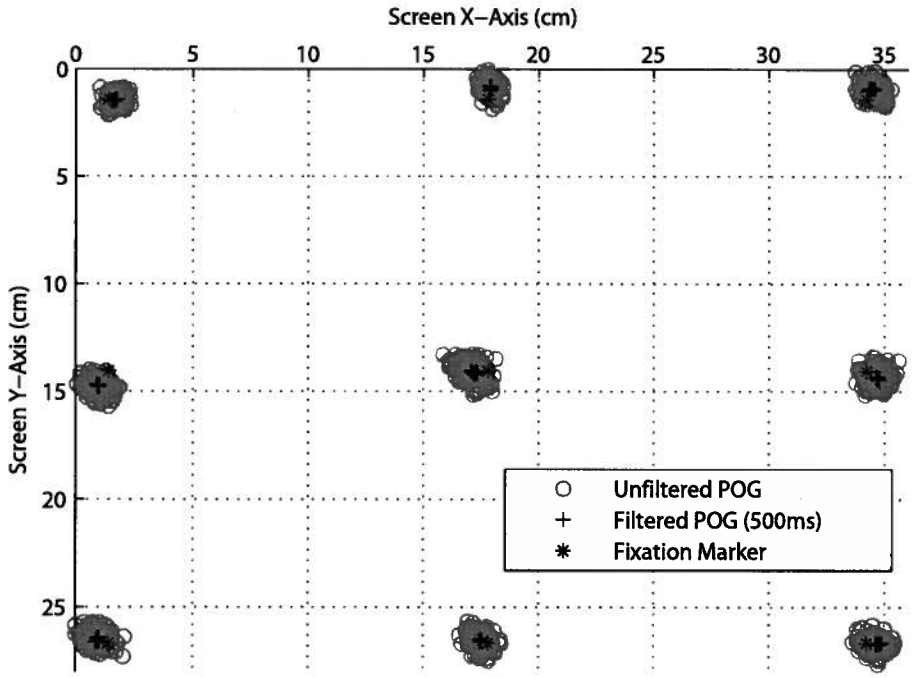
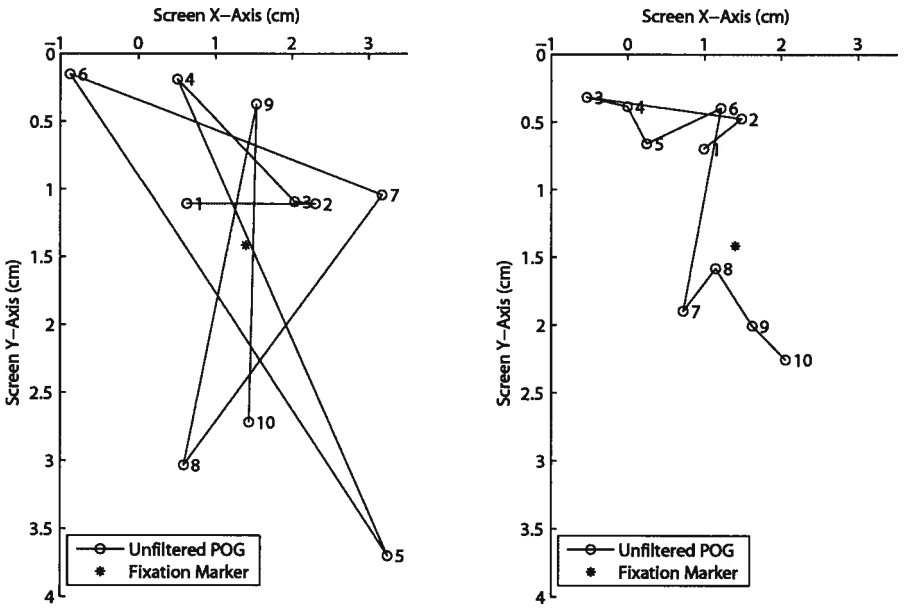


Figure 3.7: An example of the fixation task in which the user observed each of 9 points on a 3 x 3 grid. In this example the POG samples were recorded with the HS P-CR vector method and a camera frame rate of 407 Hz. The original POG data is shown along with the results of filtering with a 500 ms moving window average. The POG screen coordinates have been converted from units of pixels to centimeters in this figure.



(a) 3D POG estimation at 30 Hz

(b) 3D POG estimation at 407 Hz

Figure 3.8: A labeled sequence of 10 unfiltered POG estimates for the 3D POG estimation method are shown from a single fixation marker. Sampling sequences at two camera frame rates are illustrated, 30 Hz shown in Fig. 3.8(a) in which the 10 point sequence corresponds to a time interval of 333 ms, and 407 Hz shown in Fig. 3.8(b) which corresponds to a time interval of 25 ms.

pupil images was used for the higher camera frame rates. The achieved HS P-CR update rates for each camera frame rate along with the corresponding bright to dark pupil image ratios are listed in Table 3.2.

Table 3.2: Image sequence parameters for the HS P-CR POG method.

Camera Frame Rate (fps)	**Bright to Dark Pupil Ratio**	**Dark Pupil Period (ms)**	**POG Sampling Rate (Hz)**
30	1:1	66	15
200	9:1	50	180
407	19:1	49	386

Low pass filtering of the recorded sequence of POG screen coordinates was performed offline for each subject and each system configuration. Filtering the POG data offline allowed for comparison of various levels of filtering on a consistent set of data. The recorded X and Y POG coordinates were filtered with a rectangular window FIR filter (moving average) with filter lengths corresponding to latencies (window lengths) of 30 ms, 100 ms and 500 ms. The filter order for each system configuration was determined from the POG sampling rate and the desired latency as listed in Table 3.3. The three filter lengths were chosen to contrast the difference in fixation precision with latencies up to the duration of a typical fixation.

Table 3.3: Filter order for each sampling rate and filter length for the HS P-CR and 3D POG estimation methods.

Sampling Rate	Filter Length		
	30 ms	***100 ms***	***500 ms***
HS P-CR Method			
15 Hz	1	1	7
180 Hz	5	18	90
386 Hz	11	39	193
3D Method			
30 Hz	1	3	15
200 Hz	6	20	100
407 Hz	12	41	203

After filtering the recorded X and Y POG coordinates with each of the FIR filters, the fixation precision was determined at each of the 9 fixation

points. The standard deviation was computed on the last 500 ms of the two seconds of data recorded at each fixation point to avoid combining data points from adjacent fixations when high filter orders are used.

The reported fixation precision for each system configuration is the average of the 9 standard deviations for each of the 4 subjects and is reported in degrees of visual angle as shown in Table 3.4. To convert from units of screen pixels to degrees of visual angle the estimated POG and fixation marker reference point are first converted from pixels to centimeters with the scaling factors of 35.8 cm / 1280 pixels for the X coordinate and 29 cm / 1024 pixels for the Y coordinate. The POG error is then computed as the difference between the estimated POG (p_x, p_y) and the fixation marker reference point (r_x, r_y). It is assumed that in the worst case, the eye is located along a vector normal to the screen that extends from the midpoint of the POG error vector. The equation to convert from pixels to degrees of visual angle (θ) is then shown in (3.3) with the assumption that the average distance from eye to screen was 75 cm.

$$\theta = 2 \cdot \tan^{-1}\left(\frac{\sqrt{(p_x - r_x)^2 + (p_y - r_y)^2}}{2} \cdot \frac{1}{75}\right) \tag{3.3}$$

Table 3.4: Fixation Precision for each system configuration.

Sampling Rate	Filter Length			
	None	***30 ms***	***100 ms***	***500 ms***
HS P-CR Method				
15 Hz	0.205	0.205	0.205	0.065
180 Hz	0.258	0.173	0.112	0.051
386 Hz	0.199	0.115	0.071	0.035
3D Method				
30 Hz	0.550	0.550	0.306	0.108
200 Hz	0.390	0.288	0.200	0.074
407 Hz	0.347	0.230	0.155	0.050

Note: All units in degrees of visual angle

3.5 Discussion

Using the techniques described above, operation of the remote eye-gaze tracking system at high sampling rates was achieved. The higher sampling

rates more accurately record the faster dynamics of the eye and reduce signal aliasing. Using the Nyquist criterion the sampling rate should be at least twice the highest frequency of the micro-saccades and tremors (up to 150 Hz [88]) observed during fixations. To illustrate the effect of aliasing a labeled sequence of POG estimates is shown with a low sampling rate (30 Hz) in Fig. 3.8(a) and at a much higher sampling rate (407 Hz) in Fig. 3.8(b). For the lower sampling rate the details of the trajectory of the POG are missing as illustrated by the erratic and large displacements between subsequent POG estimates. At the higher sampling rate the trajectory of POG estimates can more clearly be seen as the displacement between estimates is smaller.

Processing the incoming images at 200 fps was achieved with the use of only the software ROI. With the addition of the hardware ROI, the camera frame rate increased to 407 fps. Using the 3D model-based POG estimation algorithm the sampling rate was equal to the camera frame rate: 30 Hz at 30 fps, 200 Hz at 200 fps and 407 Hz at 407 fps. When using the HS P-CR method for estimating the POG an update rate of only 15 Hz was achieved when operating at 30 Hz due to the requirements of the image differencing technique. With the reduced inter-frame motion at higher frame rates it was possible to enhance the P-CR method by increasing the ratio of bright to dark pupil images without losing lock on the eye. Increasing the bright to dark pupil ratio to 9:1 for the 200 fps frame rate increased the POG sampling rate to 180 Hz and increasing the ratio to 19:1 at 407 fps increased the sampling rate to 386 Hz. The POG update rates achieved for the HS P-CR and 3D methods are a significant increase over the rates achieved by similar eye-gaze tracking systems discussed in the background review of this chapter.

The fixation precision reported for the 3D model-based POG method at the lowest sampling rate (30 Hz) was 0.55°. This result is of a similar magnitude to the precision reported by Yoo and Chung [98] at 0.84° for their non-contact, free head, eye-gaze tracking system, which operated at a rate of 15 Hz. The benefit of our system is the ability to increase the POG sampling rate which then allows digital filtering to further improve fixation precision while still maintaining an acceptable latency. Using digital low pass filtering resulted in an improvement in fixation precision in all system configurations as shown in Table 3.4. In the experiments performed, the best fixation precision was achieved with the longest filter (500 ms), which for the HS P-CR method resulted in a standard deviation of 0.035° or 1.6 screen pixels and 0.050° or 2.3 screen pixels for the 3D model-based method. The relationship between filter length and fixation precision appears to be exponential as shown in Fig. 3.9. As filter length increases, a diminishing return in the

trade off between achieved precision and POG latency is achieved.

The fixation precision of the HS P-CR method was compared with the 3D model-based method at each of the camera frame rates using three one-way ANOVAs. It was found that the HS P-CR method was statistically more precise than the 3D method at 30 fps ($F(1,70) = 87.168, p < 0.001$), 200 fps ($F(1,70) = 17.939, p < 0.001$) and 407 fps ($F(1,70) = 38.273, p < 0.001$). This result is possibly due to motion of the eye between the image frames used to compute the POG in the 3D method. It is possible that the natural eye motions between image frames results in misaligned bright and dark pupil image features, increasing the variability of the estimated POG and consequently decreasing the fixation precision. Supporting this theory is the improvement in fixation precision for the 3D method when the camera frame rate increases, decreasing the time between image frames and consequently reducing the degree of potential inter-frame motion.

A comparison of accuracy between the two methods was not performed as the focus in this chapter is on fixation precision. A more detailed investigation of system accuracy is presented in [96]. While not the focus of this chapter, system accuracy was confirmed to be comparable to many contemporary remote eye-gaze tracking systems [82]. Averaged over all subjects and all operating conditions the HS P-CR method resulted in an accuracy of 0.72° while the 3D method accuracy was 1.0° of visual angle. The accuracy of the HS P-CR method appears slightly better in these experiments; however, the measurements were only recorded with the head located near the calibration position and did not explicitly exercise the free head capabilities of the 3D model-based method.

3.6 Conclusions

The precision of eye-gaze tracking systems within fixations is a key factor in determining the usability of eye-gaze tracking for human computer interaction. In this chapter the start and end of fixations were detected using position variance thresholding. The precision of a fixation was then computed as the standard deviation of the POG estimates temporally located between the beginning and end of the fixation.

Techniques were presented which enable video-based, non-contact, eye-gaze tracking systems to operate at high POG sampling rates, more adequately recording the dynamics of high speed eye movements. A high speed method for P-CR POG estimation was also presented in which the sampling rate was increased by modifying the ratio of bright pupil to dark pupil im-

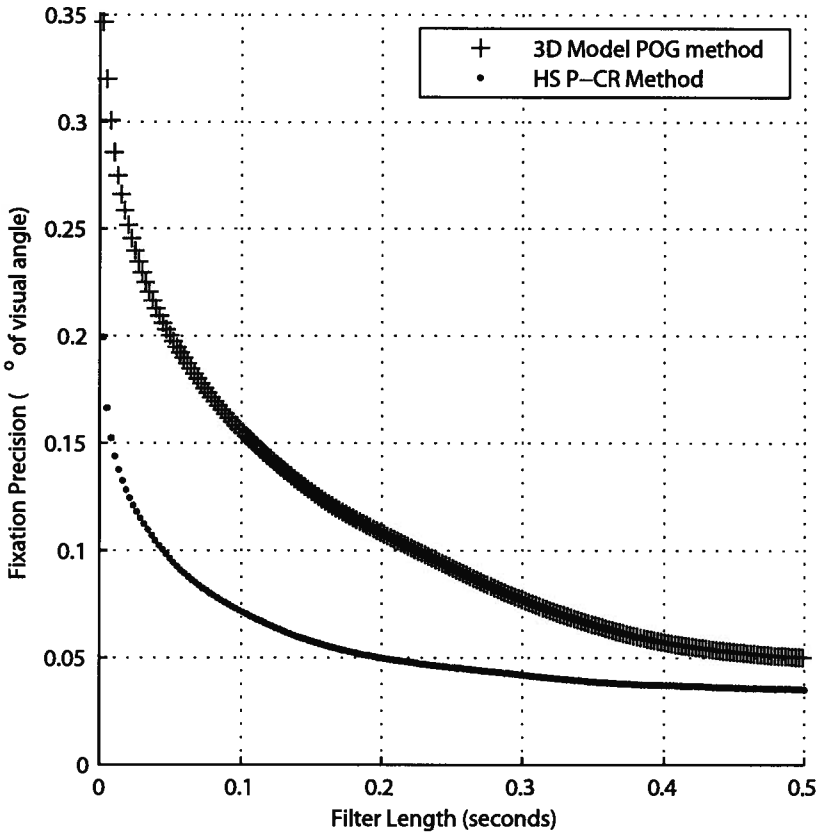


Figure 3.9: Fixation precision verses filter length is shown averaged across all four subjects indicating an exponential relationship. The POG screen coordinates were recorded with the system operating at 407 fps for both the HS P-CR and 3D POG methods.

ages. Increasing the frequency of bright pupil images increased the frequency of the images containing the features required to compute the POG.

The high speed techniques were evaluated on both the HS P-CR and 3D model-based POG methods. Within the fixations defined by the position variance thresholding, fixation precision was shown to improve through the application of low pass digital filters. Higher POG sampling rates allowed for a trade-off between fixation precision and real-time POG latency, depending on the intended user application. An exponential relationship was observed between filter order and fixation precision, indicating a diminishing incremental improvement with increasing filter orders.

A comparison between the HS P-CR POG estimation method and the 3D model-based method showed that the fixation precision for the HS P-CR method was significantly better than the 3D method at each of three camera frame rates tested. One possible explanation for this result is that the HS P-CR POG estimation method avoided the misalignment of image feature data resulting from inter-frame motion by using information from only a single image to compute the POG. Although the 3D method is shown to be less precise, it does allow a wider range of head motion [96] than the HS P-CR method [82]. In this study however, subjects were asked to maintain a comfortable, relatively stationary head-position.

Future work will focus on the evaluation of the techniques presented in this chapter on a larger sample of subjects. Integration of these methods with an eye-gaze tracking system for use in the real-world is also desirable to increase the realism of the eye-gaze tracking experiments.

Acknowledgment

The authors would like to express their appreciation for the support of the Natural Sciences and Engineering Research Council of Canada (NSERC) Chair in Design Engineering, and NSERC Discovery Grant #4924-05.

References

[77] R. Jacob, *Virtual Environments and Advanced Interface Design.* New York, NY, USA: Oxford University Press, 1995, ch. Eye tracking in advanced interface design, pp. 258–288.

[78] R. Jacob and K. Karn, *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research.* Amsterdam: Elsevier Science, 2003, ch. Eye Tracking in Human-Computer Interaction and Usability Research: Ready to Deliver the Promises (Section Commentary), pp. 573–605.

[79] S. Zhai, C. Morimoto, and S. Ihde, "Manual and gaze input cascaded (magic) pointing," in *CHI '99: Proceedings of the SIGCHI conference on Human factors in computing systems.* New York, NY, USA: ACM Press, 1999, pp. 246–253.

[80] K. S. Karn, S. Ellis, and C. Juliano, "The hunt for usability: tracking eye movements," in *CHI '99 extended abstracts on Human factors in computing systems.* New York, NY, USA: ACM Press, 1999, pp. 173–173.

[81] H. Collewijn, *Vision research: A practical Guide to Laboratory Methods.* Oxford University Press, 1999, ch. Eye movement Recording, pp. 245–285.

[82] C. H. Morimoto and M. R. M. Mimica, "Eye gaze tracking techniques for interactive applications," *Comput. Vis. Image Underst.*, vol. 98, no. 1, pp. 4–24, 2005.

[83] J. P. Hansen, D. W. Hansen, and A. S. Johansen, *Universal Access In HCI.* Lawrence Erlbaum Associates, 2001, ch. Bringing Gaze-based Interaction Back to Basics, pp. 325–328.

[84] D. J. Ward and D. J. C. MacKay, "Fast hands-free writing by gaze direction," *Nature*, vol. 418, no. 6900, p. 838, 2002.

[85] M. Ashmore, A. T. Duchowski, and G. Shoemaker, "Efficient eye pointing with a fisheye lens," in *Proceedings of the 2005 conference on Graphics interface*. School of Computer Science, University of Waterloo, Waterloo, Ontario, Canada: Canadian Human-Computer Communications Society, 2005, pp. 203–210.

[86] D. Miniotas and O. Špakov, "An algorithm to counteract eye jitter in gaze-controlled interfaces," in *Information Technology and Control*, vol. 30, no. 1, Tampere, Finland, 2004, pp. 65–68.

[87] K. Rayner, "Eye movements in reading and information processing: 20 years of research." *Psychol Bull*, vol. 124, no. 3, pp. 372–422, Nov 1998.

[88] A. Spauschus, J. Marsden, D. Halliday, J. Rosenberg, and P. Brown, "The origin of ocular microtremor in man," *Experimental Brain Research*, vol. 126, no. 4, pp. 556–562, June 1999.

[89] U. Tulunay-Keesey, "Fading of stabilized retinal images." *J Opt Soc Am*, vol. 72, no. 4, pp. 440–447, Apr 1982.

[90] M. A. Just and P. A. Carpenter, "A theory of reading: from eye fixations to comprehension." *Psychol Rev*, vol. 87, no. 4, pp. 329–354, Jul 1980.

[91] A. T. Duchowski, *Eye Tracking Methodology: Theory and Practice*. Springer-Verlag, 2003.

[92] T. Hutchinson, J. White, W. Martin, K. Reichert, and L. Frey, "Human-computer interaction using eye-gaze input," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 19, no. 6, pp. 1527–1534, 1989.

[93] S.-W. Shih and J. Liu, "A novel approach to 3-d gaze tracking using stereo cameras," *IEEE Transactions on Systems, Man and Cybernetics, Part B*, vol. 34, no. 1, pp. 234–245, Feb. 2004.

[94] T. Ohno and N. Mukawa, "A free-head, simple calibration, gaze tracking system that enables gaze-based interaction," in *Proceedings of the 2004 symposium on Eye tracking research & applications*. New York, NY, USA: ACM Press, 2004, pp. 115–122.

[95] D. Beymer and M. Flickner, "Eye gaze tracking using an active stereo head," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, 18-20 June 2003, pp. II–451–II–458.

[96] C. Hennessey, B. Noureddin, and P. Lawrence, "A single camera eye-gaze tracking system with free head motion," in *Proceedings of the 2006 symposium on Eye tracking research & applications.* New York, NY, USA: ACM Press, 2006, pp. 87–94.

[97] M. B. Stout, *Basic Electrical Measurements.* Prentice Hall, Englewood Cliffs, N.J., 1960.

[98] D. H. Yoo and M. J. Chung, "A novel non-intrusive eye gaze estimation using cross-ratio under large head motion," *Comput. Vis. Image Underst.*, vol. 98, no. 1, pp. 25–51, 2005.

[99] Z. Cherif, A. Nait-Ali, J. Motsch, and M. Krebs, "An adaptive calibration of an infrared light device used for gaze tracking," in *Proceedings of the 19th IEEE Instrumentation and Measurement Technology Conference*, vol. 2, 21-23 May 2002, pp. 1029–1033vol.2.

[100] R. Tsai, "A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses," *IEEE Journal of Robotics and Automation*, vol. 3, no. 4, pp. 323–344, Aug 1987.

[101] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000.

[102] D. A. Goss and R. W. West, *Introduction to the Optics of the Eye.* Butterworth Heinemann, 2001.

[103] Y. Ebisawa and S. Satoh, "Effectiveness of pupil area detection technique using two light sources and image difference method," in *Proceedings of the 15th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Oct 28-31, 1993, pp. 1268–1269.

[104] C. H. Morimoto, D. Koons, A. Amir, and M. Flickner, "Pupil detection and tracking using multiple light sources." *Image and Vision Computing*, vol. 18, no. 4, pp. 331–335, 2000.

[105] C. Hennessey, "Eye-gaze tracking with free head motion," Master's thesis, University of British Columbia, August 2005.

[106] A. Fitzgibbon, M. Pilu, and R. Fisher, "Direct least square fitting of ellipses," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 5, pp. 476–480, May 1999.

[107] J. Zhu and J. Yang, "Subpixel eye gaze tracking," in *Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition.* Washington, DC, USA: IEEE Computer Society, 2002, p. 131.

Chapter 4

Improving the Accuracy and Reliability of Remote System-Calibration-Free Eye-gaze Tracking [3]

4.1 Introduction

Eye-gaze tracking can be used as a human-machine interface technique for individuals with high level spinal-cord injuries or motor-neuron disorders who are unable to operate standard interface tools such as the keyboard and mouse [108]. While video-based eye-gaze tracking has great potential for improving the quality of life of these individuals, a number of key technical issues need to be improved upon. While the requirements for remote eye-gaze tracking are application dependent, in general, improvements are needed in accuracy, precision, response time, reliability, ability to tolerate head motion and simplification of system and user calibration requirements [109]. Reducing the need for system calibration simplifies the initial user setup of the system, while simplifying the user calibration reduces the time and effort required for per-user calibration. The focus of this chapter will be on increasing the reliability of eye and image feature tracking as well as improving the overall accuracy of a remote, system-calibration-free, eye-gaze tracking system.

Video-based eye-gaze tracking systems can be divided into two categories, head mounted and remote.

[3] A version of this chapter has been submitted for publication. Hennessey, C., and Lawrence, P. 2008. Improving the Accuracy and Reliability of Remote System-Calibration-Free Eye-gaze Tracking. IEEE Transactions on Biomedical Engineering

Head Mounted

Head mounted eye-gaze trackers typically use the Pupil-Corneal Reflection (P-CR) vector method for point-of-gaze (POG) estimation. The P-CR method is a relatively simple technique in which a vector is formed in the recorded images from a single reflection (commonly known as a glint) off the surface of the cornea to the center of the image of the pupil [110]. A polynomial mapping, determined through user-calibration, is then used to relate the 2D camera image vector to 2D POG screen coordinates. The accuracy of head mounted eye-gaze trackers is typically 1° of visual angle or better, though accuracy degrades as the head is displaced from the calibration position, especially in depth [109]. Binocular tracking of both eyes is common with head mounted eye-gaze trackers as two cameras can be placed on the head, one for each eye.

Using a single corneal reflection in the P-CR method can be problematic as eye rotations can result in distortion or loss of the reflection when the reflection nears the boundary between the cornea and scelera, resulting in increased error or system failure. Hua *et al* [111] recently proposed a technique for head mounted P-CR POG estimation using a symmetric arrangement of four light-emitting-diodes (LEDs) to generate a cross shaped pattern of corneal reflections. A virtual point located at the intersection of the horizontal and vertical lines connecting the matching pairs of reflections was then used in forming the P-CR vector. To compensate for the loss of reflections the two pairs of LED's must be placed orthogonally with respect to each other (i.e. vertically or horizontally) and parallel to the camera image plane.

Head mounted systems offer the benefit of fixed head-to-camera displacement, however, mounting the system on the head can result in slippage requiring recalibration. As well, if used over an extended period of time, fatigue can result, and comfort can be a concern [112] [113].

Remote

Remote eye-gaze tracking offers greater comfort and ease of use as the user is not required to wear head-mounted equipment. Early remote eye-gaze tracking systems using the P-CR technique however required a relatively motionless head, as eye motion coupled with head motion resulted in increased error [109]. A recent attempt by Cerrolaza *et al* to overcome this limitation showed promise by tracking the relative displacement of corneal reflections and normalizing the P-CR vector accordingly [114]. They found

that the normalized P-CR vector performed better than the traditional P-CR vector when the head is displaced with depth. When the head is displaced arbitrarily however, a means by which the multiple corneal reflections can be tracked is required. As will be shown later in this chapter, combined head and eye movements can lead to the loss and distortion of the corneal reflections required for P-CR normalization.

To allow for natural head motion more complex techniques for POG estimation have been developed based on models of the eye, camera and physical system. For the model-based techniques the center of the eye in 3D space was determined using multiple corneal reflections, which along with the 3D pupil center were used to form the visual axis along which the user is looking. Intersection of the visual axis with the screen, modeled as a planar surface, resulted in an estimate for the POG.

An early remote system developed by Shih and Lui [115] tracked both eyes using two remote cameras imaging at 30 Hz and mounted close to the subject's eyes. System calibration included stereo camera lens calibration [116; 117], physical system modeling of the computer screen, LEDs, and camera positions, and per user calibration to approximate parameters of the eyes. While only two corneal reflections were required for estimation of the POG, three reflections were used to provide redundancy should one be lost due to eye rotation. An average accuracy over six subjects of slightly better than 1° of visual angle was reported. For this system however only a small degree of head motion (4 x 4 cm with little depth motion) was possible due to the proximity of the cameras to the eyes and the limited depth of field of the lens.

In the system by Ohno *et al*, two cameras and a pan/tilt/zoom mechanism were used to increase the allowable range of head motion while achieving similar accuracy results to Shih *et al.* A wide angle lens camera was used to direct the narrow angle lens pan/tilt/zoom camera to track the eye, however, the mechanical tracking mechanism was too slow to keep up with fast head motions [118]. High speed galvanometer mechanisms were investigated by Beymer *et al* for providing fast mechanical tracking [119]. The tracking mechanism was significantly more complex however, leading to difficult system calibration, and the use of two pairs of stereo cameras led to a low system update rate of 10 Hz.

The system by Yoo *et al* used an eye-model along with a novel cross-ratio method for estimating the POG to reduce the required system calibration to several simple measurements [120]. The cross-ratio method requires four light sources at the four corners of the computer screen and uses the horizontal and vertical ratio of the resulting corneal reflections, along with the

pupil image center to determine the POG. The POG estimation requires all four corneal reflections. Their system used the pan/tilt/zoom technique for tracking the eye with two cameras, achieving a reported accuracy of under 1° of visual angle. The range of head motion was not specified and a 15 Hz update rate was achieved.

With systems based on mechanical tracking of the eye, typically only a single eye is tracked due to the complexity of the mechanical hardware. Tracking a single eye is in general sufficient as both eyes tend to point to the same position [121].

Present Work

In the work presented here, a suite of three novel approaches are presented for improving single camera remote eye-gaze tracking. Firstly a novel technique for tracking a pattern of corneal reflections is presented to provide redundancy for both the P-CR and model-based POG estimation techniques. Tracking the corneal reflection pattern improves the reliability of POG estimation by compensating for the loss and distortion of reflections when both head and eye rotations cause the reflections to move off the surface of the cornea. The tracking technique presented here has fewer restrictions on the placement of the light sources than the method by Hua *et al* and Yoo *et al*, as well as providing a mechanism for detecting distortion of the reflections and not just the complete loss. Secondly, it is shown how tracking the affine transformation parameters of the corneal reflection pattern can be used for enhancing the performance of the P-CR method for operation in system calibration free, remote eye-gaze tracking. The enhanced P-CR vector technique is shown to achieve the same performance as the more complex model-based method which requires considerable system calibration. Thirdly, it is shown that binocular tracking of both the left and right eyes can be achieved using a single remote camera at high speeds without mechanical tracking. A high-speed face tracking technique provides a means for distinguishing the eyes when only a single eye is visible, enlarging the effective lateral head motion range. In the event that one eye translates laterally out of the view of the camera, the other eye, which remains in view, still provides valid monocular POG data for the system.

This chapter also contributes: 1) a unique comparison of the P-CR and model-based experimental POG accuracies for displacements of the head, 2) a list of the image processing times broken down by subtask, illustrating the high speed achievable when processing only a single video stream, as well as a comparison of the processing times for the P-CR and model-based methods,

and 3) a comparison of left and right eye accuracies *vs.* the accuracy of the binocular average of both eyes.

4.2 Methods

A high level overview of the proposed system is outlined in Fig. 4.1. In this system a single camera is used to record images of the face in which both left and right eye tracking is attempted. The identified image features are labeled as coming from either the left or right eye and are then used in the POG estimation algorithm. If both eyes are visible, the POG estimates for the left and right eyes can then be averaged to provide a more accurate estimate of the POG. The image processing and POG estimation stages are described in greater detail in the following subsections.

4.2.1 Image Processing

The P-CR and model-based POG estimation algorithms require accurately identified pupil and corneal reflection image features. The purpose of the image processing stage of the eye-gaze tracking system is to extract these image features accurately and rapidly. The process flow of the image processing stage is outlined in Fig. 4.2(a). The first image processing operation is the face tracking stage outlined in Fig. 4.2(b) which identifies if a face is visible in the camera image. A search is then performed for the image features needed for POG estimation, including the pupil center and corneal reflection contours. An ellipse is fit to each image feature contour with the contour center location then identified at the center of the ellipse [122]. If valid eye features are found in the image, the first identified eye image is blanked out and a second image feature search is performed. Depending on the number of eyes found, the boundary of the identified face is then used to determine which set of detected eye features belong to either the left or right eye.

Face Tracking

When both eyes are visible, the extracted image features can easily be associated with either the left or right eye based on their relative displacements in the image. If only a single eye is visible however, it becomes difficult to determine from which eye the extracted image features originated. The loss of an eye from the extracted image may be due to head motion which positions an eye outside the field of view of the camera. When only one eye

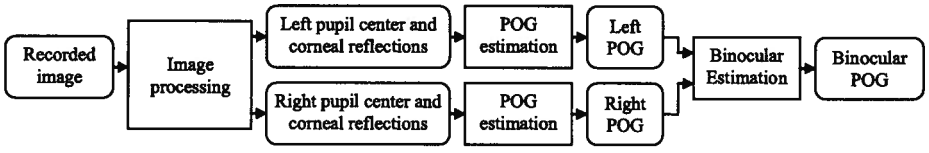
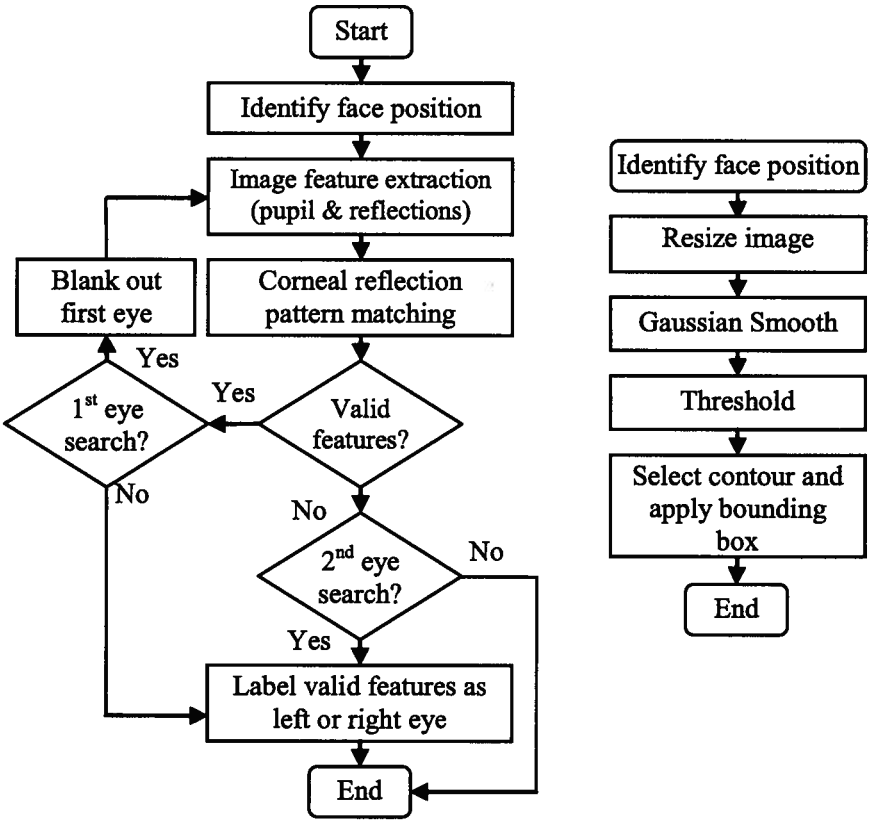


Figure 4.1: The high level binocular eye-gaze tracking system block diagram is shown. The final POG may be estimated from either the left or right eye, increasing reliability to the loss of an eye due to head motion. Alternatively, the final POG can be estimated as the average of the POG estimates from the left and right eyes providing a more accurate estimate of the true POG.



(a) Image Processing (b) Face Tracking

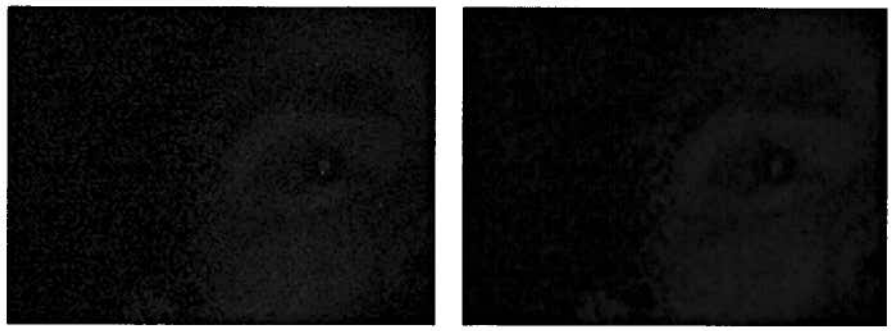
Figure 4.2: Image Processing. After identification of the position of the face in the image, a search is performed for image features from either eye. If image features are correctly identified, the corresponding image pixels are zeroed and a second search takes place for the second eye. If two sets of image features are identified, the left and right eyes are distinguished easily. If only one eye is found, the detected face position is used to determine which eye the image features belong to. If no eye features are correctly identified, the process aborts and begins again on the subsequent recorded image.

is visible, face tracking can be used to determine the position of the eye with respect to the horizontal sides of the face. If the position of the visible eye is closer to the left side of the head, the extracted image features belong to the left eye, while if the position of the eye is closer to the right side of the head the extracted image features belong to the right eye.

There have been numerous techniques developed for face and facial feature tracking, for a literature survey see Zhao *et al* [123]. While existing face tracking algorithms have been shown to operate in real-time (15-30 Hz), a faster technique is required to operate at the 200 Hz used by the eye-gaze tracking system [124]. The face detection process can be simplified however, as only the horizontal sides of the face are required. The image processing stages of the face detection algorithm are outlined in Fig. 4.2(b) with graphical examples shown in Fig. 4.3. Structured lighting is used for the image feature extraction process in which infrared (IR) light sources are used to illuminate the face while an IR filter on the camera lens prevents visible light from being recorded. The low-power IR lighting results in an illuminated face against a dark background as seen in Fig. 4.3(a). After thresholding at a fixed intensity level above the black background, the resulting binary contours are sorted by size and the largest contour identified as the face. The sides of the face visible to the camera are then determined using a bounding box fit to the identified face contour.

There are four possible situations in which the face tracking system is required for eye identification as shown in Fig. 4.4. The limited resolution of the camera used in the system presented here required a long focal length camera lens to provide enough spatial resolution for the extracted image features. The long focal length results in only a partial view of the face in the recorded image. The face detection bounding box therefore only surrounds the portion of the face visible to the camera. To determine the off-image sides of the head an assumed average head width w_h is required.

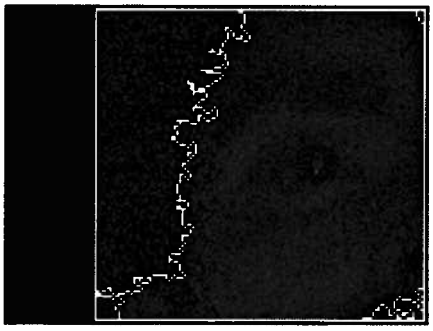
When only one side of the head is visible as in Fig. 4.4(a) and Fig. 4.4(b), it is assumed that the opposite side of the head is w_h pixels to the opposite side and identification of the left or right eye proceeds accordingly. In the event that neither side of the face is observed as in Fig. 4.4(c) and Fig. 4.4(d), it is assumed that an eye in the left half of the image is the left eye while an eye in the right half of the image is the right eye. This assumption holds, even with horizontal head motion, as by the time the left or right eye crosses the centerline of the recorded image, the corresponding side of the head also becomes visible.



(a) Resized

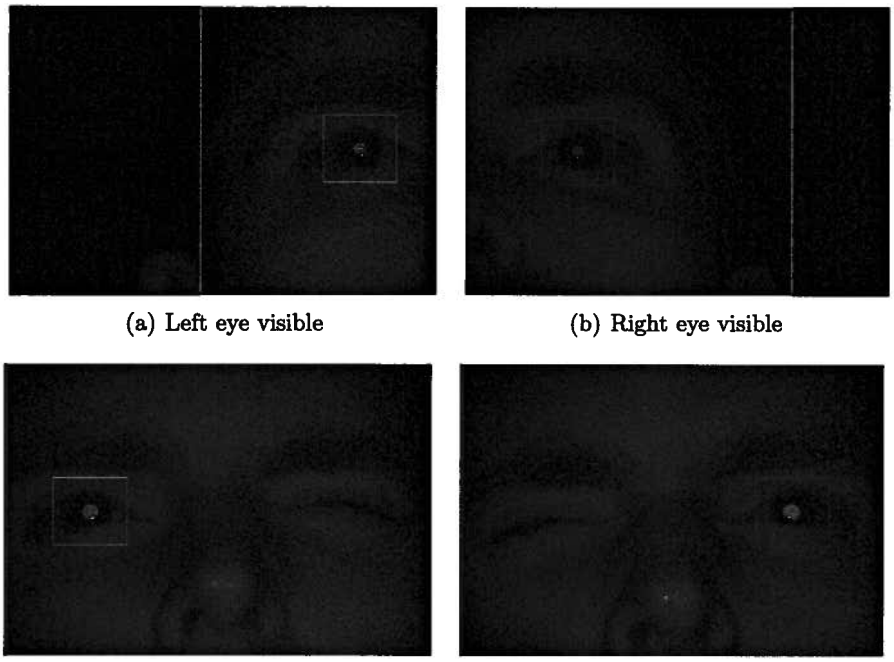
(b) Smoothed

(c) Thresholded



(d) Contour Selection

Figure 4.3: The image processing steps performed by the face tracking algorithm are shown. The algorithm operates at high speed by first reducing the size of the image to 6% of its original size (640x480 to 160x120 pixels) as shown in Fig. 4.3(a). A high gain setting required for the short exposure time results in considerable noise which is smoothed for segmentation as shown in Fig. 4.3(b). The image is then thresholded at a fixed intensity level above the black background as shown in Fig. 4.3(c). After thresholding, the resulting image contours are sorted by size and a bounding box is fit to the largest contour determined as shown in Fig. 4.3(d).



(a) Left eye visible

(b) Right eye visible

(c) Left eye visible, head centered

(d) Right eye visible, head centered

Figure 4.4: When both eyes are visible the left eye is simply the eye on the left and the right eye the eye to the right. When only a single eye is visible the bounding box surrounding the face is used to distinguishing the visible eye based on the proximity of the eye to the side of the face.

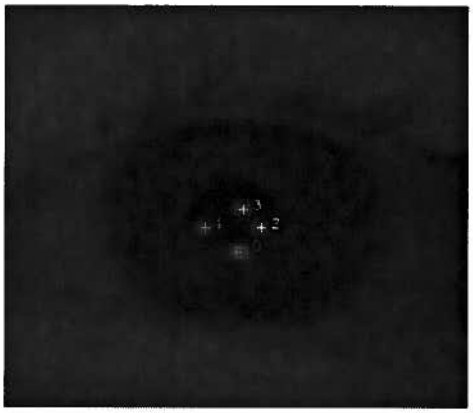
Image Feature Extraction

The image features required from the recorded images are the centers of the pupils and the locations of the corneal reflections. Infrared light is used for system illumination to enhance the performance of the feature extraction, using the bright-pupil and image-difference techniques [125] [126]. Using IR light generates the necessary reflections off the cornea, as well as reducing the sensitivity of the system to ambient lighting conditions. The image feature extraction procedure is described in greater detail in Hennessey *et al* [127].

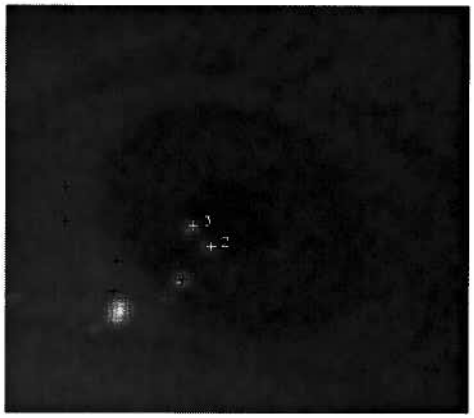
Corneal Reflection Pattern Matching

In a new approach to corneal reflection tracking, the off-axis light sources are used to generate a pattern of corneal reflections in the dark-pupil image. The corneal reflection pattern can then be used to enhance the performance of the POG estimation techniques. For the P-CR POG estimation method, a single corneal reflection is required for each eye, typically the on-axis corneal reflection. For the model-based method two corneal reflections, typically from multiple off-axis light sources, are required for triangulation of the 3D center of the cornea. Using three or more off-axis light sources to generate multiple corneal reflections can be used to provide redundancy should any reflection be corrupted or lost. Distortion or loss of corneal reflections occurs when the images of the corneal reflections lie near the boundary between the cornea and the sclera or on the sclera itself. The distortion of the reflections are due to the different radius of curvature between the sclera and the cornea, while the rougher surface of the sclera can cause valid reflections to disappear or spurious reflections to appear. A valid pattern of four corneal reflections are shown in Fig. 4.5(a) while the same pattern is shown corrupted in Fig. 4.5(b) due to eye rotation.

Using multiple corneal reflections requires a means for distinguishing the corneal reflection image points from one another, as the POG estimation methods require the correspondence between the light source and the generated reflection. Many general techniques for point pattern matching have been developed, for a literature survey see Cox and Jager [128]. The corneal reflection point-pattern matching technique described here is based on inter-point distances and is customized for corneal reflection detection. The algorithm compensates for translation, distortion, addition and deletion of corneal reflections. For proper operation, the IR point light sources must be placed such that at least two valid reflections off of the surface



(a) Valid corneal reflections



(b) Invalid corneal reflections

Figure 4.5: A set of four valid corneal reflections have been labled as shown in Fig. 4.5(a). In Fig. 4.5(b) the eye has been rotated, resulting in the loss of one of the valid corneal reflections, the distortion of another (labeled with a black cross) and the generation of a spurious reflection off the surface of the scelera.

of the cornea will always be visible to the camera, as a single reflection is insufficient for the matching procedure. In addition, unique displacements between all pairs of reflections are required to provide a means for matching the valid reflections with the corresponding IR point light sources.

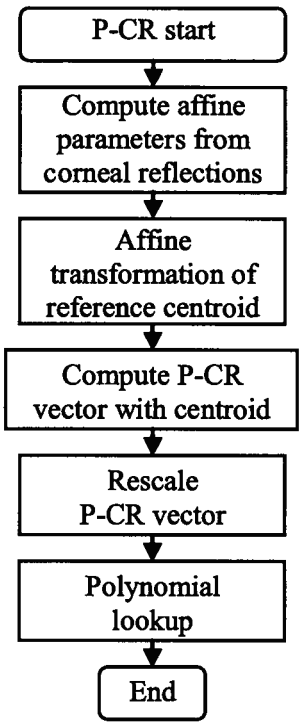
To perform the matching operation a reference pattern is required in which the valid corneal reflections are identified and associated with their corresponding IR point light sources as shown in Fig. 4.5(a). The reference pattern is created by recording a valid pattern of image reflections formed on each of the eyes and manually identifying the corresponding corneal reflections and IR light sources. Subsequent system operation extracts the coordinates of the corneal reflection image points (Q_i) and searches for matching pairwise displacements within the reference pattern points (R_i). A match is identified if a displacement is found under a certain tunable threshold value. This threshold is set to allow corneal reflections with slight distortions to pass, while larger distortions are rejected.

To reduce the pattern search space it was noted that the corneal reflection located closest to the pupil was least likely to be distorted on the boundary between the cornea and sclera. Accordingly the algorithm assumes that the corneal reflection image point located closest to the center of the pupil image will be valid. This image point is then used in each of the pairwise comparisons as described in Algorithm 1.

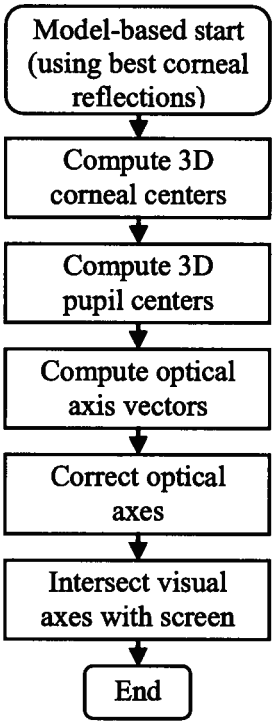
While the algorithm compensates for translation, distortion, addition and deletion of corneal reflections, it does not explicitly handle rotation or changes in scale between the reference and image point patterns. As the points are reflections off of a spherical surface, rotation of the image pattern should not be present. As well, by using the tunable threshold for the allowable distortion, the small changes in scale occurring due to changes in depth of the subject's eyes are accommodated.

4.2.2 POG estimation

The two main techniques used for POG estimation in remote eye-gaze tracking are the P-CR and model-based methods. The traditional P-CR and model-based methods have been enhanced to take advantage of the binocular eye-tracking and multiple redundant corneal reflections, enhancing both reliability and the ability to handle head motion. Each algorithm is outlined in Fig. 4.6 and will be described in the following subsections.



(a) Enhanced P-CR Method



(b) Model-Based Method

Figure 4.6: POG Estimation Stage. The processes are shown for the enhanced P-CR method in Fig. 4.6(a), while the processes for the model-based method are shown in Fig. 4.6(b). The enhanced P-CR method integrates the corneal reflection tracking for centroid estimation with the re-scaling of the P-CR vector to compensate for head motion. For the model-based method only the best available corneal reflections, as determined by the corneal reflection tracking, are used for the corneal center estimation.

Algorithm 1 Corneal reflection pattern matching

Input: P_{center} pupil center; Q_i, $i = 1..M$ image points; R_j, $j = 1..N$ reference points; $thresh$ distortion threshold

Output: Identified corresponding Q_i and R_j points

```
1: dist_min = inf
2: // Find index for closest image point Q_α to center of pupil
3: α = arg min_i ||Q_i − P_center||
4: // Identify corresponding image and reference points
5: for j = 1..N do
6:     // Translation from image to reference
7:     T_j = R_j − Q_α
8:     for i = 1..M, i ≠ α do
9:         for k = 1..N, k ≠ j do
10:            // Dist. from each image pt. to reference pt.
11:            d_k = ||(T_j + Q_i) − R_k||
12:            // Label Q_α at minimum overall dist.
13:            if d_k < dist_min then
14:                dist_min = d_k
15:                Label Q_α as R_j
16:            end if
17:        end for
18:        β = arg min_k{d_k}
19:        // Label Q_i if minimum dist. is under threshold
20:        if d_β ≤ thresh then
21:            Label Q_i as R_β
22:        end if
23:    end for
24: end for
```

Enhanced P-CR Vector

Traditionally the P-CR vector ($V = (v_x, v_y)$) is formed in the recorded bright-pupil image of the eye from the on-axis corneal reflection to the center of the pupil. Through a user calibration procedure in which the subject observes known points on the screen, the P-CR vector is mapped to the POG ($U = (u_x, u_y)$) on the screen in pixels. The mapping is usually a simple first order polynomial (4.1) where the parameters a_i and b_i are determined from calibration. A minimum of 4 calibration points are required to solve

for the 4 unknowns in each of the two separate equations.

$$\begin{aligned} u_x &= a_0 + a_1 v_x + a_2 v_y + a_3 v_x v_y \\ u_y &= b_0 + b_1 v_x + b_2 v_y + b_3 v_x v_y \end{aligned} \tag{4.1}$$

Using a single corneal reflection to create the P-CR vector can be problematic however, as the reflection may be distorted or lost during large eye rotations, as illustrated in Fig. 4.5(b). As well, after user calibration, if the head is translated in depth from the screen, the P-CR vector based on a single corneal reflection will appear to increase or decrease in scale, which would be interpreted as a change in POG position rather than just a change in head depth. To overcome these potential sources of error, the enhanced P-CR vector method uses the corneal reflections generated from multiple off-axis lights with the centroid of the corneal reflection pattern used to form the P-CR vector. The scale factor of the corneal reflection pattern can be determined and compared to the scale factor from calibration and used to re-scale the P-CR vector, reducing the effect of head motion.

The centroid of the 2D corneal reflection reference pattern R_c is first determined (4.2) as all valid 2D corneal reflection positions $R_i = (r_{ix}, r_{iy})$ are known. The 2D corneal reflection points $Q_i = (q_{ix}, q_{iy})$ are then extracted from the recorded images and matched with the reference points using Algorithm 1 and an affine transformation (4.3) is formed for the translation and scale at each point.

$$R_c = \frac{1}{N} \sum_{i=1..N} R_i \tag{4.2}$$

$$R_i = s \cdot Q_i + T \tag{4.3}$$

The scale (s) and 2D translation ($T = (t_x, t_y)$) parameters for the corneal reflection pattern can then be determined provided two or more valid image points are detected, resulting in an overdetermined set of equations for s, t_x and t_y (4.4). This equation is of the form $b = Ax$ where b and A are known and is easily solved using a least squares approach.

$$\begin{bmatrix} r_{1x} \\ r_{1y} \\ r_{2x} \\ r_{2y} \\ \vdots \end{bmatrix} = \begin{bmatrix} q_{1x} & 1 & 0 \\ q_{1y} & 0 & 1 \\ q_{2x} & 1 & 0 \\ q_{2y} & 0 & 1 \\ \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} s \\ t_x \\ t_y \end{bmatrix} \tag{4.4}$$

To compute a robust estimate of the corneal reflection pattern centroid Q_c at run-time, the original reference pattern centroid R_c is scaled and translated (4.5) according to the determined scale and translation factors. The 2D estimated centroid Q_c is then robust to loss or distortion of the corneal reflections which otherwise would have distorted the centroid calculation based on the remaining visible Q_i points alone.

$$Q_c = \frac{1}{s} \cdot (R_c - T) \tag{4.5}$$

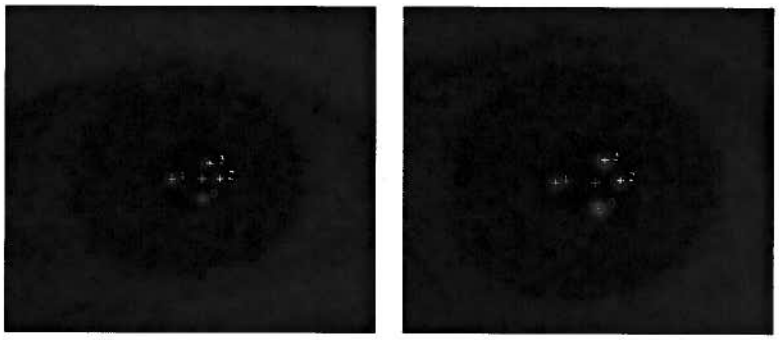
To accommodate translation of the head toward or away from the camera, resulting in changes in scale of the P-CR vector, the P-CR vector is rescaled based on the size of the corneal reflection pattern determined during user calibration. Using the ratio of the determined scale factor s and the calibration scale factor s_{cal} the 2D P-CR vector V is rescaled (4.6) where P_{center} denotes the center of the pupil image. The resulting P-CR vector is then robust to corneal reflection distortion, loss as well as depth translations of the eye as illustrated in Fig. 4.7.

$$V = \frac{s}{s_{cal}} \cdot (P_{center} - Q_c) \tag{4.6}$$

Model-Based

The second method of estimating the POG is based on 3D models of the eye and system as shown in Fig. 4.6. The model-based method for POG estimation was designed to inherently compensate for motion of the head, at the expense of an increasingly complex system configuration and algorithm. The model-based method requires 3D models of the camera and lens, computer screen and eye. In addition to the pupil image center, two corneal reflection images are also required to estimate the POG. The details of the model-based POG estimation procedure can be found in Hennessey *et al* [129].

The performance of the model-based POG estimation method was improved by using multiple corneal reflections to improve the quality and reliability of the corneal reflection image input data used in estimating the center of the cornea. The corneal reflection pattern tracking algorithm provides a means for tracking the valid reflections, and since only two reflections are required, only the corneal reflections least likely to be distorted are used. Since the distortion and loss of reflections occurs as the points approach the boundary of the cornea and scelera, the two reflections located closest to



(a) Centroid from reference pattern

(b) Centroid at near depth

(c) Centroid at far depth

(d) Centroid missing reflections

(e) Centroid missing reflections

(f) Centroid missing reflections

Figure 4.7: In the figures shown, the centroid maintains its position relative to the corneal reflection image points regardless of the scale, distortion, or loss of corneal reflections making up the pattern. In Fig. 4.7(b) and Fig. 4.7(c) the head was translated towards and away from the camera respectively. In Fig. 4.7(d) through Fig. 4.7(f) the centroid was correctly determined while up to two corneal reflection points were missing.

the center of the pupil are used in the estimation of the corneal center. This ensures that the most reliable and robust information is used in the POG estimation method.

Binocular Estimation

The P-CR and model-based POG estimation methods can be performed independently for the left and right eyes resulting in two POG estimates, one for each eye. As healthy eyes generally observe the same point in space, the left and right eye POG estimates should be located at the same position. In the event that one eye is located outside the field of view of the camera, or the POG for one eye is unable to be computed due to corrupt image features, the remaining valid POG can be used as the POG estimate. Additionally if both POG estimates are available, the average of the two can be determined, potentially reducing the overall error, as was observed by Cui *et al* for head mounted eye-gaze tracking [130].

4.3 Experimental Methods and Results

4.3.1 Experimental Hardware

The experiments were performed on the eye-gaze tracking system as shown in Figure 4.8. A single DragonFly Express camera from Point Grey Research is located below the computer screen and used to record images of the face and eyes. The single camera used had a sensor resolution of 640 x 480 pixels which streamed video over the Firewire 1394b data bus at 200 frames per second. An IR ring surrounds the camera lens and is used to generate the on-axis lighting. The off-axis light sources are comprised of six clusters of seven, 880 nm, IR LED's located around the computer screen. Only four of the six clusters were used to generate the off-axis corneal reflection pattern in the system presented here. A microcontroller is used to synchronize the camera shutter with the on and off-axis LED lighting. The computer screen is a 17" LCD with a resolution of 1280 x 1024 pixels. Extruded aluminum rails were used to create a mechanical mounting structure to which IR point light sources could be attached and the displacements between the lights, camera and screen fixed. For the model-based POG estimation method the camera lens was calibrated with the Matlab camera calibration toolbox while the physical locations of the camera, screen and LEDs were measured manually. For the P-CR method no system calibration was required. The computer used had a 2.66 GHz Intel Core 2 processor and 2 gigabytes of

RAM, and was capable of processing the single camera video stream at full frame rates.

4.3.2 Processing Time Evaluation

The frame rate at which the camera operated was 200 Hz, resulting in a time budget of 5 ms per image frame. The eye-gaze tracking algorithms were implemented in C++ and the average execution time required for each processing stage recorded as listed in Table 4.1. The recorded times were averaged over 1 second of operation and measured when both eyes were visible to the camera. Note that the sum of the sub-stages do not always equal the time required by the overall stage due to data logging and display processes used by the system.

With the high speed sampling rate of 200 Hz, filtering was used to smooth out noise from the system and the inherently jittery eye motions [127]. A rectangular FIR low pass filter (moving window average) with a filter order of 100 samples, or 0.5 seconds, was used to smooth the POG estimates. The filter was reset between fixations to prevent overlapping filter histories from merging data from two different fixations.

Table 4.1: Processing Times

Activity	Processing Time (ms)
Entire Process	2.5
Image Processing*	1.9
Face Tracking	0.25
Feature Tracking	1.4
Point Matching	0.022
POG Estimation	0.45
P-CR	0.15
Model-Based	0.30

* Image proc. time is common for both POG methods

4.3.3 Horizontal Motion Evaluation

Methods

Using binocular tracking increases the allowable head motion as the system can still operate if only one eye is visible. The face tracking system was used for distinguishing the left from right eye, when only a single eye was

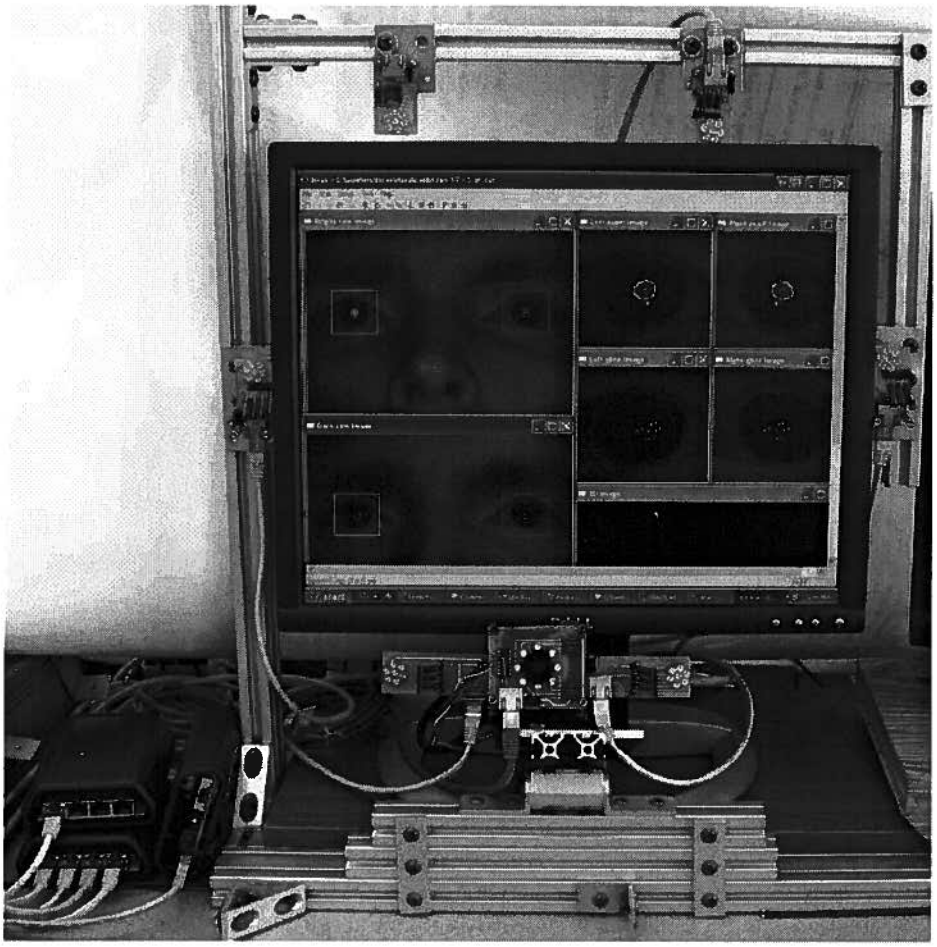


Figure 4.8: The eye-gaze tracking system is shown. The camera is located below the computer screen, with the camera lens surrounded by the ring of on-axis lighting. There are six off-axis point light sources located around the computer screen, of which four were used in the system presented here. The microcontroller used for synchronization of the on and off-axis lighting with the camera shutter is located in the lower left portion of the image.

visible. For the face tracking method used the average head width was required, which over the subjects tested in the work presented here was found to correspond to a camera image head width w_h set to 900 pixels. The increase in allowable horizontal head motion was measured for a single subject. The subject's head was initially located in a central position with both eyes visible for a five point user calibration of both the model-based and enhanced P-CR POG methods. The five points used were the four corners and the center point of the screen. The subject then performed an accuracy measurement at each of four laterally displaced head positions, similar to those shown in Fig. 4.4. At head position 1 the head was located to the extreme left with only the right eye visible at the left border of the camera image. At head positions 2 and 3 the head was located with both eyes visible, with the left eye located at the left border of the camera image at position 2, and the right eye located at the right border of the camera image at position 3. Finally at head position 4 the head was located to the extreme right with only the left eye visible at the right border of the camera image. At each head location the POG was computed with the model-based and enhanced P-CR POG estimation methods and recorded while the subject observed each of nine points located in a 3 x 3 grid across the computer screen. The model-based method was also used to determine an estimate for the location of the eye in 3D space for tracking the horizontal change in head position.

Results

Listed in Table 4.2 are the horizontal eye positions at each of the four head positions, measured from the world coordinate origin located at the lower left corner of the monitor. For this experiment the average eye to screen distance was 64 cm. Also shown in the table is the average POG estimation error on the 3 x 3 grid for each of the left, right and binocular (average of left and right POG) eyes for the model-based and enhanced P-CR POG methods. Note that since the binocular POG estimate is a 2D vector average of the left and right eye POG estimates, the magnitude of the resulting error for the binocular estimate can be lower than either the left or right eyes, as shown at head position 2 in Table 4.2.

Table 4.2: Eye position and average POG error over 3 x 3 screen grid (single subject)

Head Position	1	2	3	4
X coordinate (cm)				
Left eye	-	10.81	14.73	21.91
Right eye	11.18	17.90	21.85	-
Model-Based average POG error (cm)				
Left eye	-	1.01	0.63	1.35
Right eye	1.00	0.72	1.41	-
Binocular	-	0.71	0.64	-
Enhanced P-CR average POG error (cm)				
Left eye	-	1.30	1.01	2.04
Right eye	0.92	0.94	1.78	-
Binocular	-	0.77	1.05	-

- Not in view

4.3.4 Multi-subject Evaluation of Reliability and Accuracy

Methods

To analyze the performance of the system across a larger population sample, a multi-user experiment was evaluated on 10 different subjects. The subjects included 8 males and 2 females, with ages ranging from 24 to 31 years old. Two subjects wore contact lenses while the remaining had uncorrected vision. The ethnicity of the subjects was 5 Caucasian, 1 Hispanic, 3 Middle Eastern and 1 Indian.

The experiment was designed to provide a comparison of: 1) Reliability - the number of times any one corneal reflection was lost and the number of times the corneal reflection pattern centroid (requiring any two corneal reflections) was unable to be estimated, 2) POG method accuracy - the difference between the average accuracy of the traditional P-CR, enhanced P-CR using re-scaling, and the model-based method at three different head depths, and 3) Monocular *vs* binocular accuracy - the difference in average accuracy between the POG estimated by the left, right and average of the two eyes.

The experimental procedure had each test subject begin with the five point user calibration at the midpoint of the depth of focus of the camera lens, approximately 62 cm from the screen. After calibration, each subject

was asked to move his/her head towards the screen until just before the image features became too blurred to properly track the eyes, due to the limited depth of focus. At this point the extracted image features and the POG using each POG estimation method was recorded at each point on a 3 x 3 grid across the screen. The 9 point data collection procedure was repeated with the head located back at the middle of the depth of focus (roughly the original calibration position) and again with the head as far back as possible before the image features again became out of focus.

Results

The number of times each of the on-axis or off-axis corneal reflections were lost at each of the 9 points, at each of 3 depths, for the 10 subjects was determined from the recorded image feature data. The number of times that fewer than two valid off-axis corneal reflections were available, resulting in an inability to estimate the centroid, was also determined. The percentage of lost corneal reflections compared with the percentage of lost centroid positions (out of 270) was determined for each of the subject's left and right eyes and summarized in Table 4.3, where the off-axis corneal reflections are identified as labeled in Fig. 4.7(a). The 3D positions of the eyes were determined from the model-based method for POG estimation, which also provides estimates for the 3D position of the center of the cornea of each eye. The average eye depth from eye to screen over the 10 subjects for the close position was 58 cm, for the middle position 62 cm and for the far position 66 cm.

Table 4.3: Corneal reflection loss for each eye as a percentage of total possible at three head depths.

Corneal	Close (%)		Middle (%)		Far (%)	
Reflection	L	R	L	R	L	R
Off-axis (0)	7	28	2	9	7	14
Off-axis (1)	7	3	6	1	10	2
Off-axis (2)	14	12	10	4	10	8
Off-axis (3)	14	14	2	4	4	6
On-axis	1	6	0	2	0	1
Centroid loss	0	0	0	0	0	0

At each test position at each depth the POG was estimated for both the left and right eyes using each of the three POG estimation algorithms,

traditional P-CR, enhanced P-CR and model-based. Operating the POG estimation algorithms on the same recorded images allows for a direct comparison between the performance of the different methods. The error averaged over the 10 subjects is shown in Table 4.4. In addition to the average POG error from the left and right eyes, the binocular POG error is also shown. The average POG accuracy can be converted from centimeters to degrees of visual angle given the depths of the eyes. For the enhanced P-CR and model-based methods an accuracy of 0.71 cm was achieved at the middle position using the binocular average of the left and right eye POG, corresponding to a visual angle accuracy of 0.66°.

Table 4.4: Average error from monocular and binocular data

	Average Error (cm)		
Method	**Left**	**Right**	**Binocular**
Close Position (58 cm)			
Trad. P-CR	2.79	3.07	2.77
Enha. P-CR	1.00	1.52	1.01
Model-Based	1.03	1.11	0.91
Middle Position (62 cm)			
Trad. P-CR	1.00	1.01	0.80
Enha. P-CR	0.95	0.97	0.71
Model-Based	0.85	0.93	0.71
Far Position (66 cm)			
Trad. P-CR	2.59	2.29	2.28
Enha. P-CR	1.39	1.14	0.97
Model-Based	1.30	1.02	0.98

4.4 Discussion

When the eye is rotated to view different points on the screen, the corneal reflections will translate across the surface of the cornea. In remote eye-gaze tracking, translation of the head also results in corneal reflection translations, unlike head mounted systems where the head to camera displacements are fixed. At certain orientations of the eye and head with respect to the camera, the corneal reflections may be blocked by eyelashes, distorted on the boundary between the cornea and scelera, or lost on the rougher surface of the scelera due to diffuse reflection. This was studied in an experiment

in which the corneal reflections were tracked over 270 different head and eye positions as listed in Table 4.3. As seen from the table, over the 10 subjects tested, if only a single on-axis or off-axis corneal reflection was used to form the P-CR vector the system would have been unable to determine the POG up to 6% or 14% of the time respectively. Using the centroid, as determined by equations (4.2) through (4.5), to form the P-CR vector however, resulted in a valid POG estimate for all head positions and eye rotations tested. Consequently the use of multiple redundant corneal reflections results in a more reliable system for POG estimation in which head motion is allowed.

Tracking the corneal reflections provides an estimate of the scale and translation of the corneal reflection pattern. In the multi-subject experiment, the traditional P-CR, enhanced P-CR and model-based methods were each used to estimate the POG at the same time, using the same source image data, to compare the accuracy of the three POG estimation methods. The average error shown in Table 4.4 for the binocular (averaged) eyes using each of the three POG methods was compared using an analysis of variance (ANOVA) at each of the three depths tested.

At the close distance ($F(2,267) = 85.27$, $p<0.001$) and far distance ($F(2,267) = 45.83$, $p<0.001$) a statistically significant difference was found between the POG estimation methods. A Bonferroni *post-hoc* analysis showed that the statistically significant difference (at the 0.05 level) was between the traditional P-CR method and both the enhanced P-CR and model-based methods. The ability of the enhanced P-CR method to handle changes in head depth was shown to improve to match that of the model-based method as no statistically significant difference was observed between the two methods. The average POG accuracy of the enhanced P-CR method improved for the binocular result by over 2.3 times when compared with the traditional P-CR method, from 2.28 cm to 0.97 cm at the far distance and from 2.77 cm to 1.01 cm at the close head distance.

At the middle depth, no statistically significant differences between the accuracies of the POG estimation techniques were found ($F(2,267) = 1.26$, $p=0.286$). No improvement of the enhanced P-CR method over the traditional method at the middle depth was expected however, as the user calibration was originally performed at approximately the same depth. Overall POG estimation accuracies as good as 0.71 cm or 0.66° of visual angle were observed with the enhanced P-CR and model-based POG estimation methods at the middle depth.

The addition of face tracking provided the ability to distinguish between the left and right eyes when only a single eye was visible to the camera. The ability to track either eye increased the allowable horizontal head motion

while still maintaining an estimate for the subject's POG.

An experiment was performed on a single subject in which the POG was tracked for both eyes using the enhanced P-CR and model-based method. The horizontal coordinate for each eye, as determined by the model-based method, as well as the average POG accuracy was recorded as listed in Table 4.2. Given the resolution of the camera sensor and the spatial resolution required for image feature extraction, the allowable horizontal motion of the head while tracking both eyes was only 4 cm. If both eyes are tracked with face tracking used for distinguishing the left from right eye, the range of horizontal motion increases further to 18 cm (based on an interpupillary distance of 7.1 cm from head positions 2 and 3).

Increasing the allowable horizontal head motion increases the usability of the system as users are not required to control their heads as carefully during operation. Although a larger field of view is possible using a pan/tilt/zoom mechanism, the benefit of the system presented here is that the slower mechanical tracking is avoided as the eyes are tracked within the recorded images at high speeds. To increase the allowable headspace of a single camera eye-gaze tracking system a camera with a higher sensor resolution can be used, allowing a decrease in camera lens focal length and therefore an increase in the horizontal and vertical field of view, for the equivalent spatial resolution.

The average error for the enhanced P-CR and model-based method was also recorded as shown in Table 4.2. An ANOVA was performed for each of the four laterally displaced head positions tested, comparing the enhanced P-CR with the model-based POG estimation methods. Under typical system operation the binocular POG estimate would be used for the comparison between methods, as at head position 2 ($F(1,16) = 0.05$, $p=0.829$) and head position 3 ($F(1,16) = 4.00$, $p=0.063$). At head position 1 only the right eye POG estimate is available for comparison ($F(1,16) = 0.10$, $p=0.751$) and the left eye POG at head position 4 ($F(1,16) = 3.41$, $p=0.084$). No statistically significant difference was found between the enhanced P-CR and model-based method at any of the horizontally displaced head positions.

High speed image feature tracking was achieved with image processing routines designed to utilize a minimal amount of processing power. In the system presented here the entire processing loop for the single video stream required only 2.5 ms, of which 1.9 ms was used for image processing and 0.45 ms was used for POG estimation. The face tracking system required only 0.25 ms while the corneal-reflection pattern matching algorithm required only 0.022 ms. Given the processing requirements, the system was capable of maintaining operation at the 200 Hz camera frame rate.

Tracking both eyes increased the reliability of the remote eye-gaze tracking system by increasing the range of head motion and allowed for the loss of a single eye. Averaging of the left and right eye POG estimates can potentially be used to increase the overall system accuracy. For the remote eye-gaze tracking system presented here, the POG accuracy results of the three POG estimation methods at the middle depth shown in Table 4.4 were analyzed with an ANOVA comparing the average error of the left, right and binocular estimates. For both the traditional ($F(2,267) = 4.32$, $p=0.014$) and enhanced P-CR ($F(2,267) = 7.72$, $p=0.001$) methods, the binocular POG estimation accuracy was statistically better (at the 0.05 level) than the POG accuracy of both the left or the right eyes alone. For the model-based method ($F(2,267) = 3.48$, $p=0.032$) at the middle depth, the binocular POG accuracy was found to be statistically better than the right eye POG accuracy while no difference was found with the left eye. Binocular tracking with averaging of the left and right eye POG estimates in remote eye-gaze tracking therefore equals or improves on the accuracy of monocular tracking alone.

4.5 Conclusions

Remote eye-gaze tracking requires the ability to handle both head and eye motion since the camera-to-eye displacement is not fixed as it is with head mounted systems. With head motion comes the potential of positioning the head such that an eye lies outside of the stationary field of view of the eye-tracking camera. The eyes may also be translated with respect to the camera by head movement, with key image features such as the corneal reflections becoming occluded by eye lashes or distorted on the boundary between the cornea and scelera. A corneal reflection pattern-matching algorithm detected lost and distorted corneal reflections allowing POG estimation with greater reliability than using one or two corneal reflections alone.

A centroid estimation technique allowed for more robust detection of the P-CR vector with rescaling of the enhanced P-CR vector used to compensate for depth translations of the head, improving accuracy by over 2.3 times when compared with the traditional method. In both horizontal and depth translations of the head it was shown that the performance of the enhanced P-CR method matched that of the model-based method, while avoiding the need for complex system calibration.

With the high speed face tracking system described in this chapter the loss of an eye from the field of view does not prevent the estimation of the

POG from the remaining eye as the left and right eyes can be distinguished based on their relative displacements in the face. For the single camera used in the system presented here, an increase in horizontal head motion to 18 cm was achieved when compared with the 4 cm of horizontal motion when both eyes had to remain in view.

It was also shown that over 10 different subjects, binocular averaging of the left and right eye POG estimates resulted in an accuracy that was statistically equal to or better than the monocular performance for the traditional P-CR, enhanced P-CR and model-based POG estimation methods.

For system users who have difficulty maintaining a relatively fixed head position the ability to handle head motion is a key usability factor in eye-gaze tracking. In the system presented here, the range of allowable head motion was increased and the accuracy and reliability of tracking improved using a combination of multiple corneal reflections and binocular eye-gaze tracking. Using the techniques presented, the enhanced P-CR, system-calibration-free, POG estimation method was shown to improve to match the performance of the more complex model-based method requiring system calibration.

Acknowledgments

The authors would like to express their appreciation for the support of the Natural Sciences and Engineering Research Council of Canada (NSERC) Chair in Design Engineering, and NSERC Discovery Grant #4924.

References

[108] A. T. Duchowski, *Eye Tracking Methodology: Theory and Practice.* Springer-Verlag, 2003.

[109] C. H. Morimoto and M. R. M. Mimica, "Eye gaze tracking techniques for interactive applications," *Comput. Vis. Image Underst.*, vol. 98, no. 1, pp. 4–24, 2005.

[110] T. Hutchinson, J. White, W. Martin, K. Reichert, and L. Frey, "Human-computer interaction using eye-gaze input," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 19, no. 6, pp. 1527–1534, 1989.

[111] H. Hua, P. Krishnaswamy, and J. P. Rolland, "Video-based eyetracking methods and algorithms in head-mounted displays," *Opt. Express*, vol. 14, no. 10, pp. 4328–4350, 2006.

[112] S. K. Schnipke and M. W. Todd, "Trials and tribulations of using an eye-tracking system," in *CHI '00 extended abstracts on Human factors in computing systems.* New York, NY, USA: ACM Press, 2000, pp. 273–274.

[113] R. Jacob and K. Karn, *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research.* Amsterdam: Elsevier Science, 2003, ch. Eye Tracking in Human-Computer Interaction and Usability Research: Ready to Deliver the Promises (Section Commentary), pp. 573–605.

[114] J. J. Cerrolaza, A. Villanueva, and R. Cabeza, "Taxonomic study of polynomial regressions applied to the calibration of video-oculographic systems," in *Proceedings of the 2008 symposium on Eye tracking research & applications.* New York, NY, USA: ACM, 2008, pp. 259–266.

[115] S.-W. Shih and J. Liu, "A novel approach to 3-d gaze tracking using stereo cameras," *IEEE Transactions on Systems, Man and Cybernetics, Part B*, vol. 34, no. 1, pp. 234–245, Feb. 2004.

[116] R. Tsai, "A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses," *IEEE Journal of Robotics and Automation*, vol. 3, no. 4, pp. 323–344, Aug 1987.

[117] J. Heikkilä and O. Silvén, "A four-step camera calibration procedure with implicit image correction," p. 1106, 1997.

[118] T. Ohno and N. Mukawa, "A free-head, simple calibration, gaze tracking system that enables gaze-based interaction," in *Proceedings of the 2004 symposium on Eye tracking research & applications*. New York, NY, USA: ACM Press, 2004, pp. 115–122.

[119] D. Beymer and M. Flickner, "Eye gaze tracking using an active stereo head," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, 18-20 June 2003, pp. II–451–II–458.

[120] D. H. Yoo and M. J. Chung, "A novel non-intrusive eye gaze estimation using cross-ratio under large head motion," *Comput. Vis. Image Underst.*, vol. 98, no. 1, pp. 25–51, 2005.

[121] R. J. K. Jacob, *Eye Movement-Based Human-Computer Interaction Techniques: Toward Non-Command Interfaces*. Norwood, N.J.: Ablex Publishing Co., 1993, vol. 4, pp. 151–190.

[122] A. Fitzgibbon, M. Pilu, and R. Fisher, "Direct least square fitting of ellipses," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 5, pp. 476–480, May 1999.

[123] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Comput. Surv.*, vol. 35, no. 4, pp. 399–458, 2003.

[124] M. C. Santana, "On real-time face detection in video streams. an opportunistic approach." Ph.D. dissertation, Universidad de Las Palmas de Gran Canaria, March 2003.

[125] Y. Ebisawa and S. Satoh, "Effectiveness of pupil area detection technique using two light sources and image difference method," in *Proceedings of the 15th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Oct 28-31, 1993, pp. 1268–1269.

[126] C. H. Morimoto, D. Koons, A. Amir, and M. Flickner, "Pupil detection and tracking using multiple light sources." *Image and Vision Computing*, vol. 18, no. 4, pp. 331–335, 2000.

[127] C. Hennessey, B. Noureddin, and P. Lawrence, "Fixation precision in high-speed noncontact eye-gaze tracking," *IEEE Transactions on Systems, Man and Cybernetics, Part B*, vol. 38, no. 2, pp. 289–298, April 2008.

[128] G. Cox and G. de Jager, "A survey of point pattern matching techniques and a new approach to point pattern recognition," in *Proceedings of the 1992 South African Symposium on Communications and Signal Processing*, 11 Sept. 1992, pp. 243–248.

[129] C. Hennessey, B. Noureddin, and P. Lawrence, "A single camera eye-gaze tracking system with free head motion," in *Proceedings of the 2006 symposium on Eye tracking research & applications.* New York, NY, USA: ACM Press, 2006, pp. 87–94.

[130] Y. Cui and J. M. Hondzinski, "Gaze tracking accuracy in humans: two eyes are better than one." *Neuroscience Letters*, vol. 396, no. 3, pp. 257–262, Apr 2006.

Chapter 5

Non-Contact Binocular Eye-Gaze Tracking for Point-of-Gaze Estimation in Three Dimensions [4]

5.1 Introduction

The point of conscious attention of an individual can be used to provide insight into cognitive processes - information that may otherwise be difficult to obtain [131]. Eye movements, and the resulting point-of-gaze (POG) of a subject can be estimated automatically with an eye-gaze tracker. With the real-time capabilities of modern eye-gaze tracking systems the use of eye-gaze has expanded from a diagnostic tool to applications in which the POG is used for control as well [132].

Two dimensional (2D) displays are currently the standard method of visual display used with eye-gaze trackers. Considerable progress however has been made towards the development of stereoscopic, or three dimensional (3D) displays [133]. In addition to enhancing the realism of the viewing experience, 3D displays can be used to more readily view complex volumetric data sets in medical imaging (magnetic resonance and computed tomography for example), 3D computer-aided design, and telesurgery. Furthermore, autostereoscopic displays which do not require any contact with the viewers face have been developed [134; 135].

The ability to determine a user's POG in 3D space will become increasingly important as the use of 3D displays become more widespread. The current methods for 3D interaction typically use an electromagnetic or op-

[4] A version of this chapter has been accepted for publication and is currently in revisions. Hennessey, C., and Lawrence, P. 2008. Non-Contact Binocular Eye-Gaze Tracking for Point-of-Gaze Estimation in Three Dimensions. IEEE Transactions on Biomedical Engineering

tically tracked stylus held by the user in 3D space against gravity [136]. Using the 3D POG for 3D interaction avoids the visual disconnect when the tracked tool cannot be physically located within the environment in which it is supposed to be acting [137]. The 3D POG also requires no physical effort other than directing the gaze to the point of interest. In addition to interaction with 3D displays, the 3D POG can be used to provide a means for interaction in real world 3D spaces using only the eyes. This could be an important advance for individuals with restricted mobility such as those with high level spinal cord injuries or advanced degenerative motor neuron diseases.

Limitations of existing eye-gaze tracking systems are application dependent. In research or clinical studies of eye movement, some inconveniences (e.g. head mounted equipment, long calibration processes) may be acceptable. For other users of a system, including the general public, the same deficiencies in usability may not be acceptable. A number of significant limitations for 2D eye-gaze tracking have been listed by Morimoto *et al* [138], difficulties which are further compounded when extending from 2D to 3D. Some of these limitations include low accuracy, low sampling rates, poor precision, complex and lengthy calibrations and uncomfortable user requirements including the need to wear the system on the users head, or to maintain a fixed head position. The usability of modern eye-gaze tracking systems may be a major reason why they are most commonly found in research based environments or specialized applications and are not widely used by the general population.

One of the areas targeted for improvement has been on increasing the usability of eye-gaze trackers with the transition from head mounted to remote eye-gaze tracking [138], which mirrors the transition to autostereoscopic displays for improving the usability of 3D displays. Head mounted systems are well suited to eye-gaze tracking applications involving user mobility such as walking or active sports [139], however, users may be averse to wearing headgear in everyday computer use. In addition, slippage of the head gear can result in increased error or require recalibration. In applications where the subject is seated, eye-gaze trackers based on remote image recording can enhance the user experience by requiring no contact with the subject's face or head.

There are two main image based techniques for estimating the POG, the Pupil-Corneal Reflection (P-CR) method [140] and methods based on models of the eye and system [141; 142; 143]. The P-CR method uses the vector formed from a reflection generated off the surface of the cornea and the center of the pupil, along with a polynomial mapping (determined

through calibration [138] [144]) to determine the POG on a 2D surface such as a computer screen. The P-CR method is well suited to head mounted applications in which the distance from the eye to camera changes little, as the accuracy of the estimated POG has been shown to degrade when head motion is coupled with eye motion [138]. Model-based methods are designed to avoid the degradation in POG accuracy as head motion is implicitly compensated. With the model-based methods the image features are used to determine the position of the eyes in 3D space, the visual axis along which the user is looking, and the POG at the intersection of the visual axis and the surface of interest.

One of the first systems developed to investigate 3D POG estimation was presented by Duchowski *et al* [145] for use in a 3D virtual reality environment. A commercial Head Mounted Display (HMD) was used to provide disparity images to the left and right eyes. A commercial, binocular, head mounted eye-gaze tracker using the P-CR method for POG estimation was integrated with the HMD to determine the user's 2D POG on the left and right HMD screens. In addition to the eye-gaze tracker, an electro-magnetic tracker was attached to the head mounted apparatus to determine head position and orientation. Two stages of per-user calibration were required, the first to calibrate the eye-gaze tracker on the HMD and the second to provide estimates for the geometric parameters such as the interpupillary distance (the distance between the eyes) and the distance from the eyes to the surface of the HMD screens. Standard stereo geometry techniques [146] were then used to estimate the 3D POG based on the head pose and 2D POG estimates.

The 3D POG estimation system developed by Essig *et al* also used a binocular P-CR based head mounted eye-gaze tracker, however a neural network was used to generate the 3D POG estimates [147]. The 2D POG estimates were tracked on a remote desktop monitor and used as input to a neural network which then estimated the 3D POG. In their original work the 2D computer display used single image random dot stereograms to provide the virtual 3D display while in their later work anaglyph images were used [148]. Two stages of calibration were required, the first to calibrate the eye-gaze tracker on the desktop display and the second to train the neural network.

The system recently developed by Munn and Pelz [149] for 3D POG estimation again used a P-CR based head mounted eye-gaze tracker, however, only a single eye was used for their method. A head mounted scene camera was used to record a 2D projection of the subject's scene view, upon which the monocular 2D POG estimates were tracked. With sufficient head motion

the monocular visual axis vectors over time were intersected to determine the 3D POG, provided the head mounted camera position and orientation were also accurately tracked in 3D space.

A novel binocular system by Kwon *et al* [150] estimated the 3D POG in a virtual 3D environment with a 2D parallax barrier display. The P-CR method was used to determine eye-gaze direction, along with the relative displacements of the binocular pupil centers to estimate the depth of the 3D POG. This technique required a fixed head to camera displacement which was achieved using a chin rest.

The system we propose for 3D POG estimation follows the design goal of improving the usability of eye-gaze tracking with no contact required and no equipment mounted on the user's head. Our system uses a model-based method for estimating the 3D POG, which allows for head motion and does not require fixing the position or orientation of the head with a chin rest. The model-based method uses image features directly and avoids the intermediate stage of 2D POG estimation on a 2D surface, simplifying the per user calibration to a single stage. The system we propose also estimates the POG in a 3D real world volume in real-time and does not require large head motions as binocular eye-tracking is employed.

To the best of the authors' knowledge this chapter has three original contributions. The first is the design of the first reported binocular system for estimating the absolute X, Y, Z coordinates of where one is looking in the real 3D world. Secondly, this is the first system that uses a model-based method for 3D POG estimation and therefore requires only a single per-user calibration stage. Finally, it is the first non-contact, head-free 3D POG eye-gaze tracking system to be reported and/or evaluated in the literature. With no attachments to the user's head or use of chin rests to fix the position of the head, the system permits eye and head motions within the field of view of the camera.

5.2 Methods

The proposed system for non-contact 3D POG estimation is comprised of an image processing stage for extracting image features, a model fitting stage for computing the corneal centers and optical axes of the eyes and finally a model-based vergence algorithm for computing the 3D POG. A single per-user calibration stage is used to correct the eye models for between-subject differences.

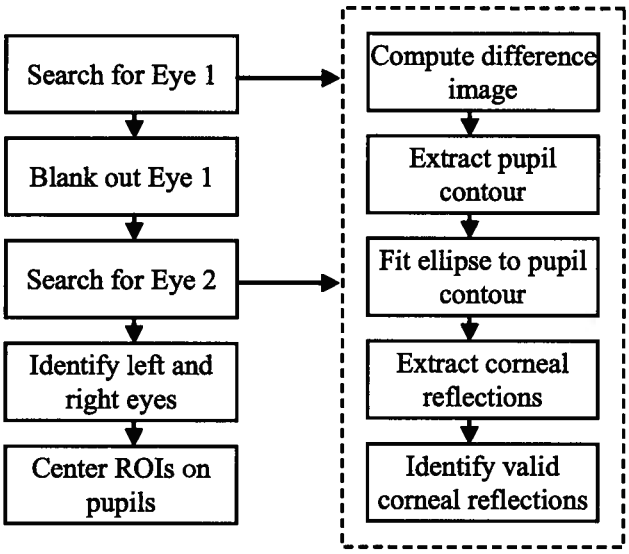
5.2.1 Image processing

The model-based POG estimation method requires accurately identified image features of both eyes from the recorded images as described in [151]. To estimate the 3D position of the cornea, the image locations of two corneal reflections are required. To determine the direction of the optical axis, the image location of the center of the pupil is required, in addition to the previously computed 3D center of the cornea. An outline of the image processing procedure is shown in Fig. 5.1 in which Fig. 5.1(a) illustrates the overall binocular tracking and Fig. 5.1(b) illustrates the image processing steps for each eye. In the event that less than two eyes are detected the system will not be able to estimate the 3D POG from vergence and the processing halts until the next image frame is recorded.

To aid the pupil tracking algorithm, the bright pupil and image differencing techniques are used to create a high contrast image of the pupil [152; 153]. The bright pupil image is taken using a light source located coaxially with the lens of the camera which results in a brightly illuminated pupil due to the retro-reflective property of the retina (the same phenomenon as red-eye in flash photography). The dark pupil image is formed by using off-axis lighting, which illuminates the face equivalently but does not generate a bright pupil. The difference image formed by subtracting the dark pupil image from the bright pupil image results in a high contrast pupil contour which is easily segmented. The roughly identified difference image is then used to identify the pupil contour in the bright pupil image [151]. Once the pupil contour has been identified, an ellipse is fit to the perimeter and the center of the ellipse is used as the center of the pupil [154].

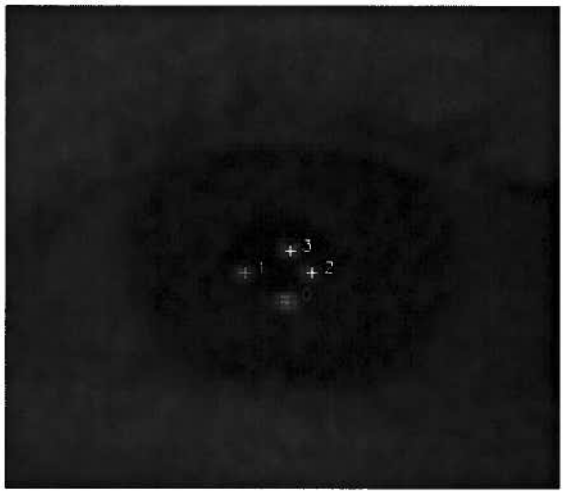
The corneal reflections are found by searching the dark pupil image for high intensity image pixels located in close proximity to the identified pupil. Significant rotation of the eyes with respect to the camera, commonly occurring in 3D POG estimation, can cause the corneal reflections to appear distorted near the boundary between the cornea and scelera, or disappear completely on the rougher surface of the scelera [155]. While the location of only two corneal reflections are required for triangulation of the location of the cornea, in the system described here a set of four off-axis light sources were used to generate four corneal reflections for redundancy. Point pattern matching is used to match a reference pattern of known valid corneal reflections, shown in Fig 5.2(a), with the remaining visible corneal reflections, shown in Fig 5.2(b) [156]. The reference pattern is formed based on the relative positions of the off-axis light sources.

To achieve the desired high speed sampling rates needed for digital fil-

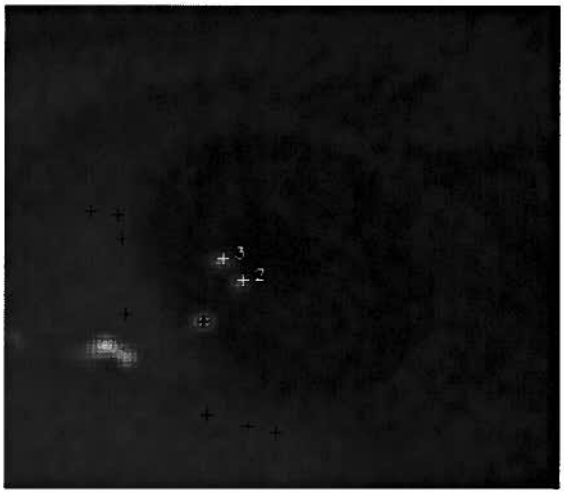


(a) Image Processing Procedure (b) 'Search for Eye N' Procedure

Figure 5.1: The overall image processing loop is shown in Fig. 5.1(a). The search for the eyes is performed sequentially and only after both eyes have been detected are they identified as either left or right. When the ROIs are applied the image search space is greatly reduced. In Fig. 5.1(b) the procedure for identifying the image features required for the next stage of model fitting is presented.



(a) Reference Pattern



(b) Pattern Matching

Figure 5.2: An example of the four valid corneal reflections used as the reference pattern is shown in Fig. 5.2(a). With the large eye rotation shown in Fig. 5.2(b) some of the corneal reflections were distorted or lost, however, two valid corneal reflections remain, which is sufficient for POG estimation.

tering, the amount of image information to process per system loop is significantly reduced by only processing the ROI's as opposed to the full image as described in [157]. To track the motion of the eyes within the recorded images, the left and right eye ROI's are continuously repositioned onto the left and right pupil image centers respectively. If either eye is lost, due to blinking or eye placement outside the field of view of the camera, the ROI's are resized to the full image. The full image ROI's are processed until each eye is re-acquired, after which the ROI's are reduced to encompass just the eyes, and high speed processing resumes.

5.2.2 Model Fitting

The model fitting algorithm uses the extracted image features, along with models of the physical system, camera and eye to estimate the 3D center of the cornea C, pupil P_c and ultimately the optical axis vector joining these two points as shown in Fig. 5.3. The 3D location of the center of the cornea is determined by a triangulation method using the images of two corneal reflections. With the known position and radius of the cornea, the 3D pupil center is found using ray-tracing from the pupil image center on the camera sensor, accounting for refraction at the surface of the cornea. The details of the 3D cornea and pupil center estimation technique have been previously described in Hennessey *et al* [151] which are an extension of earlier work by Shih and Liu [141].

The model of the physical system used here is determined through direct measurement of the locations of the camera and infrared (IR) point light sources. The camera lens is modeled as a pin-hole with the intrinsic parameters estimated using the Camera Calibration Toolbox for MATLAB [158]. The model of the eye used here is based on the schematic eye developed by Gullstrand [159] as illustrated in Fig. 5.3. In the simplified schematic eye the cornea is approximated as a uniformly spherical surface with three parameters (r, r_d, n). The three parameters vary between subjects, however, to date there has been no known method for estimating them on a per user basis based purely on remote imaging the eyes and consequently population averages are typically used. An error analysis based on the effects of these assumptions are reported in Section 5.3.6 where it is clear that system accuracy could benefit from future development of a non-contact method for estimating each of these parameters.

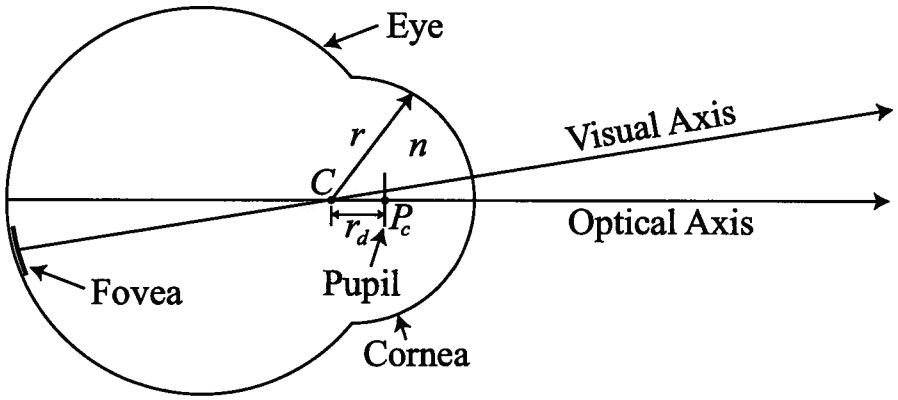


Figure 5.3: The schematic eye includes three general parameters; the radius of the corneal sphere r, the distance from the center of the corneal sphere to the center of the pupil r_d and the index of refraction n of the aqueous humor fluid. The model-based method for computing the POG is based on first determining the location of the center of the cornea. With the location of the corneal center it is then possible to compute the optical axis direction. The optical axis vector is corrected through calibration to lie along the visual axis, which is offset from the optical axis due to the displacement of the fovea on the retina.

5.2.3 Calibration

In the model fitting procedure outlined here, the optical axis can be determined based on the simplified eye model, however, the true visual axis may lie up to 5° from the optical axis depending on the location of the fovea (high resolution portion of the retina) for an individual user [160]. The offset between the optical axis and the visual axis can be compensated with a per-user calibration.

The per user calibration procedure involves having a user observe known points in 3D space while the optical axes of the eyes are computed and the offsets required to intersect the optical axes with the test positions are determined. For 2D POG estimation using the model-based method, the test points are located on the surface of the display [141] [142]. For POG estimation in 3D, the test point can be located anywhere within the workspace volume. While a single calibration point is sufficient to determine the angular offsets, multiple calibration points located throughout the workspace display (or volume for 3D) are typically used.

For each of the N calibration test positions T_i as shown in Fig. 5.4(a), each optical axis OA_i is normalized and converted to spherical coordinates (5.1) where ϕ_i and θ_i are readily determined.

$$\left[\widehat{OA_i}\right] = \frac{[OA_i]}{\|[OA_i]\|} = \begin{bmatrix} \sin\phi_i \cos\theta_i \\ \sin\phi_i \sin\theta_i \\ \cos\phi_i \end{bmatrix} \tag{5.1}$$

The angular offset corrections $\Delta\phi_i$ and $\Delta\theta_i$, between the optical axis and the visual axis, can be determined using the parametric equation of a line (5.2) with 3 equations and 3 unknowns (t, $\Delta\phi_i$, and $\Delta\theta_i$) which can be solved for explicitly.

$$T_i = C_i + t \cdot \begin{bmatrix} \sin(\phi_i + \Delta\phi_i)\cos(\theta_i + \Delta\theta_i) \\ \sin(\phi_i + \Delta\phi_i)\sin(\theta_i + \Delta\theta_i) \\ \cos(\phi_i + \Delta\phi_i) \end{bmatrix} \tag{5.2}$$

All subsequent estimated optical axis vectors OA_{curr} are normalized and corrected to the visual axis VA_{curr} using proportional weighting of the calibration parameters. The similarity between the current normalized optical axis vector $\widehat{OA}_{curr}$ and each calibration optical axis vector $\widehat{OA}_i$ as determined by the Euclidean distance (5.3), is used to generate a list of weighting factors (5.4) which are then used to weight the angular offsets $\Delta\phi_i$, and $\Delta\theta_i$

determined during calibration.

$$d_i = \left\| \widehat{OA}_{curr} - \widehat{OA}_i \right\| \tag{5.3}$$

$$w_i = \frac{1}{d_i \cdot \sum_{k=1..N} \frac{1}{d_k}} \tag{5.4}$$

The normalized optical axis $\widehat{OA}_{curr}$ is converted to spherical coordinates ϕ_{curr} and θ_{curr}, the weighted sum of the corrections applied to the spherical coordinates (5.5) and (5.6) and the resulting visual axis determined (5.7) as shown in Fig. 5.4(b).

$$\phi'_{curr} = \phi_{curr} + \sum_i w_i \cdot \Delta\phi_i \tag{5.5}$$

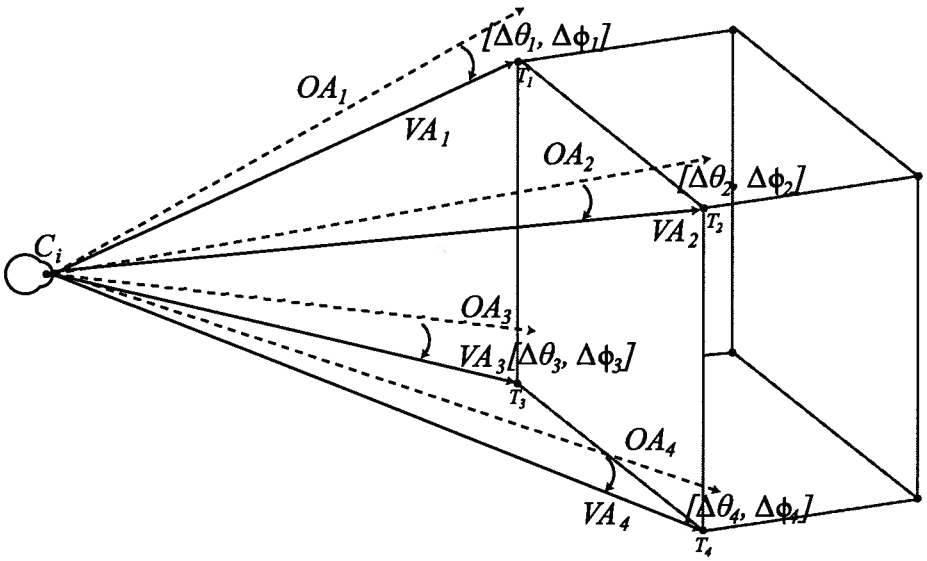
$$\theta'_{curr} = \theta_{curr} + \sum_i w_i \cdot \Delta\theta_i \tag{5.6}$$

$$VA_{curr} = \begin{bmatrix} \sin(\phi'_{curr}) \cos(\theta'_{curr}) \\ \sin(\phi'_{curr}) \sin(\theta'_{curr}) \\ \cos(\phi'_{curr}) \end{bmatrix} \tag{5.7}$$

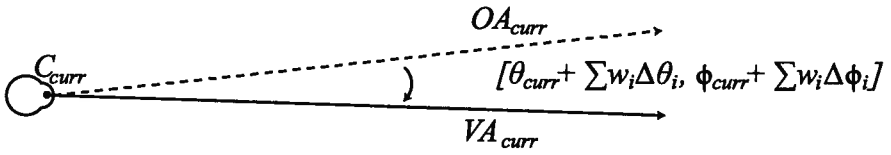
In reality the angular offsets of the eyes do not change depending on gaze direction and a single calibration point should be sufficient. However, as will be shown in the calibration experiment in Section 5.3.4, using multiple calibration positions and the proportional weighting technique proposed here provides an improvement in overall accuracy. This is due to the additional errors introduced from using a simplified eye model and population averages for the eye model parameters as discussed in Section 5.3.6. As the model of the eye is refined and techniques for determining the per-user eye model parameters are developed, it would be expected that the proportional weighting calibration procedure would then simplify to a single point calibration.

5.2.4 Model-Based Vergence

With 2D model-based eye-gaze tracking, the visual axis is traced from the center of the cornea into the 3D world and intersected with an object of



(a) Four point calibration in 3D space at a single depth plane



(b) Calibration correction of optical axis to visual axis

Figure 5.4: The calibration procedure uses calibration test positions located throughout the workspace volume. In Fig. 5.4(a) the calibration positions T_i are shown at a single depth plane for simplicity of the figure. The calibration corrections $\Delta\phi_i$ and $\Delta\theta_i$ are used to reorient all future optical axis vectors to the visual axis as shown in Fig. 5.4(b).

known geometry. This object is typically the planar surface of a desktop monitor but may be any surface, provided the location and geometry are known *a priori*. To estimate the POG in 3D space without *a priori* knowledge of the surfaces upon which the user is looking, the binocular visual axis vectors are traced from their respective corneal centers and intersected in free space. A flowchart illustrating the 3D POG estimation process is shown in Fig. 5.5.

The POG in 3D space is actually computed as the midpoint of the shortest distance between the two visual axis vectors, as the vectors are unlikely to exactly intersect as shown in Fig. 5.6 [161]. The points $P_l(s)$ and $P_r(t)$ can each be defined by a parametric equation of a line (5.8) and (5.9).

$$P_l(s) = C_l + s \cdot VA_l \tag{5.8}$$

$$P_r(t) = C_r + t \cdot VA_r \tag{5.9}$$

To minimize the distance joining the points $P_l(s)$ and $P_r(t)$, the vector W is defined from $P_l(s)$ to $P_r(t)$ and perpendicular to both VA_l and VA_r. Since W is perpendicular to both of the visual axis vectors, a system of two equations, (5.10) and (5.11), with two unknowns (the parameters s and t) can be defined and are readily determined.

$$VA_l \cdot [P_r(t) - P_l(s)] = 0 \tag{5.10}$$

$$VA_r \cdot [P_r(t) - P_l(s)] = 0 \tag{5.11}$$

Using model-based vergence to estimate the 3D POG is only valid provided the visual axis vectors of the eyes are not parallel, *i.e.* a unique solution to (5.10) and (5.11) can be found. As the distance to the point under observation increases, the visual axis vectors of the eyes increasingly approach a parallel course. Given a constant visual axis estimation accuracy, (typically 0.5 to 1.0 degree of visual angle) this means that the spatial accuracy of the estimated 3D POG will decrease with increasing depth from the eyes.

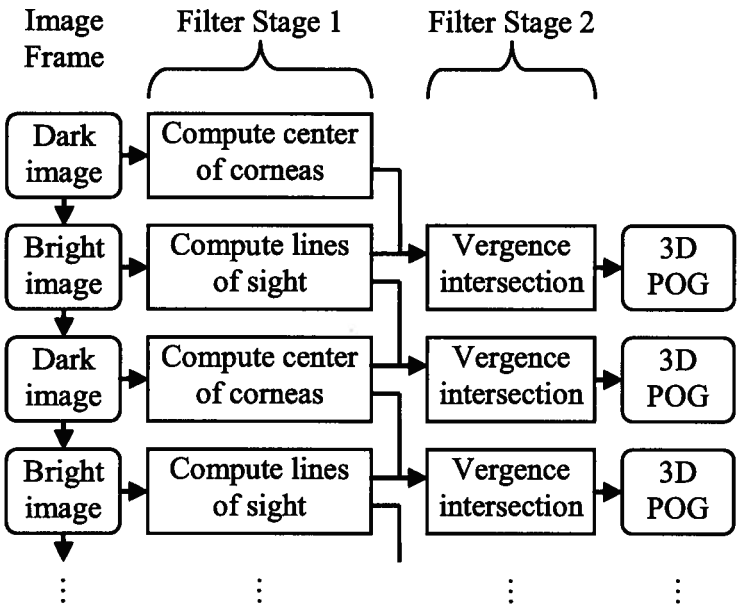


Figure 5.5: The alternating bright and dark pupil images are used to generate estimates for the centers of the corneas and the visual axis vectors. The vergence of the eyes can then be used to determine the 3D POG. Note that as a result of the image differencing technique each image frame results in an update for either the corneal centers of the visual axis vectors. The 3D POG is estimated at the full camera frame rate by using the model features from the current image frame, combined with the model features from the previous image frame.

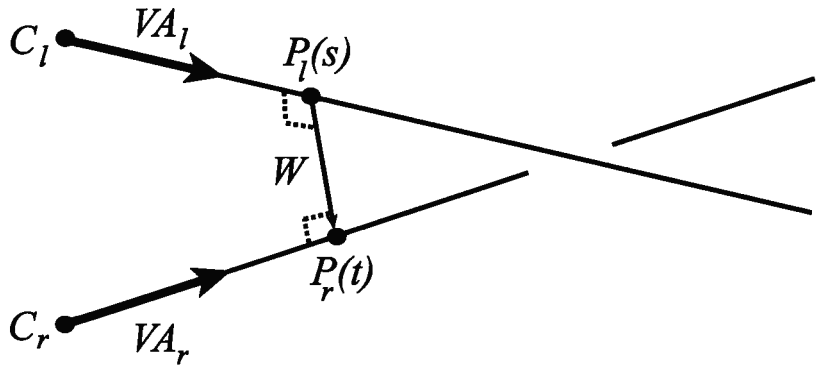


Figure 5.6: The POG in 3D space is determined by computing the points $P_l(s)$ and $P_r(t)$ on each visual axis vector which result in the closest distance between the two vectors. The 3D POG is the midpoint of the vector W formed from $P_l(s)$ to $P_r(t)$, where C_l and C_r are the locations of the left and right corneal centers and VA_l and VA_r are the left and right eye visual axes respectively.

5.2.5 Fixation filtering

The eyes are continuously in motion to keep the sensors of the eye refreshed during fixations [162]. The small motions of the eyes result in jittery visual axis vectors and can ultimately lead to poor precision in the estimated POG. With the model-based vergence technique for 3D POG estimation, increased error in the estimated visual axis vectors due to jitter can result in a much larger error in depth of the estimated 3D POG.

In the system presented here two levels of concurrent digital filtering were used to improve the precision of the 3D POG estimates as shown in Fig. 5.5. The first stage of lowpass filters (moving window averages) were used to stabilize the computed model features (corneal centers and visual axis vectors) while the second stage of filtering was used to stabilize the estimated 3D POG. The length of the filters can be used to tradeoff between precision and response time.

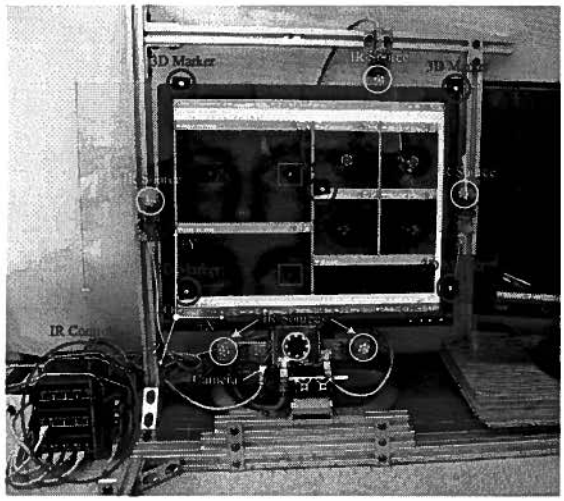
5.3 Experimental design and results

To evaluate the performance of the proposed design, the algorithms described were implemented and a set of experiments performed at the subsystem and system levels.

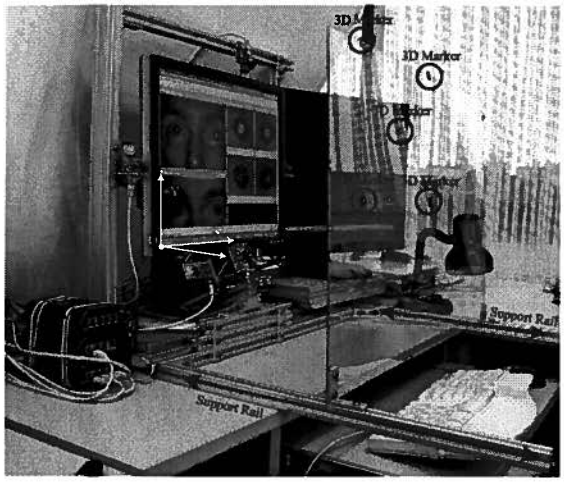
5.3.1 System Configuration

The system was comprised of multiple IR light sources, a high speed digital camera and a set of 3D POG markers as shown in Fig. 5.7. Each light source was composed of a set of seven closely spaced LED lights to approximate a point light source. The placement of the light sources were such that at least two valid reflections were formed off of the surface of the cornea at all eye rotations encountered. A microcontroller was used to synchronize the camera shutter with the alternating on-axis and off-axis LED's. The digital camera used was a monochrome DragonFly Express from Point Grey Research, capable of recording images with a resolution of 640 x 480 pixels at a frame rate of 200 Hz. The processor of the computer used for the system was an Intel 2.66 GHz Core 2 processor with 2 GB of RAM. A C++ implementation of the 3D POG estimation algorithms allowed 200 Hz real-time operation and data recording while offline analysis of the recorded data was performed in the MATLAB environment.

The 3D test point markers were placed in an X shape on a Plexiglas sheet which was mounted on aluminum rails. The corners of the X were spaced



(a) Front view of experimental setup



(b) Side view of experimental setup

Figure 5.7: Front and side views of the experimental setup are shown. In the front view the microcontroller and IR point light source expansion ports are located to the lower left of the screen. The off-axis IR point light sources are located around the frame and the on-axis IR ring is located in front of the camera lens. The 3D markers are located in an X grid of points on a clear Plexiglas sheet. The markers are a small cross on white paper, backed with black electrical tape for increased contrast for the subjects. In the side view the support rails are shown upon which the Plexiglas sheet can be translated in depth.

30 cm apart horizontally and 23 cm vertically. The rails were marked at 5 cm intervals at 6 different depths, resulting in a total workspace volume of 30 x 23 x 25 cm (width x height x depth). The total workspace volume exercised is comparable in size to modern volumetric displays [135]. An extruded aluminum structure was used to maintain the geometric positions between the camera, IR light sources and 3D position markers. The world coordinate system origin was located at an arbitrary position in 3D space. For convenience in development, it was located at the lower left corner of the monitor, with the positive X axis towards the right, the positive Y axis towards the ceiling and the positive Z axis towards the user.

5.3.2 Evaluation of filter length

The model features used to estimate the 3D POG suffer from jitter due to the natural motions of the eyes. The jittery model features can then lead to poor precision of the estimated 3D POG. To reduce the jitter and therefore increase the precision of the 3D POG, lowpass filters (moving window averages) with a user definable filter length were applied to the model features (corneal centers and visual axis vectors), as well as the final estimated 3D POG.

The accuracy and precision of the 3D POG was determined over a range of filter lengths to evaluate the effect of filtering. The experimental procedure involved a single subject, who was asked to fixate on a 3D test point located in the middle of the workspace volume while the raw image data used to compute the 3D POG was recorded.

Results

The recorded image data were then processed offline to compute the 3D POG using a variety of filter lengths. Shown in Table 5.1 are the average absolute errors, in addition to the standard deviations, over a consistent one second (200 samples) of data during the fixation. The 3D POG is listed by each coordinate (X, Y, Z) as well as the Euclidean distance error ($\sqrt{X^2 + Y^2 + Z^2}$). The maximum latency was determined as the time required for both filter histories to fill entirely with new fixation data. For example, at a sampling rate of 200 Hz, the 100 sample 3D POG filter requires 0.5 seconds, added to the 1 second for the 200 sample filter length for model features used in estimating the 3D POG, for a total latency of 1.5 seconds. For all further testing a filter length of 200 samples was used for the model features, and a filter length of 100 samples for the 3D POG as these produced the best results.

Table 5.1: Accuracy and standard deviation over varying filter lengths.

Model Length	3D POG Length	Latency (s)	Average Accuracy (cm) X	Y	Z	Euc.	Standard Dev. (cm) X	Y	Z	Euc.
1	1	0.005	0.34	0.43	3.30	3.41	0.26	0.30	2.57	2.50
20	10	0.15	0.17	0.43	1.65	1.79	0.09	0.14	1.29	1.19
100	50	0.75	0.12	0.40	1.07	1.20	0.03	0.05	0.61	0.53
200	100	1.5	0.15	0.43	0.44	0.70	0.02	0.01	0.40	0.27

5.3.3 Head Motion

Allowing the head to move naturally is a key goal of the proposed 3D POG estimation system. The ability to handle head motion is particularly important in 3D POG estimation as the head naturally moves and rotates while observing points in 3D space to reduce the strain on the extraocular muscles [163]. In this experiment, the allowable head space is such that both eyes remain in focus within the field of view of the camera. The experimental procedure involved a single subject, asked to observe a 3D test point located in the middle of the workspace volume. While observing the test point, the subject was asked to randomly position and rotate his/her head while exercising the full head space. A total of 24 different random locations and orientations were recorded. The first of the 24 positions was used as the calibration position. At each head position the estimated 3D POG was recorded, along with the positions of the left and right eyes (corneal centers) in 3D space.

Results

Accuracy was measured as the Euclidean distance between the estimated 3D POG and the actual 3D test point. The average error over the 23 head positions was found to be 1.96 cm with a standard deviation of 1.63 cm. From the calculated positions of the eyes the exercised head space spanned 3.2 cm horizontally, 9.2 cm vertically and 14 cm in depth.

5.3.4 Calibration Points

In the previous filter length and head motion experiments the subject observed a single test point which was calibrated at the same position. When extending the system to operate over the full workspace volume (30 x 23 x 25 cm), any number of 3D positions may be used as calibration points. While

a single point is sufficient to calibrate the system, the system accuracy may be increased by ensuring the 3D POG estimation algorithm is calibrated over the entire workspace volume.

The calibration experiment procedure involved a single subject, who was asked to observe each of the 30 3D test points located throughout the workspace volume. The computed corneal center and uncalibrated optical axis vectors were recorded at each test position for offline processing. The data collection procedure was repeated twice more to generate a total of three datasets. The first data set was post processed using various combinations of calibration positions to determine the optical axis angular offsets, which were then applied to the second and third datasets and the average 3D POG accuracy computed.

The calibration positions tested used 1, 5, 10, and 30 points. The single point calibration used the same mid-volume position as in the previous filter length and head motion experiments. The 5 point calibration used the 5 test positions located on the mid-volume plane. The 10 point calibration used the 5 points located on the first and last depth planes respectively. Finally, the 30 point calibration used all the data points from the complete workspace volume.

Results

The resulting average 3D POG accuracy when each calibration set was applied to the second and third datasets are shown in Table 5.2. An analysis of variance was performed to check for statistically significant differences in average accuracy between the calibration methods. Combining the second and third trials, a statistically significant difference was found between the techniques ($F(3,236)=7.273$, $p<0.001$). *Post hoc* analysis indicated that the average accuracy of the 1 and 5 point calibrations were worse than the 10 and 30 point calibrations, while there was no statistically significant difference between the 1 and 5 point calibrations or between the 10 and 30 point calibrations. The 10 point calibration procedure was therefore chosen for subsequent experiments as it maximized accuracy while minimizing the time required for calibration.

5.3.5 Multi-Subject Evaluation

An evaluation of the accuracy of the system was performed across a range of subjects to provide a more general indication of system performance. The experiment was evaluated over a total of 7 different subjects and exercised the full workspace (30 x 23 x 25 cm) for 3D POG estimation. The subjects were allowed freedom of head motion provided both eyes remained visible

Table 5.2: Average accuracy of 3D POG estimates for various calibration positions.

Calibration Points	Dataset Number	Average Accuracy (cm)	Standard Deviation (cm)
1 Point	2	5.47	4.04
1 Point	3	5.24	2.75
5 Point	2	4.84	4.58
5 Point	3	5.00	3.61
10 Point	2	3.19	2.83
10 Point	3	3.13	2.13
30 Point	2	3.22	2.76
30 Point	3	3.43	2.18

to the system camera. The subjects were all graduate students in the Electrical and Computer Engineering Department at the University of British Columbia (UBC). The subject ages ranged from 22 to 30 years old. Of the seven subjects 2 were female with 1 of 7 wearing contact lenses. The ethnicities of the subjects were 5 Caucasian and 2 Middle Eastern. The experimental procedures were certified for human experimentation by the Behavioral Research and Ethics Board of UBC under certificate H04-80920.

Each test subject was asked to observe each of the 5 points on the Plexiglas plane at the near and far depth planes to complete the 10 point calibration described in Section 5.3.4. The calibration corrections for each subject were then used to determine the subsequent 3D POG estimates. The data collection procedure required each subject to observe each of the 5 test positions on the Plexiglas sheet while the 3D POG was recorded, then move the sheet forward 5 cm, and repeated the 5 test positions until the entire workspace was exercised. The entire workspace volume was exercised twice to generate two trials per subject.

Results

The accuracy at each depth plane of the workspace volume, averaged over the two trials for all subjects is shown in Table 5.3, as well as the standard deviation. The accuracy reported is the average absolute error for the X, Y, and Z coordinates as well as the Euclidean distance error ($\sqrt{X^2+Y^2+Z^2}$). The depth of the planes are measured in centimeters from Z = 0 at the surface of the computer screen. The overall average accuracy and standard deviation for the entire workspace volume is also

shown.

Table 5.3: Average accuracy of 3D POG estimation at increasing depths from the world coordinate origin (towards the subject).

Z Depth (cm)	Average Accuracy (cm)				Standard
	X	Y	Z	Euc.	Deviation (cm)
17.5	1.28	1.20	4.04	4.61	3.14
22.5	1.28	1.27	3.64	4.28	2.81
27.5	1.26	1.13	3.36	3.98	3.11
32.5	1.10	1.04	3.20	3.75	2.96
37.5	1.31	1.13	2.55	3.35	2.59
42.5	1.38	1.46	2.60	3.62	2.14
Overall	1.27	1.20	3.23	3.93	2.83

5.3.6 Sensitivity Analysis

The potential sources of error in the system include: 1) extracted image features errors due to limited contrast and spatial resolution of the camera, 2) the simplified model of the eye with population averages for the eye model parameters, 3) errors in the camera lens calibration, and 4) errors in the physical measurement of the system. To provide an indication of the most significant sources of error, an analysis was performed of the sensitivity of the overall average accuracy with respect to both noise in the extracted image features and variations in system parameter values.

For this experiment the pupil and corneal reflection image centers were recorded rather than the computed 3D POG. The 3D POG at each data point was then recomputed offline using the raw image data, allowing evaluation of system parameter variation on a consistent data set. A single subject was asked to perform the 10 point calibration procedure as described previously. The subject then observed each of the 30 workspace points while the image data were recorded.

Results

Random Gaussian noise with zero mean and a fixed standard deviation (SD) was added to both the X and the Y coordinates of the extracted pupil center, the 3D POG computed, and the overall system accuracy was determined. The standard deviation of the noise was then increased and the process repeated. The procedure for the addition of noise was then repeated with the random noise added to both the X and the Y coordinates of the

corneal reflections. The results of the experiment are summarized in Table 5.4.

Table 5.4: Effect of noise in image feature extraction on system accuracy.

Noise SD in X & Y (pixels)	0	1	2	4
	Average Accuracy (cm)			
Pupil Center	3.78	3.92	4.16	5.59
Corneal Reflection Center	3.78	4.45	7.56	24.55

To evaluate the effect of model parameter deviations, the three eye model parameters (radius of cornea r, distance from center of cornea to center of pupil r_d, and index of refraction of the aqueous humor n) and the pinhole camera parameters (focal point f and critical point c_{px} and c_{py}) obtained through camera calibration were independently varied up to $\pm$ 10 % and the average accuracy was determined as listed in Table 5.5. The spatial coordinates of the off-axis light sources (Q) were also independently varied by up to $\pm$ 2 cm. Note that the accuracy results for the light source locations of the off-axis lights were averaged over the four lights for the X, Y, and Z coordinate variations.

Table 5.5: Sensitivity of average system accuracy to parameter variations.

Variation	-10%	-5%	0%	+5%	+10%
Eye Model		Average Accuracy (cm)			
r	5.31	4.41	3.78	3.93	4.66
r_d	5.09	3.71	3.78	5.16	6.92
n	3.77	3.53	3.78	4.45	5.17
Camera Model		Average Accuracy (cm)			
f	4.34	3.56	3.78	5.10	7.20
c_{px}	4.10	3.95	3.78	3.65	3.53
c_{py}	3.92	3.85	3.78	3.75	3.72

Variation	-2 cm	-1 cm	0 cm	+1 cm	+2 cm
Light Location		Average Accuracy (cm)			
Q (X)	4.12	3.82	3.78	3.90	4.21
Q (Y)	3.95	3.81	3.78	3.84	3.96
Q (Z)	3.84	3.81	3.78	3.79	3.80

5.4 Discussion

With rapid and robust image processing, a high speed sampling rate was achieved. Digital filtering was employed to improve precision at the expense of increased latency. In this chapter, filter lengths of 200 samples for the model features and 100 samples for the 3D POG were used. The filter lengths selected reduced the estimated POG jitter to 0.27 cm with a corresponding maximum latency of 1.5 seconds. To improve the latency of the system, fixation detection techniques may be employed to ensure that data from separate fixations are not combined in the digital filter histories, ensuring a rapid response to new fixations [160] [164].

The ability to handle head motion during 3D POG estimation is important as the head naturally reorients to reduce eye strain when observing points that require significant eye rotation. The ability to accurately estimate the 3D POG in the presence of unconstrained head motion was evaluated and an average accuracy of 1.96 cm was found over 23 different head positions and orientations. The full range of head positions spanned a head space volume of 3.2 x 9.2 x 14 cm (width x height x depth). Given the resolution of the camera sensor only a small degree of horizontal motion was possible as both eyes had to remain within the field of view of the camera. To improve the range of allowable head motion a camera with a higher resolution imaging sensor could be used to increase the field of view by decreasing the camera lens focal length without changing the effective spatial resolution.

The calibration algorithm outlined in this chapter only requires a single stage for per-user calibration. Calibration is performed by having the subject observe known positions in real world 3D space while the optical-to-visual axis offsets are determined. Statistical analysis indicated that using calibration points at only a single depth (1 and 5 points) resulted in worse accuracy than using calibration points located at different depths throughout the workspace volume (10 and 30 points). Calibration with 10 point (5 on the furthest and 5 on the closest depth planes) proved the most accurate with the shortest calibration duration.

A multi-subject experiment was performed to generalize the operation of the system over a larger population sample. The subjects were allowed to move their heads naturally while observing 3D points, provided both eyes remained within the field of view of the camera. The accuracy, averaged over all subjects, improved as expected as the distance from the eye to the 3D POG was reduced. An average accuracy of 4.61 cm at Z = 17.5 cm reduced to 3.35 cm at Z = 37.5 cm. Interestingly the error increased to

3.62 cm at Z = 42.5 cm (the plane located closest to the eyes). At the closest depth plane, the 3D test points located at the corners of the plane resulted in the most extreme eye rotations of the workspace. The increase in average 3D POG error at the nearest depth plane to the eyes is a result of the distortion of the corneal reflections when the eye is rotated to significant angles with respect to the camera. Over the entire workspace volume of 30 x 23 x 25 cm (width x height x depth) an average accuracy of 3.93 cm was determined. Given the accuracy, precision and latency achieved with the system presented here, a demonstration application was developed utilizing real-time 3D POG estimation to play a 3D game of Tic-Tac-Toe on a volumetric display in Hennessey and Lawrence [165].

To evaluate robustness and help direct further research, the sources of error leading to the average accuracy achieved were investigated by determining the effect of image feature noise and system model parameter variations. The addition of noise to the extracted corneal reflection locations considerably increased the error when compared with noise added to the pupil center as shown in Table 5.4. To reduce the effect of error in the corneal reflections, redundant off-axis light sources were used to avoid, as much as possible, the distortion that occurs when reflections approach the boundary between the cornea and scelera. Improved eye models which account for the change in curvature of the cornea may also be investigated as a means for further improvement.

Variation of the system parameters shown in Table 5.5 indicated that average accuracy was most sensitive to the eye model and the camera lens focal length parameters. Improvement of the eye model, either through increased sophistication (*i.e.* more accurately modeling the surface of the cornea) or more accurately identifying eye parameters (rather than using population averages) may lead to improved system accuracy. For remote eye model parameter estimation, the radius of the cornea and index of refraction may potentially be determined based on externally visible reflections and refraction respectively. As the distance from the center of the cornea to the center of the pupil occurs within the eye we expect this parameter to be fairly difficult to estimate from external images. One key advantage of using model-based methods for POG estimation over the P-CR or neural network based methods is that as the models of the eye improve, the accuracy of the model-based methods for both 2D and 3D POG estimation should improve as well. The desire for a higher resolution camera previously mentioned may also improve the performance of the camera calibration. Decreasing the focal length of the camera lens to increasing the field of view will also increase the perspective projection of the camera, becoming less orthographic and

increasing the needed depth information in the camera calibration images [166].

5.5 Conclusions

In this chapter techniques for a novel non-contact, head-free eye-gaze tracking system have been developed and quantitatively evaluated for 3D POG estimation in a real world scene. The 3D POG was estimated in a real world workspace volume of 30 x 23 x 25 cm and an average accuracy of 3.93 cm was achieved over seven subjects. The completely non-contact and head-free system had an allowable head space of 3 x 9 x 14 cm with the only requirement that both eyes be visible within the field of view of the camera. Through the two stages of high speed filtering the standard deviation of the unfiltered 3D POG was lowered from 2.5 cm to 0.27 cm with a corresponding maximum latency of 1.5 seconds. Reducing the maximum latency through fixation detection remains to be investigated. The use of a model-based approach for binocular eye-gaze tracking and a model-based vergence method of visual axis vector intersection allowed for a single stage of calibration. Future work will involve integration of a higher resolution camera for improving the range of free head motion, as well as researching improved models of the eye.

Acknowledgment

The authors would like to express their appreciation for the support of the Natural Sciences and Engineering Research Council of Canada (NSERC) Chair in Design Engineering, and NSERC Discovery Grant #4924.

References

[131] E. Kowler, *Eye Movements and their Role in Visual and Cognitive Processes.* Elsevier Science, 1990, vol. 4, ch. The role of visual and cognitive processes in the control of eye movement., pp. 1–70.

[132] R. Jacob and K. Karn, *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research.* Amsterdam: Elsevier Science, 2003, ch. Eye Tracking in Human-Computer Interaction and Usability Research: Ready to Deliver the Promises (Section Commentary), pp. 573–605.

[133] M. Halle, "Autostereoscopic displays and computer graphics," *SIGGRAPH Comput. Graph.*, vol. 31, no. 2, pp. 58–62, 1997.

[134] N. Dodgson, "Autostereoscopic 3d displays," *Computer*, vol. 38, no. 8, pp. 31 – 36, Aug. 2005.

[135] A. Jones, I. McDowall, H. Yamada, M. Bolas, and P. Debevec, "Rendering for an interactive 360° light field display," in *ACM SIGGRAPH.* New York, NY, USA: ACM, 2007, p. 40.

[136] K. Meyer, H. L. Applewhite, and F. A. Biocca, "A survey of position trackers," *Presence: Teleoper. Virtual Environ.*, vol. 1, no. 2, pp. 173–200, 1992.

[137] C. Ware, "Using hand position for virtual object placement," *Vis. Comput.*, vol. 6, no. 5, pp. 245–253, 1990.

[138] C. H. Morimoto and M. R. M. Mimica, "Eye gaze tracking techniques for interactive applications," *Comput. Vis. Image Underst.*, vol. 98, no. 1, pp. 4–24, 2005.

[139] D. Panchuk and J. Vickers, "Gaze behaviors of goaltenders under spatial-temporal constraints," *Human Movement Science*, vol. 25, no. 6, pp. 733–752, Dec. 2006.

[140] T. Hutchinson, J. White, W. Martin, K. Reichert, and L. Frey, "Human-computer interaction using eye-gaze input," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 19, no. 6, pp. 1527–1534, 1989.

[141] S.-W. Shih and J. Liu, "A novel approach to 3-d gaze tracking using stereo cameras," *IEEE Transactions on Systems, Man and Cybernetics, Part B*, vol. 34, no. 1, pp. 234–245, Feb. 2004.

[142] D. Beymer and M. Flickner, "Eye gaze tracking using an active stereo head," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, 18-20 June 2003, pp. II–451–II–458.

[143] 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, June 2006.

[144] W. J. Ryan, A. T. Duchowski, and S. T. Birchfield, "Limbus/pupil switching for wearable eye tracking under variable lighting conditions," in *Proceedings of the 2008 symposium on Eye tracking research & applications*. New York, NY, USA: ACM, 2008, pp. 61–64.

[145] A. T. Duchowski, V. Shivashankaraiah, T. Rawls, A. K. Gramopadhye, B. J. Melloy, and B. Kanki, "Binocular eye tracking in virtual reality for inspection training," in *Proceedings of the 2000 symposium on Eye tracking research & applications*. New York, NY, USA: ACM Press, 2000, pp. 89–96.

[146] B. K. Horn, *Robot Vision*. McGraw-Hill Higher Education, 1986.

[147] K. Essig, M. Pomplun, and H. Ritter, "Application of a novel neural approach to 3d gaze tracking: Vergence eye-movements in autostereograms," in *Proceedings of the 26thl Meeting of the Cognitive Science Society*, K. Forbus, D. Gentner, and T. Regier, Eds., 2004, pp. 357–362.

[148] ——, "A neural network for 3d gaze recording with binocular eye-trackers," *International Journal of Parallel, Emergent and Distributed Systems*, vol. 21, no. 2, pp. 79–95, April 2006.

[149] S. M. Munn and J. B. Pelz, "3d point-of-regard, position and head orientation from a portable monocular video-based eye tracker," in *Proceedings of the 2008 symposium on Eye tracking research & applications*. New York, NY, USA: ACM, 2008, pp. 181–188.

[150] Y.-M. Kwon and K.-W. Jeon, "Gaze computer interaction on stereo display," in *Proceedings of the 2006 ACM SIGCHI international conference on Advances in computer entertainment technology.* New York, NY, USA: ACM Press, 2006, p. 99.

[151] C. Hennessey, B. Noureddin, and P. Lawrence, "A single camera eye-gaze tracking system with free head motion," in *Proceedings of the 2006 symposium on Eye tracking research & applications.* New York, NY, USA: ACM Press, 2006, pp. 87–94.

[152] Y. Ebisawa and S. Satoh, "Effectiveness of pupil area detection technique using two light sources and image difference method," in *Proceedings of the 15th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Oct 28-31, 1993, pp. 1268–1269.

[153] C. H. Morimoto, D. Koons, A. Amir, and M. Flickner, "Pupil detection and tracking using multiple light sources." *Image and Vision Computing*, vol. 18, no. 4, pp. 331–335, 2000.

[154] A. Fitzgibbon, M. Pilu, and R. Fisher, "Direct least square fitting of ellipses," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 5, pp. 476–480, May 1999.

[155] H. Hua, C. W. Pansing, and J. P. Rolland, "Modeling of an eye-imaging system for optimizing illumination schemes in an eye-tracked head-mounted display," *Appl. Opt.*, vol. 46, no. 31, pp. 7757–7770, 2007.

[156] G. Cox and G. de Jager, "A survey of point pattern matching techniques and a new approach to point pattern recognition," in *Proceedings of the 1992 South African Symposium on Communications and Signal Processing*, 11 Sept. 1992, pp. 243–248.

[157] C. Hennessey, B. Noureddin, and P. Lawrence, "Fixation precision in high-speed noncontact eye-gaze tracking," *IEEE Transactions on Systems, Man and Cybernetics, Part B*, vol. 38, no. 2, pp. 289–298, April 2008.

[158] J.-Y. Bouguet, "Camera calibration toolbox for matlab," *www.vision.caltech.edu/bouguetj/.*

[159] D. A. Goss and R. W. West, *Introduction to the Optics of the Eye.* Butterworth Heinemann, 2001.

[160] R. Jacob, *Virtual Environments and Advanced Interface Design.* New York, NY, USA: Oxford University Press, 1995, ch. Eye tracking in advanced interface design, pp. 258–288.

[161] D. H. Eberly, *3D Game Engine Design.* Academic Press, 2001.

[162] R. J. K. Jacob, *Eye Movement-Based Human-Computer Interaction Techniques: Toward Non-Command Interfaces.* Norwood, N.J.: Ablex Publishing Co., 1993, vol. 4, pp. 151–190.

[163] R. S. Laramee and C. Ware, "Rivalry and interference with a head-mounted display," *ACM Transactions on Computer Human Interaction*, vol. 9, no. 3, pp. 238–251, 2002.

[164] A. T. Duchowski, *Eye Tracking Methodology: Theory and Practice.* Springer-Verlag, 2003.

[165] C. Hennessey and P. Lawrence, "3d point-of-gaze estimation on a volumetric display," in *Proceedings of the 2008 symposium on Eye tracking research & applications.* New York, NY, USA: ACM, 2008, pp. 59–59.

[166] X. Huang, J. Gao, and R. Yang, *Computer Vision*, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2007, vol. 4843, ch. Calibrating Pan-Tilt Cameras with Telephoto Lenses, pp. 127–137.

Chapter 6

Conclusions

In Chapters 2 through 4 the subsystems required for 3D POG estimations were described in detail, culminating in the 3D POG system presented and evaluated in Chapter 5. In Section 6.1 the conclusions of the previous chapters are summarized and discussed in the context of the overall thesis objectives. The strengths and weaknesses of the systems are outlined in Section 6.3 with a discussion of potential future work presented in Section 6.4.

6.1 Discussion

6.1.1 Model-Based POG Estimation Method

The first thesis objective was achieved with the development of a novel model-based method for monocular 2D POG estimation presented in Chapter 2. The model-based system was designed to operate remotely without making contact with the subject and to allow for free head motion. The model-based technique developed used a single camera, with models of the camera lens, eye and physical system.

In comparison of our system with the model-based method developed by Shih *et al* [167] the overall system accuracy achieved was slightly better, ranging from 0.46° to 0.90° of visual angle, compared with the 1° reported by Shih. The range of head motion reported by Shih *et al* was 4 x 4 cm with little depth due to a narrow depth of focus. For the 1024 x 768 pixel resolution camera tested with our system, a significantly larger range of head motion of 14 x 12 x 20 cm (width x height x depth) was achieved. A 640 x 480 pixel resolution camera was also tested with a corresponding head space of 7.5 x 5.5 x 19 cm. The allowable head space of the image based tracking technique presented is lower than the potential allowable head space achieved with the mechanical tracking systems by Beymer *et al* [168] and Ohno *et al* [169]. Using the image-based tracking system we developed however allows for faster tracking of head motion within the images, requires fewer cameras (2 in the system by Shih, 3 in the system by Ohno, and 4 in the system by Beymer), and relies on no moving components. As higher

resolution cameras become available the allowable head space of the image based tracking technique will increase accordingly.

When evaluating our technique with the 640 x 480 pixel resolution camera, a 30 Hz frame rate was achieved, comparable to that achieved by Shih and Ohno. The 1024 x 768 pixel resolution camera operated at 15 Hz, more comparable to the 10 Hz system operation reported by Beymer.

Contributions

The novel contributions of this work include a single camera, remote (non-contact), model based method for monocular eye-gaze tracking that allows for free head motion. The model based method provides information about the position and orientation of the eye in 3D space which is needed for 3D POG estimation based on the vergence of the eyes. The image based feature tracking system required only 10 ms to process each image and estimate the POG, resulting in a theoretical possible system update rate of 100 Hz.

6.1.2 Fixation Precision Enhancement

In the system presented in Chapter 3 the image-based tracking method used both software and hardware regions-of-interest (ROI). The software ROI greatly reduced the quantity of image information to process while the hardware ROI reduced the quantity of image information sent from the camera to the computer. Using the combination of hardware and software ROI's a high speed sampling rate of 407 Hz was achieved. The sampling rate achieved is considerably faster than the systems reported by Shih *et al* [167] and Ohno *et al* [169] both of which operated at 30 Hz. By the Nyquist criterion a sampling rate of over 300 Hz is desirable to avoid aliasing due to the low amplitude eye movements during fixations, with frequency components of up to 150 Hz [170].

Eye-gaze tracking systems frequently use low-pass filtering to improve precision by reducing the effect of the jittery eye movements during fixations. While the degree of filtering used was not reported, the precision of the remote model-based system by Yoo *et al*, was reported to be 0.84° of visual angle when operating at 15 Hz [171]. At the same frame rate, a fixation precision of a similar order of magnitude was observed in the system we developed at 0.55° of visual angle for the model-based method and a 0.205° for the P-CR method. When operating at a camera frame rate of 407 Hz with a filter length of 0.5 seconds however, the standard deviation was reduced to 0.05° and 0.035° of visual angle for the model-based and P-CR POG

estimation methods respectively.

Low-pass filtering of the POG estimates at low sampling rates can result in an increase in the latency or lag in the motion of the POG when the eye is reoriented to a new POG. However, based on the properties of the movements of the eyes, fast response times were maintained by tracking the beginning and end of the fixations. When the end of a fixation was detected due to the larger motion of a saccade, the history of the averaging filter was cleared. When the start of the following fixation was detected the filtering was begun anew, with the resulting filtered POG estimates based solely on the current fixation.

Contributions

A high speed image processing technique using a combination of software and hardware regions of interests was used to achieve POG estimation rates significantly higher than previously reported. With high speed POG estimation, aliasing of the sampled signal is avoided. Filtering of the high-speed POG estimates during fixations improved the precision by a factor of 11 times for the model-based POG method and 5.8 times for the P-CR method. The improvement in precision will become increasingly important, as the vergence technique for 3D POG estimation significantly magnifies the jitter in the depth.

6.1.3 Binocular Eye-gaze Tracking

Monocular tracking of a single eye is typically used in eye-gaze tracking as a means for reducing system complexity. Tracking a single eye is typically sufficient as both eyes generally point to the same position [172]. Binocular tracking however has a number of key advantages, including increased reliability to the loss of an eye through head motion, increased range of allowable head movement, potential improvement in POG accuracy and the ability to estimate the POG in 3D through the vergence of the eyes. The previously developed system was extended to binocular eye-gaze tracking in Chapter 4, while still maintaining the advantages of remote, non-contact operation with high speed image based tracking using only a single camera.

Tracking the position of the eyes within the face provides a means for determining which eye is visible when only a single eye is in the field of view of the camera. Contemporary face tracking techniques typically operate at 15 to 30 Hz [173], with commercial systems up to 60 Hz [174], and are unable to operate at the high rate required by our system. A face tracking

technique was therefore developed to track only the sides of the face, required for left and right eye differentiation, at very high speeds. The face tracking technique developed is based on background segmentation and requires only 0.2 ms to process operating at the full camera frame rate of 200 Hz.

With face tracking for eye differentiation, the range of horizontal head movement is effectively increased as the POG can still be determined even if a single eye translated out of the field of view of the camera. With a 640 x 480 pixel resolution camera, an allowable horizontal head motion of 4 cm was possible if both eyes had to remain within the field of view of the camera, the same as the binocular system by Shih *et al* [167]. Tracking a single eye (left or right) allowed up to 11 cm of horizontal head motion, however, using face tracking to differentiate the remaining visible eye allowed up to 18 cm of motion. The increase in allowable horizontal motion greatly increases the usability of the system by allowing for a more natural range of allowable head motions.

In Chapter 4 a novel technique was developed for tracking a pattern of multiple corneal reflections on the eye which allows for larger head and eye movements. Multiple off-axis light sources are used to generate the corneal reflection patterns, which are tracked using point pattern matching [175]. The algorithm developed for corneal reflection matching detects lost and distorted corneal reflections up to a user defined distortion threshold. The entire matching process is performed at high speed, requiring only 0.02 ms per eye. The corneal reflection pattern matching algorithm presented here has fewer restrictions on the placement of the off-axis light source than the method proposed by Hua *et al* [176], which required a symmetric arrangement of opposing light sources, coplaner with the surface of the camera sensor. The method presented by Hua *et al* also compensated only for the loss of corneal reflections and did not take into account the possible distortion of the reflections.

For 2D POG estimation, the P-CR method has the desirable characteristic of being system calibration free. Changes in the depth of the head however results in scaling of the P-CR image vector and therefore increased system error. The relative displacement of corneal reflections can be used to track the change in scale due to head motion. The system by Cerrolaza *et al* [177] used two corneal reflections to normalize the P-CR vector and showed the average accuracy of the POG estimation method degraded little over three head depths. The pattern matching technique described above can also be used to track the affine translation and scale parameters of the corneal reflection pattern. An experiment with 10 subjects was performed using the centroid of the corneal reflection pattern in the P-CR vector, as

well as normalizing by the inverse of the scale of the pattern. The results of the experiment showed that the enhanced P-CR method performed as well as the model-based method for POG estimation over several head displacements. One possible alternative technique to point pattern matching is to temporally sequence the recording of the corneal reflections [178] with one corneal reflection per recorded image. The issues with this proposed technique is the resulting decrease in POG estimation rate required to acquire all the corneal reflection image frames.

When both eyes are visible the POG can be estimated for both the left and right eyes independently. It has been observed that averaging of the binocular 2D POG estimates may result in a more accurate POG estimate [179]. In the 10 subject experiment the POG accuracy of the left, right and averaged POG estimates were compared. It was found that averaging the left and right eye POG estimates resulted in a binocular POG estimate that was statistically equal to or better than the monocular estimates alone for both the model-based and P-CR POG estimation methods.

Contributions

The novel contributions of this work include:

- High speed face tracking: A high speed face tracking technique was developed for eye differentiation when only a single eye is visible within the field of view of the camera.
- Corneal reflection pattern matching: A high speed technique for corneal reflection pattern tracking was developed for tracking redundant corneal reflections
- Enhanced P-CR: The P-CR method was enhanced using the corneal reflection pattern centroid and scaling of the P-CR vector. The enhanced P-CR method matched that of the model-based method for 2D POG estimation in the presence of head motion.

The extension from monocular to binocular model-based eye-gaze tracking allows for 3D POG estimation based on the vergence of the eyes. The corneal reflection pattern matching technique improves image feature reliability when larger head and eye rotations occur as a result of observing points in a 3D volume rather than a 2D screen.

6.1.4 3D POG Estimation

In the previous Chapters a number of key technical achievements have been accomplished to enabled remote 3D POG estimation. The key requirements include; 1) the model-based method tracking method for estimation of the center of the cornea and visual axis vector of the eye in 3D space, 2) high speed operation with filtering used to stabilize the eye estimates, 3) binocular eye-gaze tracking for vergence estimation of the 3D POG and 4) multiple corneal reflection tracking needed to ensure valid image features when large head and eye movements are used to observe 3D points in a volumetric space.

Calibration

The model-based method was used to determine the 3D position of the center of both the left and right eyes, as well as the visual axes along which the user was looking. As in the 2D case, the visual axis vectors were corrected from the optical axes through calibration. Unlike the 2D system however, the calibration test positions were located throughout a calibration volume of 30 x 25 x 25 cm (width x height x depth). It was found that using calibration points on the closest and furthest depth planes of the volume required the least user-calibration effort while still accurately calibrating the systems. The closest point of approach determined between the two 3D visual axes vectors was used as the estimate for the 3D POG within the workspace volume of 30 x 25 x 25 cm. The volume available for 3D POG tracking is equal to or greater than the volume encompassed by a number of current volumetric displays [180].

Head motion

As with the 2D POG estimation techniques developed previously, the 3D POG system was developed to operate remotely, requiring no contact with the user. The range of head motion was determined to be 3.2 x 9.2 x 14 cm. In this system the relatively low allowable horizontal head motion is due to the requirement that both eyes be visible in the camera at all times. Unfortunately this requirement reduced the range of allowable head motion, requiring the system users to increase their awareness of maintaining both of their eyes within the field of view of the camera. Provided the eyes remained within the field of view of the camera however, the users were able to perform significant head and eye rotations to comfortably observe points within the 3D workspace volume.

Precision and Latency

The binocular eye-gaze tracking system operated with a camera frame rate of 200 Hz, with a resulting latency of 5 ms between 3D POG estimates. The standard deviation of the resulting unfiltered 3D POG estimates during a fixation was found to be 2.55 cm. The 3D POG estimates were filtered with two stages of low pass filters to reduce the observed jitter due to the natural fluctuations of the eye. As in Chapter 3, a trade-off between latency and precision of the 3D POG estimates was observed, which can be appropriately selected depending on the intended application. For the evaluation of the 3D POG estimation system, a filter length of 1 second was used for smoothing of the estimated model features (corneal centers and visual axis vectors) with an addition 0.5 second filter used to smooth the resulting 3D POG estimates. The two stage filter resulted in a 3D POG latency of 1.5 seconds. However, when new fixations are detected, one does not need to wait 1.5 seconds. At the start of a new fixation, the filter memory can be immediately cleared and the filter begun on the new data. With low pass filtering the precision of the filtered 3D POG estimates during fixations was improved to a standard deviation of 0.26 cm.

Accuracy

An experiment was performed in which 7 different subjects observed 30 points throughout the workspace volume. Over all subjects and all positions an average accuracy of 3.93 cm was determined. Due to the nature of the vergence intersection for 3D POG estimation the resulting accuracy of 3D POG estimates decreases as the depth between the user and the POG target increases. The average accuracy error increased from 3.35 cm to 4.61 cm over a 20 cm increase in depth. It was also observed that at the nearest depth plane tested, the accuracy error was also increased to 3.62 cm. The increase in accuracy error at the closest depth tested was due to observing points at the extreme corners of the workspace volume in which even the remaining valid corneal reflections began to distort near the boundary between the cornea and sclera.

Sources of Error

Given the performance achieved by the 3D POG estimation system, an analysis of the sources of error was undertaken to determine potential directions for system improvement. The extracted image feature positions used for POG estimation as well as the parameter values for the models of the

system, camera and eye were varied and the resulting sensitivity in overall system accuracy determined.

For the image features, 2D Gaussian noise was added to the pupil center and corneal reflections which increased the error in the estimated 3D POG as shown in Table 5.4. For an equivalent increase in the standard deviation of the error added to the extracted image features, the error in the corneal reflections resulted in significantly larger decreases in system accuracy than the pupil center. As well, the pupil is a much larger image feature and therefore smaller image feature extraction errors would be expected. The error in the extracted corneal reflections is therefore the more significant source of error and bears further investigation for system improvement. The corneal reflections suffer from distortion due to the change in radius of the corneal surface towards the scelera. A more accurate model of the surface of the cornea may help to compensate for the corneal reflection image distortion. As the distortion appears radial in nature, the techniques for radial distortion compensation [181] may potentially be employed.

The model parameters of the system were also varied and the system accuracy determined as shown in Table 5.5. The largest changes in overall system accuracy were due to changes in the eye model parameters and camera focal length. Population averages were used for the eye model parameters which do not exactly match the individual users eye. Calibration to determine the eye model parameters on a per-user basis may help to improve the accuracy of the system. For the camera focal length the standard camera calibration checkerboard procedure was used, however the use of long focal lengths has been known to result in less accurate estimation of the intrinsic camera lens parameter values [182]. Increasing to a higher resolution camera sensor will allow for a decrease in focal length with equivalent spatial resolution, increasing the accuracy of the focal length estimation through camera calibration. The 3D measured locations of the off-axis light sources used to generate the corneal reflections did not prove to be a significant source of error. Small errors in the 3D positions of the light sources do not appear to lead to significant changes in the positions of the corneal reflection images on the surface of the camera sensor and consequently do not appear to have a large effect on the overall 3D POG estimation accuracy.

System Comparison

Previous research has been performed into the investigation of 3D POG estimation using commercial head mounted eye-gaze tracking systems. These systems used the traditional P-CR POG estimation method for determining

the POG of each eye on a 2D surface. In Duchowksi *et al* [183] the 2D POG estimates were tracked on individual left and right eye screens in a head mounted display (HMD) virtual reality system. A geometric vergence intersection method was used, combined with head pose information from an electromagnetic head tracker to determine the 3D POG in the virtual scene. For the system by Essig *et al* [184] the 2D POG estimates were tracked on a remote desktop display which used anaglyph images to create a virtual 3D environment.

In both systems, multiple user calibrations were required, both to calibrate the head mounted eye-gaze tracker, as well as to calibrate or train the 3D POG estimation systems. In the remote, non-contact system we present, the model-based method is used to estimate the 3D POG directly and therefore only a single user calibration stage is required. In addition the resulting 3D POG estimates are computed in a real-world coordinate system for potential real-world applications or use with volumetric displays, rather than the virtual displays used by Duchowski *et al* and Essig *et al.*

Contributions

There are four novel contributions for the development of the 3D POG estimation system. The first is the development of a system for estimating the 3D POG in the real 3D world rather than on a virtual display. The second contribution is the first non-contact, head-free 3D POG eye-gaze tracking system to be reported and/or evaluated in the literature. Thirdly, it is the first 3D POG estimation system that uses a model-based method for tracking the position of the eyes in 3D space and therefore requires only a single calibration stage. Finally, this is the first reported 3D POG system that requires no display while estimating the location of points in the real 3D world, as the system can be calibrated and operate in the 3D real-world.

Using the model-based method for 3D POG estimation also provide more insight into the operation of the system when compared with the non-parametric neural network approach. Additionally, improvements in the eye-model will likely lead to further improvements in the accuracy of the 3D POG estimation.

6.2 Application of 3D POG

To demonstrate the use of 3D POG estimation an application was developed illustrating the potential of the technique for human computer interaction [185]. A simple volumetric display was created using a 3 x 3 x 3 grid of

green and red LED lights to form a 28 x 23 x 22 cm workspace volume. A game of 3D Tic-Tac-Toe was then implemented using the 3D POG to select the desired 3D position to play. An experiment was performed in which a system user played a series of 10 games against the computer in which a total of 56 different positions were played. All 56 positions played were correctly identified by the system. For each position played the estimated 3D POG was also recorded. The average accuracy over all positions played over the 10 games was found to 3.2 cm. The latency of the system was such that the 3D POG was able to transit from diagonally opposite positions of the workspace volume (42 cm) in under 0.58 seconds.

6.3 Strengths and Weaknesses

The model-based tracking method has the benefit of allowing for free head motion without requiring contact with the user. For 2D and 3D POG estimation only a single calibration stage is required. Using models of the system also provides insight into the operation of the system when compared with non-parametric techniques such as neural networks. An investigation into the sources of error in Chapter 5 however indicated that the inaccurate population averages used for the eye model parameters may be a significant source of the estimation error. The population averages for the eye model parameters were used as it was not possible to determine the parameter values for each subject based on single camera images. Additionally the model of the eye itself may be a source of error, as the model is only a simplified version of the real eye. The assumption of a spherical corneal surface increasingly breaks down as the corneal reflections translate towards the boundary between the cornea and scelera which has a different radius of curvature.

The software and hardware ROI techniques allowed for high speed image processing and therefore rapid 3D POG estimation. As an artifact of the implementation of the hardware ROI by the camera manufacturer however, changing the hardware ROI position or size results in the aborting the currently exposed image before exposing again with the modified ROI. The aborted image is still transmitted to the computer and must be discarded, resulting in a slight increase in latency. To reduce the number of changes to the hardware ROI, the ROI size was increased slightly to allow the software ROI to track the position of the eye within the hardware ROI image, which was then only modified when the eye made larger, less frequent, changes in position.

The background segmentation technique used for face tracking allowed for fast processing and detection. Given the structured lighting of the system the background was kept relatively uniformly dark. To operate in an environment with a more cluttered background however, a more sophisticated segmentation method, possibly involving background subtraction and motion tracking, may be developed, rather than simple fixed level thresholding currently used.

The multiple corneal reflection pattern tracking technique operates at high speed and was shown to function well over many large eye movements. A distortion threshold parameter allowed a selectable level of distortion to compensate for small changes in scale of the corneal reflection pattern. Larger changes in the depth of the eye would result in larger scale changes which may not be detected by the technique. The depth of focus of the lens is currently the limiting factor at this point and the changes in scale are tracked within the allowable depth of field.

Finally the 3D POG estimation method was shown to operate well within the workspace and headspace volumes specified. The system operated remotely, requiring no contact with the user, and at high speed. While the user was free to move his/her head, the limited resolution of the imaging sensor allowed only a small degree of horizontal head movement as both eyes were required in the field of view of the camera. A simple and fast single stage calibration was used to compensate for the differences in foveal positions between system users. An application was developed successfully demonstrating the integration of the 3D POG as an interface tool in a simple game on a 3D volumetric display. While the accuracy achieved was sufficient for the Tic-Tac-Toe game developed, further increases in accuracy are be desirable as increased pointing resolution will further expands the range of potential applications.

6.4 Future Work

The system presented here is the first remote system for real-world 3D POG estimation of its kind. The performance of the system was characterized and the operation demonstrated with an application on a 3D Volumetric display. As 3D displays become more mainstream, the use of 3D POG as an interface tool will become increasingly important and the requirements for the performance of the technique will increase accordingly. A number of potential areas of future work for further developing the 3D POG estimation technology are as follows.

- Higher Resolution Camera: As both eyes are required for binocular eye-gaze tracking the range of horizontal head motion is diminished. Increasing the resolution of the camera sensor will allow for a large field of view and accordingly a larger range of allowable head motion. In addition, multiple copies of the camera system may be located about the 3D volume allowing for 3D POG tracking as the system user moves around the 3D volume.

- Improved Face Tracking: The face tracking algorithm may be improved to track facial features as well, such as the eyes. Tracking the eyes based on facial features may be used to aid the image differencing technique currently used for tracking the eyes in the images.

- Multiple Corneal Reflection Tracking: Extending the multiple corneal reflection pattern matching algorithm to implicitly compensate for scale may allow for greater changes in depth of the subjects head.

- Improved Eye Models: Finally the accuracy of 3D POG estimation may be improved through the development of models of the eye that more accurately reflect the true geometry of the eye. Once an improved model of the eye is developed, techniques for optimal identification of the eye model parameters on a per-user basis should be investigated.

References

[167] S.-W. Shih and J. Liu, "A novel approach to 3-d gaze tracking using stereo cameras," *IEEE Transactions on Systems, Man and Cybernetics, Part B*, vol. 34, no. 1, pp. 234–245, Feb. 2004.

[168] D. Beymer and M. Flickner, "Eye gaze tracking using an active stereo head," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, 18-20 June 2003, pp. II–451–II–458.

[169] T. Ohno and N. Mukawa, "A free-head, simple calibration, gaze tracking system that enables gaze-based interaction," in *Proceedings of the 2004 symposium on Eye tracking research & applications*. New York, NY, USA: ACM Press, 2004, pp. 115–122.

[170] A. Spauschus, J. Marsden, D. Halliday, J. Rosenberg, and P. Brown, "The origin of ocular microtremor in man," *Experimental Brain Research*, vol. 126, no. 4, pp. 556–562, June 1999.

[171] D. H. Yoo and M. J. Chung, "A novel non-intrusive eye gaze estimation using cross-ratio under large head motion," *Comput. Vis. Image Underst.*, vol. 98, no. 1, pp. 25–51, 2005.

[172] R. J. K. Jacob, *Eye Movement-Based Human-Computer Interaction Techniques: Toward Non-Command Interfaces*. Norwood, N.J.: Ablex Publishing Co., 1993, vol. 4, pp. 151–190.

[173] M. C. Santana, "On real-time face detection in video streams. an opportunistic approach." Ph.D. dissertation, Universidad de Las Palmas de Gran Canaria, March 2003.

[174] faceLAB, Seeing Machines, Canberra, Australia 2008.

[175] G. Cox and G. de Jager, "A survey of point pattern matching techniques and a new approach to point pattern recognition," in *Proceedings of the 1992 South African Symposium on Communications and Signal Processing*, 11 Sept. 1992, pp. 243–248.

[176] H. Hua, P. Krishnaswamy, and J. P. Rolland, "Video-based eyetracking methods and algorithms in head-mounted displays," *Opt. Express*, vol. 14, no. 10, pp. 4328–4350, 2006.

[177] J. J. Cerrolaza, A. Villanueva, and R. Cabeza, "Taxonomic study of polynomial regressions applied to the calibration of video-oculographic systems," in *Proceedings of the 2008 symposium on Eye tracking research & applications*. New York, NY, USA: ACM, 2008, pp. 259–266.

[178] J. D. Smith, R. Vertegaal, and C. Sohn, "Viewpointer: lightweight calibration-free eye tracking for ubiquitous handsfree deixis," in *Proceedings of the 18th annual ACM symposium on User interface software and technology*. New York, NY, USA: ACM Press, 2005, pp. 53–61.

[179] Y. Cui and J. M. Hondzinski, "Gaze tracking accuracy in humans: two eyes are better than one." *Neuroscience Letters*, vol. 396, no. 3, pp. 257–262, Apr 2006.

[180] A. Jones, I. McDowall, H. Yamada, M. Bolas, and P. Debevec, "Rendering for an interactive 360° light field display," in *ACM SIGGRAPH*. New York, NY, USA: ACM, 2007, p. 40.

[181] A. Nowakowsk and W. Skarbek, "Lens radial distortion calibration using homography of central points," in *EUROCON, 2007. The International Conference on "Computer as a Tool"*, Sept 2007, pp. 340–343.

[182] N. Daucher, M. Dhome, and J. T. Lapreste, "Camera calibration from spheres images," pp. 449–454, 1994.

[183] A. T. Duchowski, E. Medlin, N. Cournia, H. Murphy, A. Gramopadhye, S. Nair, J. Vorah, and B. Melloy, "3d eye movement analysis," *Behavior Research Methods, Instruments, & Computers (BRMIC)*, vol. 34, no. 4, pp. 573–591, Nov 2002.

[184] K. Essig, M. Pomplun, and H. Ritter, "A neural network for 3d gaze recording with binocular eyetrackers," *International Journal of Parallel, Emergent and Distributed Systems*, vol. 21, no. 2, pp. 79–95, April 2006.

[185] C. Hennessey and P. Lawrence, "3d point-of-gaze estimation on a volumetric display," in *Proceedings of the 2008 symposium on Eye tracking research & applications*. New York, NY, USA: ACM, 2008, pp. 59–59.

Appendix A

Research Ethics Approval

The University of British Columbia
Office of Research Services
Behavioural Research Ethics Board
Suite 102, 6190 Agronomy Road, Vancouver, B.C.
V6T 1Z3

CERTIFICATE OF APPROVAL- MINIMAL RISK RENEWAL

PRINCIPAL INVESTIGATOR:	DEPARTMENT:	UBC BREB NUMBER:
Peter D. Lawrence	UBC/Applied Science/Electrical and Computer Engineering	H04-80920

INSTITUTION(S) WHERE RESEARCH WILL BE CARRIED OUT:

Institution	Site
UBC	Vancouver (excludes UBC Hospital)

Other locations where the research will be conducted:
N/A

CO-INVESTIGATOR(S):

Craig Hennessey

SPONSORING AGENCIES:

Natural Sciences and Engineering Research Council of Canada (NSERC) - "Sensing and Signal Processing in the Telerobot Human Interface"

PROJECT TITLE:
Sensing and Signal Processing in the Telerobot Human Interface

EXPIRY DATE OF THIS APPROVAL: March 17, 2009

APPROVAL DATE: March 17, 2008

The Annual Renewal for Study have been reviewed and the procedures were found to be acceptable on ethical grounds for research involving human subjects.

Approval is issued on behalf of the Behavioural Research Ethics Board

Dr. M. Judith Lynam, Chair
Dr. Ken Craig, Chair
Dr. Jim Rupert, Associate Chair
Dr. Laurie Ford, Associate Chair
Dr. Daniel Salhani, Associate Chair
Dr. Anita Ho, Associate Chair

Figure A.1: Behavioral research ethics board approval renewal for 2008-2009.