A NOVEL HEAD-FREE POINT-OF-GAZE ESTIMATION SYSTEM

by

Elias Daniel Guestrin

A thesis submitted in conformity with the requirements
for the degree of Master of Applied Science
Graduate Departments of Edward S. Rogers Sr. Department of Electrical
and Computer Engineering
and Institute of Biomaterials and Biomedical Engineering,
University of Toronto

A Novel Head-Free Point-of-Gaze Estimation System

Elias Daniel Guestrin

Master of Applied Science

Edward S. Rogers Sr. Department of Electrical and Computer Engineering

University of Toronto

2003

Abstract

A novel real-time head-free point-of-gaze (POG) estimation system was developed. The system uses a video camera that captures images of one eye, and two IR light sources that illuminate the eye and produce two corneal reflections (glints). A geometrical-optical model is used to compute the POG on a computer monitor from the estimated image coordinates of pupil and glints centers. The sensitivity of the novel POG estimation method to errors in model parameters, noise in pupil and glints centers estimation, and fixation errors during calibration, is discussed. A non-spherical corneal model was developed to evaluate sub-optimal algorithms that improve the accuracy of the method. Experiments showed that the r.m.s. estimation error of the POG on a computer monitor is less than 1º. The POG estimation system was used in a pilot study to develop a visual aid for subjects with macular degeneration.

Acknowledgements

I would like to thank my supervisor Dr. Moshe Eizenman for his invaluable support and for giving me an almost unlimited freedom. The technical assistance provided by Mr. Tony Jares and Mr. Franz Schuh in the construction of the prototype of the point-of-gaze estimation system is highly appreciated. Dr. Esther González is to be thanked for her insights regarding macular degeneration, which were very valuable in the development of the visual aid proposed in this thesis. I am also grateful to Ms. Sarah Cherian (Graduate Admissions and Programs Officer of the Edward S. Rogers Sr. Department of Electrical and Computer Engineering) and Profs. Alf Dolan and Kim Woodhouse (former and current Graduate Coordinators of the Institute of Biomaterials and Biomedical Engineering) for their invaluable assistance. I would also like to acknowledge the financial support received from the Universidad Tecnológica Nacional (Argentina) and the University of Toronto.

I wish to express my gratitude to my former and present fellow lab members, Thas Yuwaraj, Lawrence Yu, Horace Lee, Brian Lui, Carmen Poon, Bryon Braymore and Ruth Sapir for enriching my graduate experience. In particular, I would like to emphasize that working with Brian Lui was an invaluable experience. I want to thank him for his friendship and his always-positive attitude.

I want to extend my gratitude to my landlords Michele Silver and Gary Burton for making me feel at home, and to my uncle, Héctor, and my aunt, Delfina, for their encouragement and support.

Finally, and most important, I am extremely grateful to my parents, Ester and Oscar, my sister, Analía, my brothers, Abraham and David, and my grandmother, Mecha, for providing an stimulating environment and for their unflagging support and encouragement over the years. For this and much more, I dedicate this thesis to them.

Table of contents

List of tables

List of figures

1. Introduction

When a person looks at an object, the eyes are oriented in such a way that the image of that object falls on the fovea of each eye. The fovea is the highest acuity region of the retina and its size corresponds to about 0.6 to 1° of visual angle [1]. In general, the point-of-gaze, i.e. what the person is looking at, is formally defined by the intersection of the visual axes of both eyes with the 3D scene. The visual axis is the line defined by the nodal point of the eye and the center of the fovea (Fig. 1.1). If the scene is a plane, such as a projection screen or a computer monitor, the point-of-gaze can be then defined as the intersection of the visual axis of one eye with the scene plane.

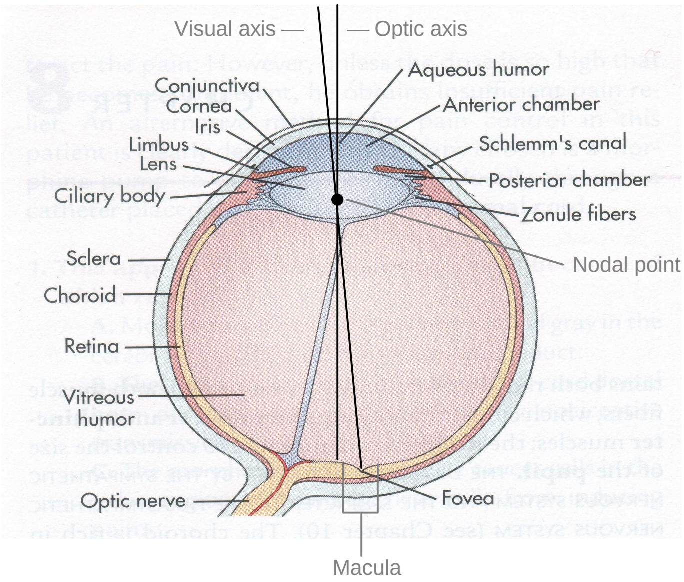


Figure 1.1: Schematic diagram of the right eye as viewed from above

(Adapted from [2])

A system that provides point-of-gaze information can be used in a large variety of applications that require the analysis of visual scanning patterns, and as an input device in human-machine interfaces. Visual scanning patterns give information regarding shifts in attentional focus and provide insights into human cognitive processes. Several examples of studies of visual scanning patterns in different areas can be given. One example, in which our research group was involved, is the study on visual selective attention for the quantification of mood disorders [3, 4, 5]. In such study, each subject was presented with a sequence of slides; each slide contained four images pertaining to neutral, dysphoric, threatening and social themes. Results showed that depressed subjects had an attentional bias towards dysphoric images, i.e. the mean time they spent looking at sad images was significantly larger than that of normal control subjects. Another example is the study of visual scanning patterns in attention, perceptual or learning disorders. In [6], an eye tracking system was used to study exploratory eye movements to pictures in schizophrenic, attention-deficit/hyperactivity disorder (ADHD), and normal children. In [7], eye movement patterns in linguistic and non-linguistic tasks in dyslexic children and adolescents were studied.

Point-of-gaze tracking technologies can be used in driving research [8, 9, 10, 11] and have the potential of being used as part of driver safety systems in the future [12, 13]. A driver’s visual scanning pattern can be analyzed in different driving conditions to get information about the focus of attention and provide insight into the reaction to road hazards, the effects of different dashboard layouts and sources of distraction such as cell phones, and the change in behavior due to different levels of fatigue. A point-of-gaze tracking system can also be used to measure and analyze the visual scanning patterns of pilots in the cockpit of an aircraft [14]. By comparing the scanning behavior of expert and amateur pilots in standard operational sequences, efficient scanning strategies can be identified, and inferences can be made between visual scanning patterns and hazard perception abilities.

Another area where a point-of-gaze tracking system can be a useful tool is in ergonomics, for the design and evaluation of human-machine interfaces. Well-organized computer interfaces result in shorter scan paths and more efficient search behavior than poorly designed interfaces [15].

The use of the point-of-gaze as an input modality in human-machine interfaces is a growing application field, from its integration in multimodal human-computer interfaces [16] to specific aids for motor disabled persons [17, 18]. For persons that can neither move nor talk but retain their intellectual capacities and can move their eyes, such as victims of spinal cord injuries or advanced ALS (Amyotrophic Lateral Sclerosis or Lou Gehrig's Disease), such a system may be the only means to communicate and control the environment through an eye-controlled computer interface. This list of applications is by no means exhaustive as the possible fields of applications of a point-of-gaze estimation system are only limited by imagination.

1.1 Methods for the estimation of the point-of-gaze

In general, in order to estimate the point-of-gaze, both the position and orientation of the eye in space must be known. A survey on general methods for estimating gaze direction is given in [1]. Some techniques, such as electro-oculography and contact lens methods [1] are not practical for general applications. Electro-oculography is based on the recording of potential differences on the skin around the eye, which are a function of eye orientation relative to the head. Although it has a large measurement range of up to ±70º, the accuracy is limited to about 1.5-2º, it requires electrodes to be placed on the skin of the subject and is sensitive to artifacts due to the activity of nearby facial muscles. Contact lens methods use special contact lenses that are tightly attached to the eye with negative pressure and have an embedded device that can be tracked. That embedded device can be, for example, a mirror that can be optically tracked, or a pair of mutually orthogonal miniature wire coils that pick up an induced voltage that is a function of eye orientation. The induced voltage is due to a magnetic field generated by two large perpendicular electromagnetic coils surrounding the subject. Even though contact lens methods can be very accurate (up to 2 seconds of arc), they cannot be used for long periods of time. The special contact lenses used can be cumbersome and very uncomfortable, often requiring the use of a local anesthetic.

Most modern eye tracking systems are video based and use images of various landmarks of the eye to estimate gaze. Video based eye tracking systems can be broadly classified as head-mounted systems and system that do not require head attachments. In head-mounted systems, the video camera that captures images of the eye is fixed relative to the head, permitting large and fast head movements without sacrificing the accuracy of the estimation of the gaze direction relative to the head. In order to estimate the point-of-gaze, in general, the 3D head pose (position and orientation) has to be known. There are various types of transducers that are used to sense head position, of which the most common is the magnetic position transducer [19]. The limitations associated with this approach include the additional noise present in the head position estimates, the restriction of the mobility of the subject to the range of the magnetic transmitter, and the susceptibility to errors due to electromagnetic interference and the presence of ferrous materials. The video based head-mounted eye tracking system developed by Dr. Eizenman's group at the University of Toronto (pupil center-corneal reflection method with dark pupil effect, described later) currently uses a different methodology to estimate head pose. A head mounted camera is used to record the scene in front of the subject. Visual cues extracted from the resulting image sequence are used to determine the head position relative to the observed scene [3, 4]. This eye tracking system has been used successfully in the analysis of visual scanning patterns of pilots in the cockpit of an aircraft [14], driving research [8, 11], and in psychiatry to study mood disorders [3, 4, 5].

Despite the advantages of head-mounted eye tracking systems, they are not suitable for applications that require monitoring gaze over long periods of time such as in the case of aids for motor disabled persons. Also, in applications that involve children, avoiding head attachments is usually desired. Current eye tracking technologies that do not require head attachments exhibit, in general, a trade-off between gaze estimation accuracy and allowed range of head movement. However, if head movements are limited, these technologies are suitable and attractive for applications in psychiatry and psychology, and for aids for motor disabled persons, where the point-of-gaze on a computer monitor is to be estimated. The discussion that follows concentrates on methods that are suitable for point-of-gaze estimation without head attachments (some of them are also suitable for head-mounted systems), since we are interested in the development of a system that will be less obtrusive and easier to use than the current head mounted system.

With the evolution of computer technology, video-based methods that combine the estimation of pupil center or iris center with a head tracker have been receiving great attention, as they are non-intrusive and head-free. In these methods, one [13,20] or two remote cameras (stereo vision) [9, 21, 22] are used to track facial features to estimate the 3D pose of the head. From the 3D head pose, the position of the center of the eyeballs can be estimated and by combining this information with the estimated positions of pupil centers or iris centers, the optic axes of the eyes can be reconstructed and, eventually, the point-of-gaze can be estimated. Given that these approaches use full-face images, the resolution of the pupil or iris is limited and the estimate of the point-of-gaze has an accuracy of only 3º [22]. This approach is useful in driver-safety systems in which the driver's focus of attention (e.g. front, rear-view mirrors, dashboard, etc.) is to be monitored.

An approach that uses the perspective projection of the iris-sclera boundary (limbus) to estimate the point-of-gaze has been suggested in [23, 24]. The general idea can be summarized as follows. The limbus can be modeled as a circle. The projection of the limbus obtained in a camera image is in general an ellipse (perspective projection of a circle). Theoretically, if the radius of the iris is known, the image of the limbus can be back projected to space and the 3D position of the iris center and the 3D orientation of the normal to the plane of the limbus can be reconstructed. Using this information, the point-of-gaze can be estimated. It should be noted that even though there are two possible solutions for the back projection problem and, consequently, two solutions for the position of the iris center and the orientation of the limbus plane, using appropriate physical constraints the spurious solution can be discarded. The main issue with this type of approach is image resolution. In general the top part and/or the bottom part of the iris is occluded by the eyelids and only the lateral edges of the iris can be used to fit an ellipse and apply the above methodology. In order to be able to fit an ellipse reliably in such conditions, a high-resolution image of the iris is required. This issue has been addressed by the use of two remote cameras, a pose camera that has a field of view somewhat larger than the whole face and a gaze camera mounted on a pan and tilt unit that takes zoom-in images of both eyes [23] or only one eye [24]. The pose camera is used to track the face and estimate the 3D head pose. The information about the head pose is used to aim the gaze camera to the eye(s). Point-of-gaze estimation errors of less than 1º of visual angle are reported in [24].

A popular type of approach to gaze estimation that has been considered since the early 60s [25] consists of using the pupil center and one or more corneal reflections. The front corneal surface resembles a convex mirror and reflection of light from any bright object (e.g. a discrete light source) forms a virtual image of the bright object behind the corneal surface. This virtual image appears as a bright spot that is called the first Purkinje image and usually referred to as corneal reflection or glint. The bright objects are usually near infrared (IR) light sources that are used to illuminate the eye while not interfering with normal vision (invisible to the human eye).

A fully operational video-based remote gaze estimation system using the pupil center and one corneal reflection was reported in the early 70s in [26]. In that system, a vidicon TV camera with a telephoto lens and a near-IR filter was used to obtain images of one eye. The near-IR illumination was collinear with the optic axis of the camera. The field of view of the camera was just slightly larger than an eye in order to be able to estimate the centers of the pupil and the corneal reflection as accurately as possible. Two versions of the system were described: one that allowed for eye translation in a volume of one cubic inch and one that allowed for eye translation in a volume of one cubic foot. The cubic-foot version incorporated moving mirrors and a moving lens to bring the image of the eye to the camera sensor and keep the image in focus. The basic gaze estimation principle can be described as follows. If the corneal surface is assumed to be a convex spherical mirror and the illumination is collimated, the distance between the orthographic projections of the pupil center and the virtual image of the light source on a fixed plane is proportional to the sine of the angle between the optic axis of the eye and the optic axis of the camera (collinear with the illumination). Furthermore, if the orthographic projection plane is parallel to the camera image plane (perpendicular to the illumination), the distance between the images of the centers of pupil and corneal reflection is equal to the product of the distance between the pupil center and the center of corneal curvature and the sine of the angle between the optic axis of the eye and the optic axis of the camera.

Using this idea, it is possible, in principle, to estimate the gaze direction relative to the camera axis after the system has been calibrated for each subject. However, the image obtained by the camera is a perspective projection rather than an orthographic projection and this can be an important source of error. If eye translation is restricted to only a few millimeters relative to the position adopted during calibration, the assumption of an orthographic projection is reasonable. In contrast, if eye translation is on the order of several centimeters, large estimation errors can occur unless the translation component is measured and compensated for. One example about how eye translation affects the estimation of gaze direction is when the distance between the eye and the camera changes. Suppose that the gaze direction relative to the camera axis remains constant but the distance between the eye and the camera changes, i.e. the size of the eye image changes. This affects the distance between the centers of pupil and corneal reflection in the eye image and leads to the wrong conclusion that the gaze direction has changed.

The estimation of the point-of-gaze is less straightforward than the estimation of gaze direction. In general, to estimate the point-of-gaze, a point in space through which the line of gaze passes is also needed. Again, if eye translation is restricted to only a few millimeters relative to the position adopted during calibration, it is possible to obtain a relatively accurate estimate of the point-of-gaze using the pupil center-corneal reflection vector. However, this is not the case in general. In the case of the cubic-foot version of the gaze estimation system [26], although the mathematical model is not given explicitly, it is inferred that it had the capability to estimate the position of the eye in space from the orientation of the mirrors and the range of the focusing system, and hence it could compensate for eye translation within the cubic foot of space. Accuracy of about 1° of visual angle was reported in that work.

In the light of the above discussion, it is not difficult to understand why when the vector from the pupil center to the corneal reflection alone is used to estimate the point-of-gaze, large estimation errors result when the head moves relative to the position adopted during calibration [17]. Several different solutions with different degrees of complexity have been proposed in order to deal with the problem of head displacement. In [27] an approach that can compensate for lateral head displacement, to some extent, but not for variation of the distance between the eye and the camera, is suggested. In this approach, the eye is assumed to be on a fixed plane parallel to the computer screen where the point-of-gaze is to be estimated. One light source is mounted coaxially with the optic axis of the camera, horizontally centered below the monitor. A second light source is mounted, vertically centered, on one side of the monitor. The combination of the two light sources is used to approximately mimic a single light source at the center of the screen in order to exploit system symmetries. The horizontal image coordinate of the glint corresponding to the light source mounted below the monitor, and the vertical image coordinate of the glint corresponding to the side mounted light source, are used to approximate the coordinates of a virtual glint that would be produced by a light source mounted at the center of the screen. The absolute coordinates of this virtual glint are used as an indication of eye translation in the plane parallel to the screen, while the vector from the pupil center to the virtual glint is used as an estimate of eye rotation. In that way, the sensitivity to lateral eye translation is reduced although the problem of variation of the distance between the eye and the camera is not solved.

Solutions that use information about the distance between the eye and the camera, and yield accuracies of about 1° of visual angle for the estimation of the point-of-gaze on a plane scene, have been proposed. In [28, 29, 30, 31] an ultrasonic distance meter mounted on the camera is used. In [12] a method using an asymmetric aperture to control an autofocus system and estimate the eye-camera distance is described. An asymmetric aperture built into the camera lens causes the image of the corneal reflection to vary in size and orientation as the eye moves forward and backward. The magnitude of the corneal reflection blur provides information about how far out-of-focus the camera is, and the orientation of the corneal reflection image provides polarity information as to whether the camera is focused too near or too far. This concept is implemented in a commercial system [32]. Another commercial system [33] utilizes a magnetic head tracker in order to compensate for error induced by head displacement relative to the position adopted during system calibration. However, the use of the magnetic head tracker means that a device has to be attached to the head of the subject. In [34], a second remote camera is used to track facial features and estimate the position of the eye in space.

An attempt to avoid the problem of estimating the position of the eye in space is presented in [35]. This approach uses multiple corneal reflections and exploits the invariance property of cross ratios in projective geometry. However, results are not accurate since the use of cross ratios assumes that the cornea is a plane mirror, which is incorrect.

The systems described in [12, 17, 26, 27, 28, 29, 32, 33] use the bright pupil effect. In these cases, the illumination (near-IR) is collinear with the optic axis of the camera. The IR light enters the eye and the retina reflects it back to the camera, making the pupil brighter than the surrounding iris. In these systems, the amount of light that enters the eye and is reflected back from the retina is a function of pupil diameter. For small pupil diameters (less than 4 mm) the contrast between the bright pupil and the iris is poor, making the extraction of the pupil boundary from the eye image very difficult or impossible. For this reason, systems using the bright pupil effect, in general, have to be used in dim light conditions [17, 26].

Additionally, experience has shown that in about 5 to 10 % of the population the near-IR reflectivity of the retina is not enough for the bright pupil effect to work properly [17, 36].

When the IR illumination is not coaxial with the optic axis of the camera, the pupil appears as a dark circle against the surrounding iris (dark pupil effect). This is because the light reflected by the retina is not collected back by the camera. The dark pupil effect is not affected by the pupil diameter and therefore it can be used in a larger variety of applications and environments. One of the papers cited above that uses the dark pupil effect is [34]. One reason for the interest in the bright pupil effect is that it usually allows for more compact implementations of the gaze estimation system.

An approach that combines the bright and dark pupil effects is used in [30, 31, 35]. In this approach, on-axis illumination (bright pupil) is used for one video field and off-axis illumination (dark pupil) is used for the other video field. The image difference between the two fields removes the effect of ambient IR illumination while the pupil appears as a bright disk. This technique allows for easy detection of the pupil boundary. A specific description of this method can be found in [37, 38, 39, 40].

Another method for the estimation of gaze direction, although not practical for a head-free system, is described briefly here for the sake of completeness. It is called the double Purkinje image method [1]. As light passes through the eye, reflections occur at each optical interface where the index of refraction changes. In addition to the corneal reflection (first Purkinje image) produced by the front surface of the cornea, there is a second image produced by the rear surface of the cornea, a third one produced by the front surface of the lens and a fourth image produced by the rear surface of the lens. The second Purkinje image is relatively dim and the third Purkinje image is formed in a plane far from the other images, hence they are not used in this method. This method uses the relation between the first and fourth Purkinje images to estimate the gaze direction relative to the axis of the collimated illumination. More specifically, the distance between the first and fourth Purkinje images is proportional to the sine of the angle between the optic axis of the eye and the illumination axis. The drawbacks of this method are that it has a limited tracking range and requires bright illumination in order to be able to detect the fourth Purkinje image reliably. In addition, an elaborate and carefully aligned optical system has to be placed close to the eye.

1.2 Proposed system specifications

In this thesis, a novel model-based monocular point-of-gaze estimation system that does not require any attachment to the head and allows for some head movement is presented. The system consists of a computer monitor to present visual information, two near-IR light sources at the sides of the monitor that illuminate the eye and produce two corneal reflections (glints), and a video camera with a near-IR filter, centered under the monitor, that captures images of the subject’s eye. Since the illumination is off-axis, the pupil appears darker than the surrounding iris. A mathematical model that uses the estimated coordinates of pupil and glints centers in each image of the eye to compute the point-of-gaze on the computer screen was developed. Using a set of model parameters that is estimated through a calibration procedure, the visual axis of the eye in space can be obtained and the point-of-gaze on the computer screen can be estimated.

In order to derive the specifications of the proposed point-of-gaze estimation system, the applications of interest were considered. In particular, in this thesis, the proposed system is applied as a tool in the development of a visual aid for persons with macular degeneration. This visual aid would allow them to read from a computer monitor. Aged-related macular degeneration is the leading cause of legal blindness in North America, affecting about 1.6 million persons over 50 years of age [41]. This disease affects the high-resolution part of the retina and results in impaired or complete central vision loss, dramatically affecting the quality of life. For this application, preliminary experiments suggest that an accuracy of 1º of visual angle and an estimation rate of 10 Hz are adequate. These requirements are consistent with other applications of interest for this system: aid for motor disabled persons and studies in psychiatry and psychology.

Eye-controlled computer interfaces for the motor disabled, including a word processor using hierarchical menu trees with as little as six options per screen, have been reported in the literature [17, 18]. In those interfaces, the number of options per screen was limited by the accuracy of the estimation of the point-of-gaze. Better accuracy permits an increase in the amount of information presented per screen thus reducing the number of steps needed to get to a desired menu option. Ideally, in the case of word processing, a full on-screen keyboard would be used. For this application, an accuracy of 1º of visual angle is adequate as an initial goal since it is in general possible for the subject to shift his/her gaze slightly to compensate for small estimation errors so that a pointer can be moved to the desired location on the screen. A minimum of ten estimates per second is considered adequate for most applications of this type.

In a typical experiment for the assessment of mood disorders [3, 4, 5], an accuracy of 1° of visual angle is satisfactory as it allows not only for the determination of which picture is the focus of attention but also for the identification of the area or areas of interest inside each picture. Unless the analysis of fast eye movements that shift the point-of-gaze is needed, an estimation rate of 15-30 Hz is adequate.

The above discussion suggests that an accuracy of 1° and an estimation rate of at least 10 Hz (15-30 Hz for analysis of visual scanning patterns) are satisfactory for the applications of interest.

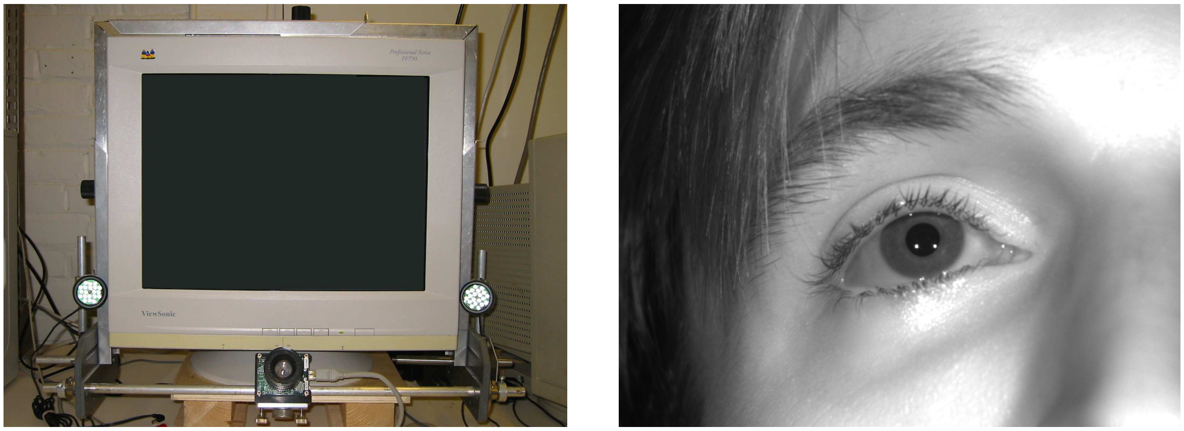
1.3 Organization of this document

Chapter 2 develops a mathematical model that relates the coordinates of pupil and glints centers in each image of the eye to the point-of-gaze on the computer screen. Chapter 3 describes the estimation of system and eye parameters that are required for the calculation of the point-of-gaze. It also discusses the sensitivity of the mathematical model to errors in each of the system and eye parameters, the effects of errors in the estimation of the coordinates of pupil and glints centers, and the effects of fixation errors during calibration. Chapter 4 discusses real corneas, that are not exactly spherical, and describes solutions that improve the estimation accuracy. Several simulation and experimental results are included. Finally, Chapter 5 proposes a novel visual aid to allow patients with macular degeneration to read from a computer monitor and shows preliminary experimental results where the point-of-gaze estimation system was used to simulate the disease in healthy subjects. The contributions of this thesis and future work are summarized at the end of the chapter.

2. Mathematical model

2.1 Introduction

The system consists of a computer monitor to present visual information, two near-IR light sources[1] at the sides of the monitor that illuminate the eye and produce two corneal reflections (glints), and a video camera with a near-IR filter, centered under the monitor, that captures images of the subject's eye (Fig. 2.1). This chapter describes a mathematical model that uses the estimated coordinates of the centers of the pupil and glints in each image of the eye to compute the subject's point-of-gaze on the computer monitor.



(a) System setup showing the video camera and the IR light sources

(b) Sample eye image showing the pupil and the two corneal reflections (glints)

Figure 2.1: System setup and sample eye image.

Section 2.2 details the assumptions used to develop the mathematical model. Section 2.3 defines the coordinate systems involved in the model. Section 2.4 develops the geometrical-optical model, which is the core of the mathematical model. Sections 2.5 and 2.6

---

[1] Each light source has 20 near-IR (850 nm) LEDs.

discuss the transformations between the different coordinate systems. Section 2.7 addresses the implementation of the equations to estimate the point-of-gaze from the coordinates of the centers of pupil and glints, while Section 2.8 considers the implementation of the equations to solve the inverse problem.

2.2 Model assumptions

The following set of assumptions was used to develop the mathematical model:

- The corneal bulge resembles a convex spherical mirror [1].
- The center of corneal curvature and the pupil center, as well as the center of rotation of the eye, are on the optic axis of the eye.
- The nodal point of the eye is assumed to be coincident with the center of corneal curvature.
- The video camera is modeled as a pinhole camera [42].
- The system configuration is fixed and its parameters (i.e. position of the light sources and position and orientation of the camera relative to the monitor) can be measured accurately.
- The parameters of the eye (i.e. radius of curvature of the cornea, distance between the center of corneal curvature and pupil center, and horizontal and vertical angles between the visual axis and the optic axis of the eye) are subject specific and can be estimated accurately through a calibration procedure (Chapter 3).
- The coordinates of the centers of the pupil and glints in the image of the eye can be estimated accurately for each image.

The validity of the assumption that the corneal bulge resembles a convex spherical mirror is discussed in Chapter 4.

2.3 Coordinate systems

Three coordinates systems are involved in the mathematical model: world coordinate system, camera coordinate system and image coordinate system.

2.3.1 World coordinate system

The world coordinate system is a right-handed 3-D Cartesian coordinate system whose axes are represented by upper-case letters (X, Y, Z). The XY-plane is coincident with the plane of the monitor and the origin is at the center of the screen, with the X-axis horizontal (Fig. 2.2(a)). These coordinates are measured in units of length, more specifically, in mm.

2.3.2 Camera coordinate system

The camera coordinate system is a right-handed 3-D Cartesian coordinate system whose axes are represented by lower-case letters (x, y, z). The xy-plane is coincident with the plane of the CCD[2] image sensor and the origin is at its center. The x-axis is in the direction of the rows, and the z-axis is coincident with the optic axis of the camera (Fig. 2.2(b)). These coordinates are also measured in units of length, typically in mm.

2.3.3 Image coordinate system

The image coordinate system is a 2-D coordinate system whose axes represent the row and column coordinates measured in pixels from the upper-left corner of the image (Fig. 2.2(c)). These coordinates are represented by lower-case letters (r, c).

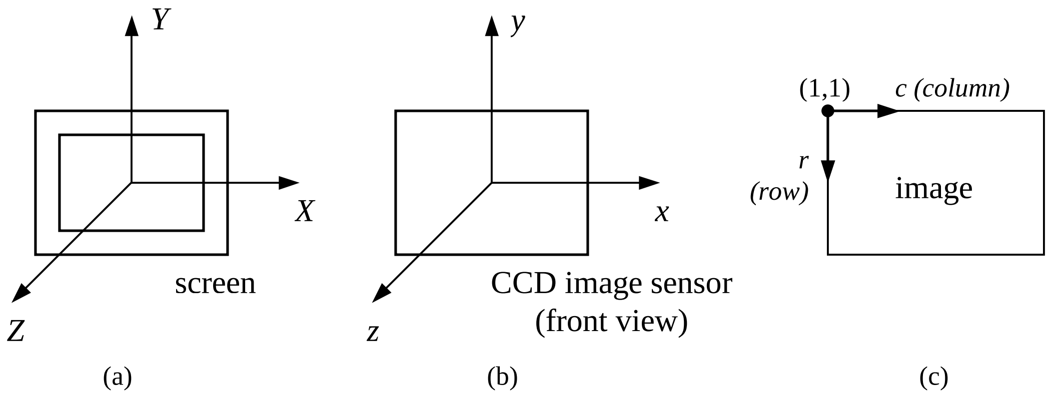


Figure 2.2: Coordinate systems: (a) world coordinate system, (b) camera coordinate system, and (c) image coordinate system.

---

[2] CCD = Charge-Coupled Device

2.4 Geometrical-optical model

Based on the laws of specular reflection and a pinhole camera model, a ray-tracing diagram can be drawn (Fig. 2.3) and a system of equations that describes the point-of-gaze as a function of the coordinates of the centers of the pupil and glints in the camera image can be constructed. Throughout this discussion, all vectors are column vectors in the world coordinate system, unless stated otherwise.

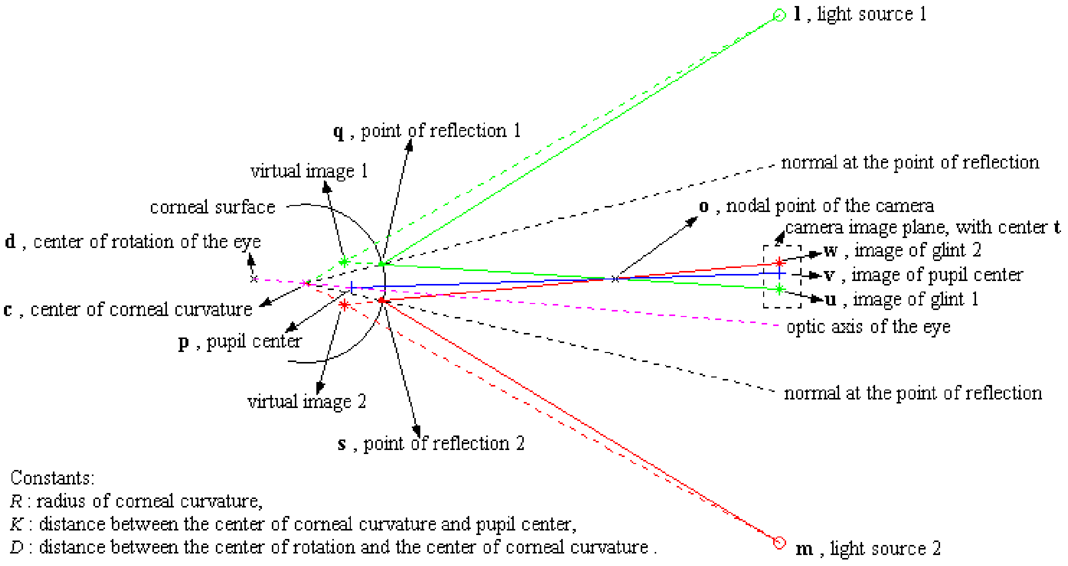


Figure 2.3: Ray tracing diagram (not to scale).

Consider a ray that comes from the center of light source 1, **l**, and reflects at a point **q** on the corneal surface such that the reflected ray passes through the nodal point of the camera, **o**, and intersects the camera image plane at **u**. As seen in Fig. 2.3, the condition that **q** is on the corneal surface, can be represented as

$$\left\|\mathbf{q}-\mathbf{c}\right\| = R\,. \qquad \textbf{(2.1)}$$

The laws of reflections state that the angle of incidence (i.e. the angle between the incident ray and the normal at the point of reflection) and the angle of reflection (i.e. the angle between the reflected ray and the normal at the point of reflection) are equal, and that the incident ray, the reflected ray and the normal at the point of reflection are coplanar. Noting that vector (**q** – **c**) is normal to the spherical surface at **q**, the two previous conditions can be expressed as

$$(\mathbf{l}-\mathbf{q})\bullet(\mathbf{q}-\mathbf{c})\cdot\|\mathbf{o}-\mathbf{q}\|=(\mathbf{o}-\mathbf{q})\bullet(\mathbf{q}-\mathbf{c})\cdot\|\mathbf{l}-\mathbf{q}\| \qquad \mathbf{(2.2)}$$

and

$$(\mathbf{l}-\mathbf{q})\times(\mathbf{o}-\mathbf{q})\bullet(\mathbf{q}-\mathbf{c})=0\ , \qquad \mathbf{(2.3)}$$

respectively.

Since the reflected ray passes through the nodal point of the camera, we have that the points **q**, **o** and **u** are collinear, which can be expressed in parametric form as

$$\mathbf{q}=\mathbf{o}+k_q(\mathbf{o}-\mathbf{u})\ ,\ \text{for some } k_q\,. \qquad \mathbf{(2.4)}$$

Similarly, for light source 2, the following equations result:

$$\|\mathbf{s}-\mathbf{c}\|=R\ , \qquad \mathbf{(2.5)}$$

$$(\mathbf{m}-\mathbf{s})\bullet(\mathbf{s}-\mathbf{c})\cdot\|\mathbf{o}-\mathbf{s}\|=(\mathbf{o}-\mathbf{s})\bullet(\mathbf{s}-\mathbf{c})\cdot\|\mathbf{m}-\mathbf{s}\|\ , \qquad \mathbf{(2.6)}$$

$$(\mathbf{m}-\mathbf{s})\times(\mathbf{o}-\mathbf{s})\bullet(\mathbf{s}-\mathbf{c})=0\ , \qquad \mathbf{(2.7)}$$

$$\mathbf{s}=\mathbf{o}+k_s(\mathbf{o}-\mathbf{w})\ ,\ \text{for some } k_s\,. \qquad \mathbf{(2.8)}$$

For the pupil center, **p**, we can write two equations, corresponding to the distance *K* between the pupil center and the center of corneal curvature, **c**, and the fact that the ray coming from **p** and passing through the nodal point, **o**, intersects the camera image plane at **v**. These conditions can be written as

$$\|\mathbf{p}-\mathbf{c}\|=K \qquad \mathbf{(2.9)}$$

and

$$\mathbf{p}=\mathbf{o}+k_p(\mathbf{o}-\mathbf{v})\ ,\ \text{for some}\, k_p\ , \qquad \mathbf{(2.10)}$$

respectively.

Now, consider the problem of reconstructing the optic axis of the eye as the line defined by points **c** and **p**, i.e. the center of corneal curvature and the pupil center. Let's assume that the parameters of the system **l**, **m** and **o**, the parameters of the eye *R* and *K*, and the positions of the images of the centers of pupil and glints in world coordinates **u**, **v** and **w** are known. Equations **(2.1)** thru **(2.10)** constitute a system of 7 scalar equations and 3 vector equations that are equivalent to a total of 16 scalar equations. The unknowns are the four points **c**, **p**, **q** and **s** and the scalar parameters k_q, k_s and k_p. At 3 scalar components per point, the total number of unknowns amount to 15. Consequently, there are enough equations to solve for the unknowns and reconstruct the optic axis of the eye, as explained in detail in Section 2.7. If only one glint is considered, the number of scalar unknowns reduces to 11, while the total number of scalar equations reduces to 10, making it impossible to reconstruct the optic axis of the eye without using some other constraints.

The point-of-gaze is defined by the intersection of the visual axis with the scene. The visual axis is the line defined by the nodal point of the eye[3] and the center of the fovea (i.e. the high-acuity region of the retina corresponding to 0.6 to 1° of visual angle), and is deviated from the optic axis [1]. In a typical adult human eye, the fovea falls about 4-5° temporally and about 1.5° below the point of intersection of the optic axis and the retina [43].

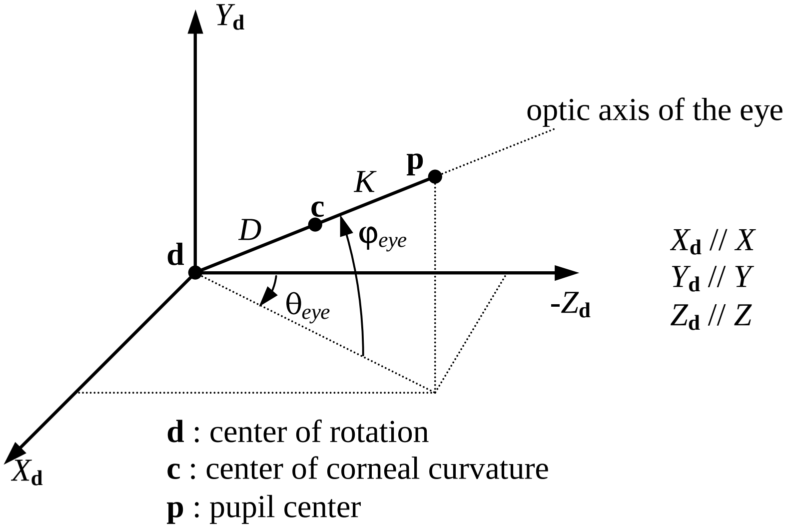


Figure 2.4: Orientation of the optic axis of the eye.

Since the system of equations that was developed at the beginning of this section allows for the reconstruction of the optic axis rather than the visual axis, the relation between the two axes has to be modeled. The orientation of the optic axis of the eye can be described by the pan (horizontal) angle θ_{eye} and the tilt (vertical) angle φ_{eye} defined in Fig. 2.4, where the world coordinate system is translated to the center of rotation of the eye, **d**. If the horizontal and vertical angles between the optic and visual axes are given by α_{eye} and β_{eye}, respectively, the orientation of the visual axis can be expressed by the pan angle $(\theta_{eye} + \alpha_{eye})$

---

[3] Actually, the eye has two nodal points, 0.3 mm apart. For the sake of simplicity, a single nodal point is considered.

and the tilt angle ($\varphi_{eye}+\beta_{eye}$), where all angles are signed. As it can be seen in Fig. 2.4, the angles θ_{eye} and φ_{eye} can be obtained from **c** and **p** by solving the following equation:

$$\mathbf{p}=\mathbf{c}+K\begin{bmatrix}\cos\varphi_{eye}\sin\theta_{eye}\\ \sin\varphi_{eye}\\ -\cos\varphi_{eye}\cos\theta_{eye}\end{bmatrix}. \qquad \textbf{(2.11)}$$

The optic axis and the visual axis intersect at the nodal point of the eye. The nodal point moves relative to the center of corneal curvature with different degrees of eye accommodation but the distance between them remains within 1 mm [1]. For this reason and for the sake of simplicity, the nodal point is assumed to be coincident with the center of corneal curvature. Fig. 2.5 shows a schematic cross section of the right eye seen from above.

From this discussion, the visual axis can then be described as

$$\mathbf{g}=\mathbf{c}+k_g\begin{bmatrix}\cos(\varphi_{eye}+\beta_{eye})\sin(\theta_{eye}+\alpha_{eye})\\ \sin(\varphi_{eye}+\beta_{eye})\\ -\cos(\varphi_{eye}+\beta_{eye})\cos(\theta_{eye}+\alpha_{eye})\end{bmatrix} \qquad \textbf{(2.12)}$$

for all k_g. Since by definition the scene is at Z = 0, the point-of-gaze is given by the previous equation for a value of k_g such that the Z-component of **g**, g_Z, equals 0, that is,

$$k_g=\frac{c_Z}{\cos(\varphi_{eye}+\beta_{eye})\cos(\theta_{eye}+\alpha_{eye})}. \qquad \textbf{(2.13)}$$

It should be noted that $\alpha_{eye}<0$ for the right eye and $\alpha_{eye}>0$ for the left eye, while $\beta_{eye}>0$.

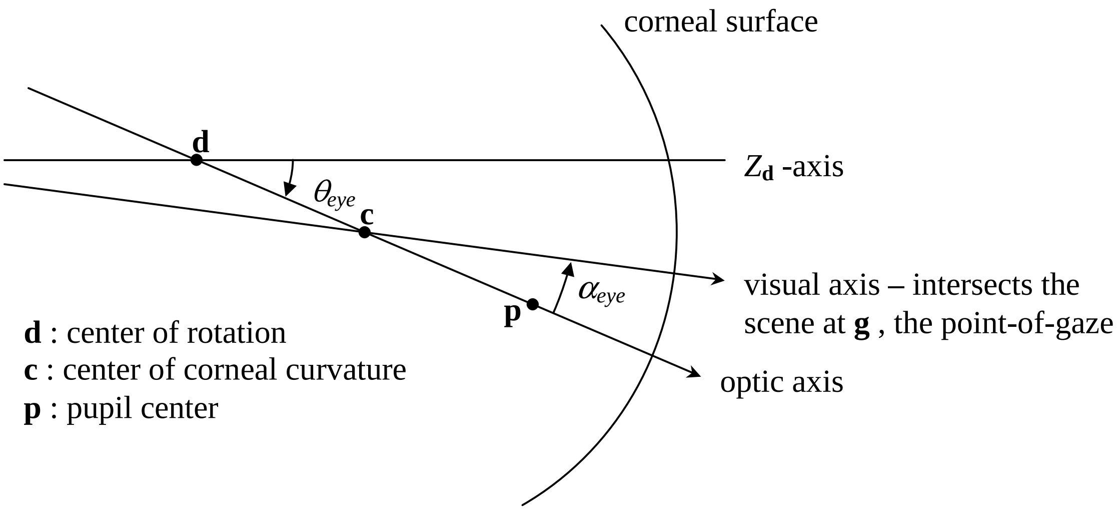


Figure 2.5: Schematic cross-section of the right eye seen from above, showing the optic and visual axes.

The development of this section assumed that the positions of the centers of the images of pupil and glints are expressed in world coordinates. However, in general they are given in pixels in the image coordinate system and need to be transformed to camera coordinates and then to world coordinates in order to estimate the point-of-gaze. The following two sections describe the relation among these three coordinate systems (see definitions in Section 2.3).

2.5 Perspective transformation and transformation between world and camera coordinate systems

As mentioned previously, the camera is modeled as a pinhole camera [42]. The perspective transformation, i.e. the projection of 3-D points onto the camera image plane, is roughly depicted in Fig. 2.3. An example of perspective transformation is shown for clarity in Fig. 2.6 for an object point **q**, an image point **u**, and a camera with the center of its image plane at **t** and its nodal point at **o**, where all these points are in world coordinates. The nodal point of the camera corresponds to the center of its lens.

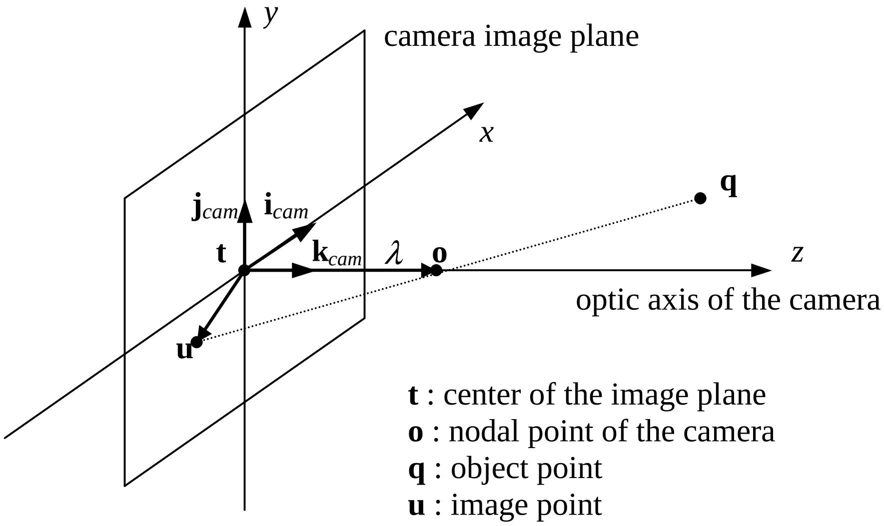


Figure 2.6: Perspective transformation.

The camera geometry can be characterized either by the center of the image plane, **t**, and the nodal point, **o**, or by the center of the image plane, the orientation of the optic axis given by the pan and tilt angles θ_{cam} and φ_{cam} and the distance between the nodal point and the image plane, represented by λ. Both characterizations are similar, as shown in Fig. 2.7, where a translated copy of the world coordinate system is placed with its origin at **t** as a reference. Formally, this relation is given by the following equation:

$$\mathbf{o} = \mathbf{t} + \lambda \begin{bmatrix} \cos\varphi_{cam} \sin\theta_{cam} \\ \sin\varphi_{cam} \\ \cos\varphi_{cam} \cos\theta_{cam} \end{bmatrix} . \qquad \textbf{(2.14)}$$

The value of λ is obtained from the Gaussian lens formula

$$\frac{1}{distance\ object-lens} + \frac{1}{distance\ image-lens} = \frac{1}{focal\ length} , \qquad \textbf{(2.15)}$$

where λ = *distance image–lens* and the *distance object–lens* is a typical value of the distance between the eye and the camera lens. As it will be shown later, the accuracy of the value of λ is not critical.

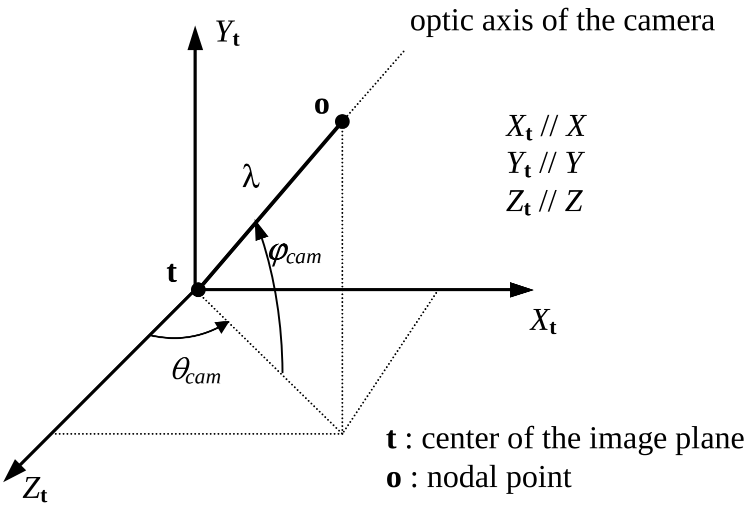
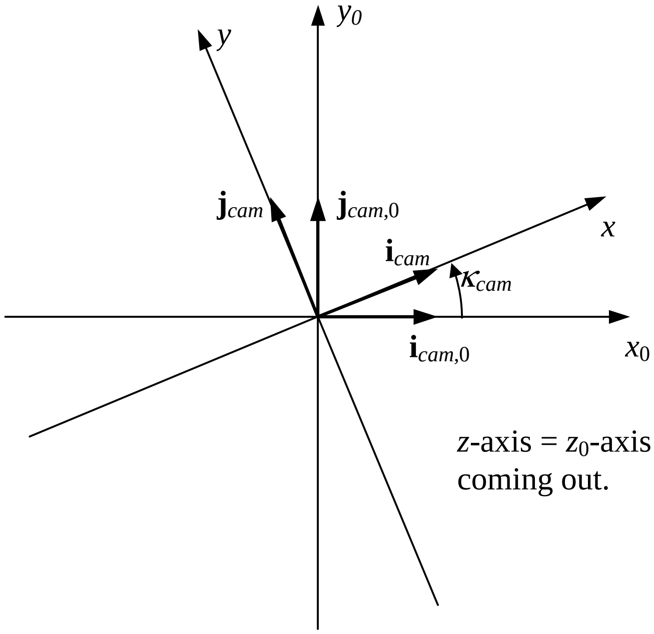


Figure 2.7: Camera geometry.

In order to provide a more general description, a rotation of the camera around its optic axis by an angle κ_{cam} is allowed, as shown in Fig. 2.8. In that figure, the x_0-axis is perpendicular to the Y-axis.



z-axis = z_0-axis : perpendicular to the page, coming out.

Figure 2.8: Rotation of the camera around its optic axis by an angle κ_{cam}.

The relation between the world coordinate system and the camera coordinate system can be analyzed in two equivalent ways. One approach is described here and the other in Appendix A. A point defined by $\mathbf{u} = [X_u\ Y_u\ Z_u]^T$ in the world coordinate system is represented by (x_u, y_u, z_u) in the camera coordinate system. Note that for points in the image plane $z_u = 0$ and therefore points in the image plane can be described in the camera coordinate system as (x_u, y_u).

The transformation between the world coordinate and the camera coordinate systems stems naturally from Fig. 2.6: a point $\mathbf{u}$ in world coordinates can be transformed into camera coordinates just by projecting the vector $(\mathbf{u} - \mathbf{t})$ on each one of the camera coordinate axes. Formally,

$$x_u = (\mathbf{u} - \mathbf{t}) \bullet \mathbf{i}_{cam}\ , \tag{2.16}$$

$$y_u = (\mathbf{u} - \mathbf{t}) \bullet \mathbf{j}_{cam}\ , \tag{2.17}$$

$$z_u = (\mathbf{u} - \mathbf{t}) \bullet \mathbf{k}_{cam}\ , \tag{2.18}$$

where $\mathbf{i}_{cam}$, $\mathbf{j}_{cam}$ and $\mathbf{k}_{cam}$ are the unit vectors in the direction of the x, y and z axes, respectively, expressed in world coordinates. Note that equation **(2.18)** is included for the sake of completeness. Conversely,

$$\mathbf{u} = x_u \mathbf{i}_{cam} + y_u \mathbf{j}_{cam} + z_u \mathbf{k}_{cam} + \mathbf{t} \ . \tag{2.19}$$

Note that the previous equations can be written as

$$\begin{bmatrix} x_u \\ y_u \\ z_u \end{bmatrix} = \begin{bmatrix} \mathbf{i}_{cam}{}^T \\ \mathbf{j}_{cam}{}^T \\ \mathbf{k}_{cam}{}^T \end{bmatrix} (\mathbf{u} - \mathbf{t}) = [\mathbf{i}_{cam} \quad \mathbf{j}_{cam} \quad \mathbf{k}_{cam}]^T (\mathbf{u} - \mathbf{t}) \tag{2.20}$$

and

$$\mathbf{u} = \begin{bmatrix} X_u \\ Y_u \\ Z_u \end{bmatrix} = [\mathbf{i}_{cam} \quad \mathbf{j}_{cam} \quad \mathbf{k}_{cam}] \begin{bmatrix} x_u \\ y_u \\ z_u \end{bmatrix} + \mathbf{t} \ , \tag{2.21}$$

respectively.

The unit vectors $\mathbf{i}_{cam}$, $\mathbf{j}_{cam}$ and $\mathbf{k}_{cam}$ are obtained as follows. From Figs. 2.6 and 2.7, and equation **(2.14)**, it should be clear that

$$\mathbf{k}_{cam} = \begin{bmatrix} \cos\varphi_{cam} \sin\theta_{cam} \\ \sin\varphi_{cam} \\ \cos\varphi_{cam} \cos\theta_{cam} \end{bmatrix} . \tag{2.22}$$

The unit vector $\mathbf{i}_{cam,0}$ (Fig. 2.8) is perpendicular to $\mathbf{k}_{cam}$ and $\mathbf{j}$ (the unit vector in the direction of the Y-axis, namely $[0\ 1\ 0]^T$), hence

$$\mathbf{i}_{cam,0} = \frac{\mathbf{j} \times \mathbf{k}_{cam}}{\left\| \mathbf{j} \times \mathbf{k}_{cam} \right\|} \ . \tag{2.23}$$

Then,

$$\mathbf{j}_{cam,0} = \mathbf{k}_{cam} \times \mathbf{i}_{cam,0} \ . \tag{2.24}$$

It can be observed from Fig. 2.8 that

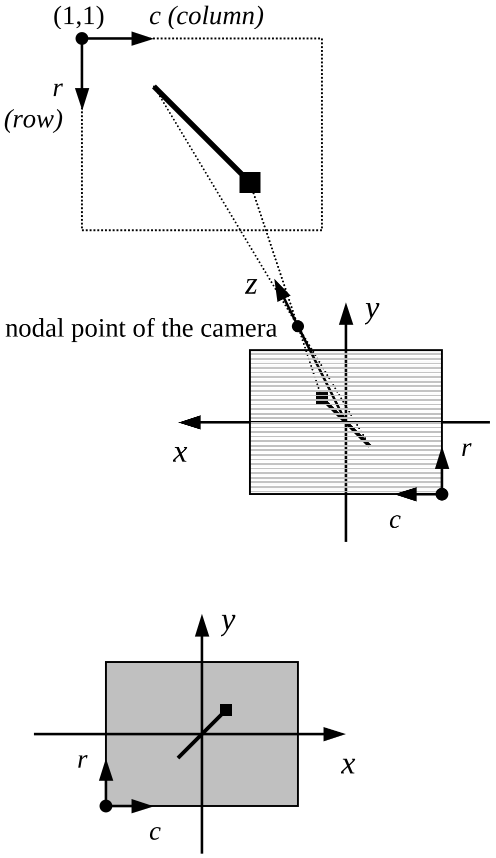
$$\mathbf{i}_{cam} = \cos\kappa_{cam} \mathbf{i}_{cam,0} + \sin\kappa_{cam} \mathbf{j}_{cam,0} \ , \tag{2.25}$$

$$\mathbf{j}_{cam} = -\sin\kappa_{cam} \mathbf{i}_{cam,0} + \cos\kappa_{cam} \mathbf{j}_{cam,0} \ . \tag{2.26}$$

Substitution of equations **(2.22)**, **(2.25)** and **(2.26)** into equation **(2.21)** allows for the transformation between the camera coordinate system and the world coordinate system.

2.6 Transformation between the image coordinate system and the camera coordinate system

In order to estimate the point-of-gaze in the computer screen, the centers of pupil and glints in the image of the eye, that are given in pixels in the image coordinate system, have to be converted to the camera coordinate system and then to the world coordinate system. The relations between the camera coordinate system and the image coordinate system are illustrated in Fig. 2.9.



object as seen from the camera and image captured by the camera

image on the camera image plane (as seen from the back of the image plane – looking towards the object)

image on the camera image plane (as seen from the front of the image plane – looking from the object)

Figure 2.9: Relations between the camera coordinate system and the image coordinate system.

Formally, for a point with row coordinate r and column coordinate c in pixels, the corresponding coordinates (x, y) in units of length in the camera coordinate system are obtained as

$$\begin{bmatrix} x \\ y \end{bmatrix} = pixel\,pitch \cdot \left(\begin{bmatrix} c \\ r \end{bmatrix} - \begin{bmatrix} c_{center} \\ r_{center} \end{bmatrix} \right) , \qquad \textbf{(2.27)}$$

where c_{center} and r_{center} are, respectively, the column coordinate and row coordinate of the center of the image in pixels and *pixel pitch* is the distance in units of length between adjacent pixels. It is assumed that the pixels are square, i.e. the pixel pitch is the same along the columns and along the rows. As an example, for a typical image of 640 pixels x 480 pixels, c_{center} = 320.5 and r_{center} = 240.5.

Correspondence between light sources and glints in the image of the eye

It is worthwhile to express the correspondence between the light sources and the glints explicitly as shown in the following table:

Light source	Point of reflection on the cornea	Image in world coordinates	Image in camera coordinates	Position of the image point in the image of the eye
l	**q**	**u**	(x_u, y_u)	Right
m	**s**	**w**	(x_w, y_w)	Left

Table 2.1: Correspondence between light sources and glints in the image of the eye.

2.7 Implementation of the equations to estimate the point-of-gaze

Assume that the positions of the light sources (**l** and **m**), the extrinsic parameters of the camera (**t**, θ_{cam}, φ_{cam}, κ_{cam} and λ), the intrinsic parameters of the camera (*pixel pitch*, c_{center} and r_{center}) and the parameters of the eye (R, K, α_{eye} and β_{eye}), are known. The position of the nodal point of the camera (**o**) is then calculated using equation **(2.14)**. The measurement and estimation of the above parameters are discussed in Chapter 3.

It was shown in Section 2.4 that the coordinates of the center of corneal curvature, **c**, and the pupil center, **p**, can be computed from the coordinates of the images of the centers of pupil and glints and then this information can be used to reconstruct the optic axis of the eye. From the optic axis, the visual axis can be reconstructed, and the point-of-gaze can be found as the intersection of the visual axis with the computer screen using equations **(2.12)** and **(2.13)**. This section shows how to rearrange the equations in Section 2.4 to reduce the number of unknowns and improve computation speed.

The development in Section 2.4 assumes that the coordinates of the images of the centers of pupil and glints are given in world coordinates. Since the coordinates of the centers of pupil and glints in the digitized image of the eye are expressed in pixels in the image coordinate system, they have to be transformed first to camera coordinates as explained in Section 2.6 (equation **(2.27)**) and then to world coordinates as described in Section 2.5 (equation **(2.19)** or **(2.21)**). After these two transformations, the images of the centers of pupil and glints are described in world coordinates by the points **v**, **u** and **w**, respectively.

The laws of reflection and the pinhole camera model imply that points **l**, **q**, **o**, **u** and **c**, are all in the same plane; this fact is formally described by equations **(2.3)** and **(2.4)** together. Similarly, points **m**, **s**, **o**, **w** and **c**, all share another plane; this fact is formally expressed by equations **(2.7)** and **(2.8)** together. These conditions imply that points **o** and **c** belong to both planes and, consequently, to the line that results from the intersection of both planes. The normal of the first plane can be written as $(\mathbf{u}-\mathbf{o})\times(\mathbf{l}-\mathbf{o})$, given that **l**, **o** and **u** are known. The normal of the second plane can be written as $(\mathbf{m}-\mathbf{o})\times(\mathbf{w}-\mathbf{o})$, given that **m**, **o** and **w** are known. Hence, the direction of the line that results from the intersection of both planes is given by

$$\mathbf{b}_{norm} = \frac{\mathbf{b}}{\|\mathbf{b}\|} \ , \ \text{where} \ \ \mathbf{b} = [(\mathbf{u}-\mathbf{o})\times(\mathbf{l}-\mathbf{o})]\times[(\mathbf{m}-\mathbf{o})\times(\mathbf{w}-\mathbf{o})] \ . \qquad \textbf{(2.28)}$$

If $\mathbf{b} \neq \mathbf{0}$, the center of corneal curvature **c** can be expressed in parametric form as

$$\mathbf{c} = \mathbf{o} + k_c \mathbf{b}_{norm} \ , \qquad \textbf{(2.29)}$$

where k_c is positive for the system configuration shown in Fig. 2.1, and represents the distance between the center of corneal curvature and the nodal point of the camera. The condition that $\mathbf{b} \neq \mathbf{0}$ is easily met for the system configuration shown in Fig. 2.1, where the light sources and the camera form a visible “V”.

Equations **(2.1)**, **(2.4)** and **(2.29)** can be combined to get

$$\|\mathbf{q}-\mathbf{c}\| = \|k_q(\mathbf{o}-\mathbf{u}) - k_c \mathbf{b}_{norm}\| = R \ . \qquad \textbf{(2.30)}$$

This equation can be re-written as

$$\|k_q(\mathbf{o}-\mathbf{u}) - k_c \mathbf{b}_{norm}\|^2 = [k_q(\mathbf{o}-\mathbf{u}) - k_c \mathbf{b}_{norm}] \bullet [k_q(\mathbf{o}-\mathbf{u}) - k_c \mathbf{b}_{norm}] = R^2 \qquad \textbf{(2.31)}$$

or

$$\|\mathbf{o}-\mathbf{u}\|^2 k_q{}^2 - 2k_c(\mathbf{o}-\mathbf{u}) \bullet \mathbf{b}_{norm} k_q + k_c{}^2 - R^2 = 0 \ , \qquad \textbf{(2.32)}$$

which is a quadratic in k_q. Solving for k_q and keeping the negative sign of the square root (the positive sign corresponds to a spurious solution “behind” the center of corneal curvature, on the extension of the spherical corneal surface), we obtain

$$k_q = \frac{k_c(\mathbf{o}-\mathbf{u})\bullet\mathbf{b}_{norm} - \sqrt{k_c{}^2[(\mathbf{o}-\mathbf{u})\bullet\mathbf{b}_{norm}]^2 - \|\mathbf{o}-\mathbf{u}\|^2(k_c{}^2 - R^2)}}{\|\mathbf{o}-\mathbf{u}\|^2} \; . \qquad \textbf{(2.33)}$$

This equation can be substituted into equation **(2.4)** to express **q** as a function of the parameter k_c.

Similarly, equations **(2.5)**, **(2.8)** and **(2.29)** can be combined and solved for k_s, while keeping the negative sign of the square root, to get

$$k_s = \frac{k_c(\mathbf{o}-\mathbf{w})\bullet\mathbf{b}_{norm} - \sqrt{k_c{}^2[(\mathbf{o}-\mathbf{w})\bullet\mathbf{b}_{norm}]^2 - \|\mathbf{o}-\mathbf{w}\|^2(k_c{}^2 - R^2)}}{\|\mathbf{o}-\mathbf{w}\|^2} \; . \qquad \textbf{(2.34)}$$

This equation can be substituted into equation **(2.8)** to express **s** as a function of the parameter k_c.

In a similar way, equations **(2.9)**, **(2.10)** and **(2.29)** can be combined and solved for k_p, while keeping the negative sign of the square root, to obtain

$$k_p = \frac{k_c(\mathbf{o}-\mathbf{v})\bullet\mathbf{b}_{norm} - \sqrt{k_c{}^2[(\mathbf{o}-\mathbf{v})\bullet\mathbf{b}_{norm}]^2 - \|\mathbf{o}-\mathbf{v}\|^2(k_c{}^2 - K^2)}}{\|\mathbf{o}-\mathbf{v}\|^2} \; . \qquad \textbf{(2.35)}$$

This equation can be substituted into equation **(2.10)** to express **p** as a function of the parameter k_c.

By using the above equations, the unknowns **q**, **s**, **c** and **p** are expressed as functions of the parameter k_c. The expressions for **q** and **c** as functions of the parameter k_c are substituted into equation **(2.2)**, which is solved numerically for k_c, to obtain a solution k_{c1}. The starting point for this numerical solution is taken to be a typical value of the distance between the eye and the camera lens. A second solution k_{c2} is obtained by substituting the expressions for **s** and **c** as functions of the parameter k_c into equation **(2.6),** which is also solved numerically for k_c. In this case, for faster convergence, the value k_{c1} is used as the starting point for the numerical solution. Then, the value for the parameter k_c is taken as the arithmetic mean of the two solutions k_{c1} and k_{c2}. Using the computed value of the parameter k_c, **c** is calculated with equation **(2.29)** and **p** is obtained using equations **(2.10)** and **(2.35)**. Hence, the optic axis of the eye can be reconstructed.

Having obtained **c** and **p**, the orientation of the optic axis of the eye, given by the pan angle θ_{eye} and the tilt angle φ_{eye} can be computed from equation **(2.11)** as

$$\varphi_{eye} = \arcsin \frac{p_Y - c_Y}{K} \tag{2.36}$$

and

$$\theta_{eye} = \arcsin \frac{p_X - c_X}{K \cos \varphi_{eye}} = -\arctan \frac{p_X - c_X}{p_Z - c_Z} \ . \tag{2.37}$$

Using these values of θ_{eye} and φ_{eye}, the point-of-gaze on the computer screen, **g**, is finally calculated with equations **(2.12)** and **(2.13)**.

The methodology developed so far in this chapter to estimate the point-of-gaze from the coordinates of the centers of pupil and glints in an image of the eye can be thought of as a function

$$\hat{\mathbf{g}} = \hat{\mathbf{g}}(\mathbf{f}, \boldsymbol{\rho}) \tag{2.38}$$

where

$$\mathbf{f} = [r_{left\ glint} \quad c_{left\ glint} \quad r_{pupil\ center} \quad c_{pupil\ center} \quad r_{right\ glint} \quad c_{right\ glint}] \tag{2.39}$$

is the vector of coordinates of the centers of pupil and glints in the image of the eye expressed in pixels in the image coordinate system, and $\boldsymbol{\rho}$ is the vector of model parameters used to estimate the point-of-gaze. The notation $\hat{\mathbf{g}}$ emphasizes that it is an estimate of the point-of-gaze.

Note that the information obtained through the computation of the point-of-gaze allows for the estimation of the position of the center of rotation of the eye **d** from

$$\mathbf{c} = \mathbf{d} + D \begin{bmatrix} \cos \varphi_{eye} \sin \theta_{eye} \\ \sin \varphi_{eye} \\ -\cos \varphi_{eye} \cos \theta_{eye} \end{bmatrix} , \tag{2.40}$$

as it can be seen from Fig. 2.4. In this way, the methodology of this chapter can be also thought of as a function

$$[\hat{\mathbf{g}} \quad \hat{\mathbf{d}}] = \mathcal{G}(\mathbf{f}, \boldsymbol{\rho}) \ , \tag{2.41}$$

where the notation “^” stands for estimate.

2.8 The inverse problem

Solving the inverse problem allows to calculate the coordinates of the centers of pupil and glints in an image of the eye from the position of the center of rotation of the eye, **d**, and the point-of-gaze, **g**, on the computer screen. This supports studies of the sensitivity of the estimation of the point-of-gaze to three type of error: (a) errors in the model parameters, (b) errors in the estimation of the coordinates of the centers of pupil and glints in each image of the eye, and (c) fixation errors during calibration.

Assuming that the system and eye parameters, including the distance between the center of rotation and the center of corneal curvature, D, are known, the inverse problem can be solved in several steps. The first step is to calculate the position of the center of corneal curvature, **c**. The value of **c** could be computed using equation **(2.40)** but the values of θ_{eye} and φ_{eye} are needed. For this purpose, equations **(2.12)** and **(2.40)** can be combined to get

$$\mathbf{g}=\mathbf{d}+D\begin{bmatrix}\cos\varphi_{eye}\sin\theta_{eye}\\ \sin\varphi_{eye}\\ -\cos\varphi_{eye}\cos\theta_{eye}\end{bmatrix}+k_g\begin{bmatrix}\cos(\varphi_{eye}+\beta_{eye})\sin(\theta_{eye}+\alpha_{eye})\\ \sin(\varphi_{eye}+\beta_{eye})\\ -\cos(\varphi_{eye}+\beta_{eye})\cos(\theta_{eye}+\alpha_{eye})\end{bmatrix}, \qquad \mathbf{(2.42)}$$

which can be solved numerically for θ_{eye}, φ_{eye} and k_g, thus allowing to obtain **c** from equation **(2.40)**. In order to reduce computation time, note that in the context of Fig. 2.5 and equations **(2.12)** and **(2.40)**, **c** can be computed from **d** and **g** going through the following steps:

(i) Take a first approximation to the visual axis as the line that goes through **d** and **g**. This is a reasonable approximation since the distance between **c** and **g** is two orders of magnitude greater than the distance between **d** and **c**. With this idea, we can write that

$$\begin{bmatrix}\cos(\varphi_{eye}+\beta_{eye})\sin(\theta_{eye}+\alpha_{eye})\\ \sin(\varphi_{eye}+\beta_{eye})\\ -\cos(\varphi_{eye}+\beta_{eye})\cos(\theta_{eye}+\alpha_{eye})\end{bmatrix}\approx\frac{\mathbf{g}-\mathbf{d}}{\|\mathbf{g}-\mathbf{d}\|} \qquad \mathbf{(2.43)}$$

and take the first approximation to the pan and tilt angles of the optic axis $\theta_{eye,0}$ and $\varphi_{eye,0}$ from

$$\begin{bmatrix}\cos(\varphi_{eye,0}+\beta_{eye})\sin(\theta_{eye,0}+\alpha_{eye})\\ \sin(\varphi_{eye,0}+\beta_{eye})\\ -\cos(\varphi_{eye,0}+\beta_{eye})\cos(\theta_{eye,0}+\alpha_{eye})\end{bmatrix}=\frac{\mathbf{g}-\mathbf{d}}{\|\mathbf{g}-\mathbf{d}\|} \qquad \mathbf{(2.44)}$$

obtaining

$$\varphi_{eye,0} = \arcsin\frac{g_Y - d_Y}{\|\mathbf{g}-\mathbf{d}\|} - \beta_{eye} \tag{2.45}$$

and

$$\theta_{eye,0} = \arcsin\frac{g_X - d_X}{\|\mathbf{g}-\mathbf{d}\|\cos(\varphi_{eye,0} + \beta_{eye})} - \alpha_{eye} = -\arctan\frac{g_X - d_X}{g_Z - d_Z} - \alpha_{eye} \ . \tag{2.46}$$

(ii) Using these values of $\theta_{eye,0}$ and $\varphi_{eye,0}$ compute the first approximation to the center of corneal curvature as

$$\mathbf{c}_0 = \mathbf{d} + D\begin{bmatrix} \cos\varphi_{eye,0}\sin\theta_{eye,0} \\ \sin\varphi_{eye,0} \\ -\cos\varphi_{eye,0}\cos\theta_{eye,0} \end{bmatrix} . \tag{2.47}$$

(iii) Set an iteration counter $i = 1$.

(iv) Take a further approximation to the visual axis as the line that goes through $\mathbf{c}_{(i\text{-}1)}$ and **g** as

$$\begin{bmatrix} \cos(\varphi_{eye,i} + \beta_{eye})\sin(\theta_{eye,i} + \alpha_{eye}) \\ \sin(\varphi_{eye,i} + \beta_{eye}) \\ -\cos(\varphi_{eye,i} + \beta_{eye})\cos(\theta_{eye,i} + \alpha_{eye}) \end{bmatrix} = \frac{\mathbf{g} - \mathbf{c}_{(i-1)}}{\|\mathbf{g} - \mathbf{c}_{(i-1)}\|} \tag{2.48}$$

to obtain

$$\varphi_{eye,i} = \arcsin\frac{g_Y - c_{Y,(i-1)}}{\|\mathbf{g} - \mathbf{c}_{(i-1)}\|} - \beta_{eye} \tag{2.49}$$

and

$$\theta_{eye,i} = \arcsin\frac{g_X - c_{X,(i-1)}}{\|\mathbf{g} - \mathbf{c}_{(i-1)}\|\cos(\varphi_{eye,i} + \beta_{eye})} - \alpha_{eye} = -\arctan\frac{g_X - c_{X,(i-1)}}{g_Z - c_{Z,(i-1)}} - \alpha_{eye} \ . \tag{2.50}$$

(v) Using these values of $\theta_{eye,i}$ and $\varphi_{eye,i}$ compute a new approximation to the center of corneal curvature as

$$\mathbf{c}_i = \mathbf{d} + D\begin{bmatrix} \cos\varphi_{eye,i}\sin\theta_{eye,i} \\ \sin\varphi_{eye,i} \\ -\cos\varphi_{eye,i}\cos\theta_{eye,i} \end{bmatrix} . \tag{2.51}$$

(vi) Increment the iteration counter i by 1. Steps (iv) and (v) can be repeated until $\|\mathbf{c}_i - \mathbf{c}_{(i-1)}\| \le \varepsilon$ where ε is a specified tolerance. However, in practice, the result for $i = 2$ is used. When $\mathbf{c} = \mathbf{c}_2$ is adopted and the coordinates of the images of the centers of pupil and glints in the image of the eye are calculated as explained in this Section, the distance between the estimate of the point-of-gaze computed as discussed in Section 2.7 and the actual point-of-gaze is less than 10^{-4} mm, which can be neglected.

Following the calculation of the value of **c**, it is now possible to find the points of reflection, **q** and **s**, independently of each other. The point of reflection **q** can be obtained from the system of equations given by equations **(2.1)** thru **(2.3)**. In that case there are three unknowns, namely, the three scalar components of vector **q**. However, calculations are sped up if those equations are rearranged such that the number of unknowns is reduced to one. For this purpose, note that equation **(2.3)** can be rewritten as

$$(\mathbf{l}-\mathbf{c})\times(\mathbf{o}-\mathbf{c})\bullet(\mathbf{q}-\mathbf{c})=0 \qquad \textbf{(2.52)}$$

and equation **(2.1)** can be expressed as (Fig. 2.10)

$$(\mathbf{o}-\mathbf{c})\bullet(\mathbf{q}-\mathbf{c})=\|\mathbf{o}-\mathbf{c}\|R\cos\nu_q \qquad \textbf{(2.53)}$$

and

$$(\mathbf{l}-\mathbf{c})\bullet(\mathbf{q}-\mathbf{c})=\|\mathbf{l}-\mathbf{c}\|R\cos(\omega_q-\nu_q)\ , \qquad \textbf{(2.54)}$$

where

$$\omega_q=\arccos\frac{(\mathbf{l}-\mathbf{c})\bullet(\mathbf{o}-\mathbf{c})}{\|\mathbf{l}-\mathbf{c}\|\,\|\mathbf{o}-\mathbf{c}\|}\ . \qquad \textbf{(2.55)}$$

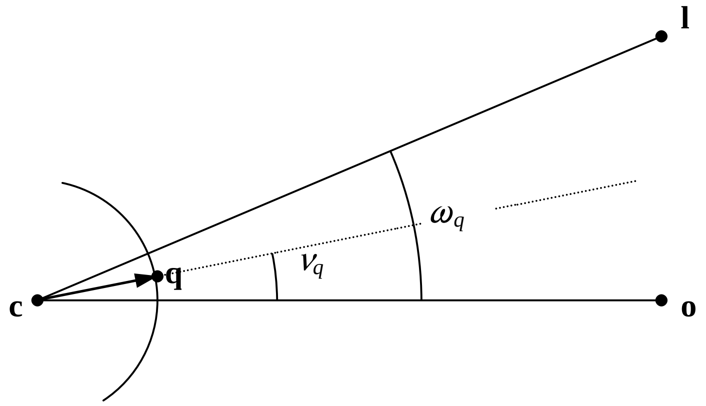


Figure 2.10: Representation of **q** in parametric form.

Equations **(2.52)** thru **(2.54)** can be written in matrix form as

$$\mathbf{Q}(\mathbf{q}-\mathbf{c})=\begin{bmatrix}0\\ \|\mathbf{o}-\mathbf{c}\|R\cos\nu_q\\ \|\mathbf{l}-\mathbf{c}\|R\cos(\omega_q-\nu_q)\end{bmatrix}, \quad \textbf{(2.56)}$$

where

$$\mathbf{Q}=\begin{bmatrix}[(\mathbf{l}-\mathbf{c})\times(\mathbf{o}-\mathbf{c})]^T\\ (\mathbf{o}-\mathbf{c})^T\\ (\mathbf{l}-\mathbf{c})^T\end{bmatrix} \quad \textbf{(2.57)}$$

and hence, from equation **(2.56)**, **q** can be expressed as a function of the single parameter ν_q as

$$\mathbf{q}=\mathbf{c}+\mathbf{Q}^{-1}\begin{bmatrix}0\\ \|\mathbf{o}-\mathbf{c}\|R\cos\nu_q\\ \|\mathbf{l}-\mathbf{c}\|R\cos(\omega_q-\nu_q)\end{bmatrix}. \quad \textbf{(2.58)}$$

Equation **(2.58)** is substituted into equation **(2.2)**, which is solved numerically for ν_q, taking $\nu_{q,0}=\omega_q/2$ as the starting point. Then, **q** is obtained using equation **(2.58)**.

Similarly, **s** is expressed in parametric form as

$$\mathbf{s}=\mathbf{c}+\mathbf{S}^{-1}\begin{bmatrix}0\\ \|\mathbf{o}-\mathbf{c}\|R\cos\nu_s\\ \|\mathbf{m}-\mathbf{c}\|R\cos(\omega_s-\nu_s)\end{bmatrix}, \quad \textbf{(2.59)}$$

where

$$\mathbf{S}=\begin{bmatrix}[(\mathbf{m}-\mathbf{c})\times(\mathbf{o}-\mathbf{c})]^T\\ (\mathbf{o}-\mathbf{c})^T\\ (\mathbf{m}-\mathbf{c})^T\end{bmatrix} \quad \textbf{(2.60)}$$

and

$$\omega_s=\arccos\frac{(\mathbf{m}-\mathbf{c})\bullet(\mathbf{o}-\mathbf{c})}{\|\mathbf{m}-\mathbf{c}\|\,\|\mathbf{o}-\mathbf{c}\|}. \quad \textbf{(2.61)}$$

Equation **(2.59)** is substituted into equation **(2.6)**, which is solved numerically for ν_s, taking $\nu_{s,0}=\omega_s/2$ as the starting point. Then, **s** is obtained using equation **(2.59)**.

The pupil center, **p**, can be computed directly using equation **(2.11)**. Once **p**, **q** and **s** are known, we can find the respective images **v**, **u** and **w** in the camera image plane. Towards this end, consider the image of the pupil center and rewrite equation **(2.10)** as

$$\mathbf{v} = \mathbf{o} + k_v(\mathbf{o} - \mathbf{p}) \ . \tag{2.62}$$

Given that the center of the camera image plane is given by point **t** and vector (**o** – **t**) defines the optic axis of the camera, normal to the image plane, the condition that **v** is in the image plane can be written as

$$(\mathbf{v} - \mathbf{t}) \bullet (\mathbf{o} - \mathbf{t}) = 0 \ . \tag{2.63}$$

Substituting equation **(2.62)** into **(2.63)**, solving for k_v and substituting back into **(2.62)**, we obtain

$$\mathbf{v} = \mathbf{o} + \frac{(\mathbf{o} - \mathbf{t}) \bullet (\mathbf{o} - \mathbf{t})}{(\mathbf{p} - \mathbf{o}) \bullet (\mathbf{o} - \mathbf{t})}(\mathbf{o} - \mathbf{p}) \ . \tag{2.64}$$

Similarly,

$$\mathbf{u} = \mathbf{o} + \frac{(\mathbf{o} - \mathbf{t}) \bullet (\mathbf{o} - \mathbf{t})}{(\mathbf{q} - \mathbf{o}) \bullet (\mathbf{o} - \mathbf{t})}(\mathbf{o} - \mathbf{q}) \ , \tag{2.65}$$

$$\mathbf{w} = \mathbf{o} + \frac{(\mathbf{o} - \mathbf{t}) \bullet (\mathbf{o} - \mathbf{t})}{(\mathbf{s} - \mathbf{o}) \bullet (\mathbf{o} - \mathbf{t})}(\mathbf{o} - \mathbf{s}) \ . \tag{2.66}$$

Finally, the coordinates of the centers of pupil and glints in world coordinates **v**, **u** and **w** are transformed to camera coordinates as explained in section 2.5 (equations **(2.16)** and **(2.17)** or equation **(2.20)**) and then to image coordinates as shown in section 2.6 (equation **(2.27)**).

The methodology developed in this section can be thought of as a function

$$\mathbf{f} = \mathbf{f}(\mathbf{g}, \mathbf{d}, \boldsymbol{\rho}) \tag{2.67}$$

where **g** is the point-of-gaze, **d** is the position of the center of rotation of the eye, $\boldsymbol{\rho}$ is the vector of model parameters and **f** is the vector of coordinates of the centers of pupil and glints in the image of the eye (equation **(2.39)**).

3. Estimation of system and eye parameters

3.1 Introduction

In order to apply the methodology developed in the previous chapter to estimate the point-of-gaze on the computer screen, several system and eye parameters must be known. The system parameters, namely the positions of the light sources and the position and orientation of the camera relative to the monitor, are measured. Given that the system configuration is fixed, these measurements are carried out only once. The system was configured in such a way that it is relatively easy to measure these parameters accurately. The positions of the centers of the light sources (**l** and **m**) are obtained by direct measurements using rulers and calipers. The tilt angle of the camera (φ_{cam}) is measured with a goniometer that is part of the system setup. The position of the center of the camera image plane (**t**) is measured indirectly, as it is calculated from the position of the axis of rotation of the camera holder relative to the screen, the position of the center of the image plane relative to the axis of rotation and the tilt angle of the camera. The pan angle of the camera (θ_{cam}) and the angle of rotation around the optic axis of the camera (κ_{cam}) are typically set to 0^o, although these two parameters are calculated as part of a calibration procedure in order to provide two additional degrees of freedom that contribute to improved estimation accuracy under real experimental conditions. The distance between the nodal point of the camera and the camera image plane (λ) is calculated from the focal length of the camera lens and a typical value of the distance between the cornea and the nodal point of the camera using equation (**2.15**). The typical value of the distance between the cornea and the nodal point of the camera is not critical, as it will be shown later. The remaining system parameters, which are needed for the transformation from image coordinates in pixels to camera and world coordinates, are the coordinates of the center of the image in pixels and the pixel pitch. If a digital camera is used, these intrinsic parameters are obtained from information provided by the manufacturers of the camera and the image sensor. However, if an analogue camera is used, information about the image acquisition board (frame grabber) is also required.

The parameters of the eye, namely the radius of curvature of the cornea (R), the distance between the center of corneal curvature and pupil center (K), and the horizontal and vertical angles between the visual axis and the optic axis (α_{eye} and β_{eye}), are subject specific and cannot be measured easily. Consequently, these four parameters are obtained through a calibration procedure that is performed once for each subject.

The nominal values of the system parameters and typical values of the eye parameters are given in Appendix B. The following section formulates the calibration problem. Section 3.3 presents a calibration example with experimental data. In Section 3.4 the sensitivity of the point-of-gaze estimation methodology to errors in model parameters is studied. Section 3.5 shows how the calibration procedure can help to compensate for errors in parameters that are not directly estimated by the calibration procedure. In Section 3.6 the effects of errors in the estimation of pupil and glints centers in the eye images are studied. Section 3.7 discusses the effects of fixation errors during the calibration procedure and summarizes the observations of this chapter.

3.2 Calibration procedure

The four eye parameters R, K, α_{eye} and β_{eye}, and the system parameters θ_{cam} and κ_{cam} are optimized through a calibration procedure. In this procedure, the subject is asked to fixate a sequence of known points $\mathbf{g}_i$, $i = 1, 2, \ldots, N$ on the screen (typically, 9 evenly distributed points). In order to stabilize the point-of-gaze at the desired position, a stimulus consisting of a circle that reduces its diameter and collapses on itself is used.

Typically, for each point $\mathbf{g}_i$, 100 images of the eye are acquired and the positions of the pupil and glints centers are estimated for each image. The results are then averaged to reduce the estimation error, producing the corresponding vectors of estimated coordinates of pupil and glints centers $\hat{\mathbf{f}}_i$, $i = 1, 2, \ldots, N$. These vectors have the form

$$\mathbf{f} = [r_{left\ glint} \quad c_{left\ glint} \quad r_{pupil\ center} \quad c_{pupil\ center} \quad r_{right\ glint} \quad c_{right\ glint}] \ . \tag{3.1}$$

Using these data, the values of the parameters are chosen such that the sum of the square errors between the estimated and intended points of gaze is minimized. Towards the formulation of the calibration problem, equation **(2.38)** is rewritten as $\hat{\mathbf{g}} = \hat{\mathbf{g}}(\mathbf{f}, [\boldsymbol{\rho}_{cal} \;\; \boldsymbol{\rho}_{fixed}])$ where $\boldsymbol{\rho}_{fixed}$ is the vector of parameters that are fixed and $\boldsymbol{\rho}_{cal}$ is the vector of parameters that are optimized by the calibration. The calibration problem is then formally stated as

$$\min_{\boldsymbol{\rho}_{cal}} \sum_{i=1}^{N} \left\| \hat{\mathbf{g}}(\hat{\mathbf{f}}_i, [\boldsymbol{\rho}_{cal} \;\; \boldsymbol{\rho}_{fixed}]) - \mathbf{g}_i \right\|^2 \quad , \quad \boldsymbol{\rho}_{cal} = [R \quad K \quad \alpha_{eye} \quad \beta_{eye} \quad \theta_{cam} \quad \kappa_{cam}] \; . \qquad \textbf{(3.2)}$$

This minimization problem is a multiparametric nonlinear least squares optimization problem.

In practice, a set of constraints based on the knowledge of eye anatomy and the physical characteristics of the system defines a region of feasibility for the solution. These constraints are bounds on the six parameters that are optimized by the calibration procedure and on a measure of the distance between the eye and the camera for each of the calibration points. Note that as part of the procedure to estimate the point-of-gaze, the position of the center of corneal curvature **c** is computed for each $\hat{\mathbf{f}}_i$, i = 1, 2, …, N, hence $\left\| \mathbf{c}_i - \mathbf{o} \right\|$ can be computed as a measure of the distance between the eye and the nodal point of the camera. The starting point and the bounds for the optimization problem are detailed in table 3.1. Note that the starting point is given by the typical/nominal values of the parameters that are optimized by the calibration procedure. Also note that the optimization constraints are rather loose, however they prevent the optimization algorithm from seeking a local optimum in a region that has no physical meaning.

Parameter	Starting point	Min	Max
R	7.8 mm	3 mm	20 mm
K	4.75 mm	2 mm	15 mm
α_{eye}	–5° for the right eye, 5° for the left eye	–10°	10°
β_{eye}	1.5°	–5°	5°
θ_{cam}	0°	–8°	8°
κ_{cam}	0°	–5°	5°
$\left\| \mathbf{c}_i - \mathbf{o} \right\|, \; i = 1, 2, \ldots, N$	N/A	400 mm	1000 mm

Table 3.1: Starting point and bounds for the calibration problem.

The calibration problem is solved as a general constrained optimization problem using a Sequential Quadratic Programming (SQP) algorithm [45, 46]. In the implementation of the SQP algorithm [46], gradients are calculated using an adaptive finite difference method. The computation of the finite difference gradients involves perturbing each of the variables in turn and calculating the rate of change in the function being optimized (objective function). The size of these perturbations is set adaptively within an interval whose extremes are chosen by the user. The performance of the optimization algorithms can be sensitive to truncation and round-off errors in the calculation of the finite difference gradients. If the minimum perturbation size is too small, it may result in significant error in the computation of the finite different gradients. If the minimum perturbation size is too large, the approximation to the gradient may be too coarse. In some problems, an adequate range of the perturbation size can help to overcome the effects of small discontinuities on the calculation of the finite difference gradients. In the case of the calibration problem, small discontinuities can arise from error propagation through the numerical computation of the objective function.

The performance of the optimization algorithm for the calibration problem, expressed as in equation **(3.2)**, was evaluated with several simulated examples with known solution and it was observed that a minimum step size on the order of 10^{-6} and a maximum step size of 10^{-1} led to a reasonable performance. In practice, under the assumption that the actual solution is close to the typical values of the calibration parameters, a change of variables was introduced with the aim of improving convergence, both in terms of speed and accuracy. Although this change of variables is not fully justified in general, it shows good behavior even if the solution is not close to the typical values of the parameters. Increased convergence speed cannot be guaranteed in all cases but the scaling associated with the change of variables is favorable for the computation of the finite difference gradients and the adoption of the search direction, thus making it attractive. Such change of variables is described in Appendix C.

3.3 Calibration example with experimental data

A calibration example with fixation data for subject E.G. without optical correction, is presented in Fig. 3.1. In this example, calibration data was collected as the subject fixated

9 points on the computer screen (table 3.2). These points are indicated by diamonds in Fig. 3.1 (the radius of the dashed circles centered on the diamonds is 10 mm). In this way, a set of vectors of coordinates of pupil and glints centers was obtained and used to optimize the 6 calibration parameters. Using these parameters, the point-of-gaze estimate was computed for each fixation point on the screen. These estimation results are indicated by asterisks in Fig. 3.1. A second independent set of vectors of coordinates of pupil and glints centers was collected in the same manner and the calibration parameters obtained with the first set (C.S.) were used to estimate the point-of-gaze. The corresponding estimates are indicated by pentagrams in Fig. 3.1. It is clear that the results obtained for each data set are consistent with each other, and the errors associated with them are very similar. This figure also shows that there is a small underestimation error in the vertical direction, while in the horizontal direction there is a marked overestimation error when the gaze is above the screen center and underestimation error below the screen center. The point-of-gaze estimates show a clear pattern that opens upwards/closes downwards.

There are several possible sources of error: (a) errors in the measurement and estimation of system and eye parameters, (b) errors in the estimation of the coordinates of pupil and glints centers, (c) errors in the fixation points during the calibration procedure, and (d) modeling errors. Since the results obtained from two independent data sets are consistent with each other, it is likely that the second and third sources of error are not playing a major role. The above sources of error, except modeling errors, are analyzed in the following sections. Modeling errors will be discussed in Chapter 4. This analysis provides insight into the effects of each type of error on the point-of-gaze estimation.

i	g_X (mm)	g_Y (mm)	g_Z (mm)
1	-132	100	0
2	0	100	0
3	132	100	0
4	-132	0	0
5	0	0	0
6	132	0	0
7	-132	-100	0
8	0	-100	0
9	132	-100	0

Table 3.2: Coordinates of the points of gaze used for the calibration under experimental conditions.

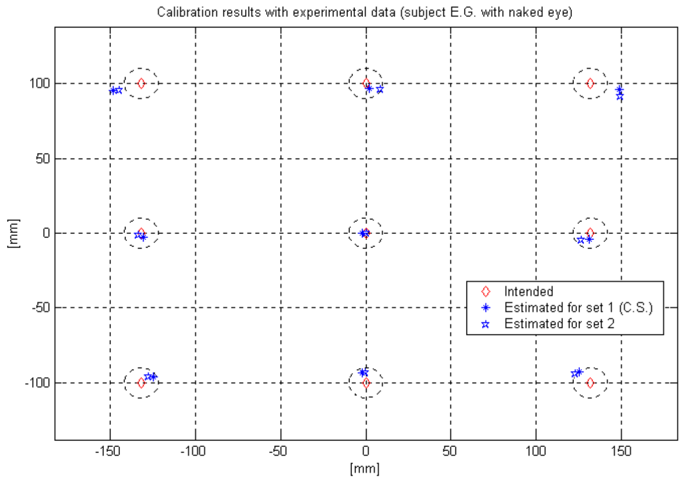


Figure 3.1: Calibration example with fixation data for subject E.G. without optical correction.

3.4 Sensitivity of the point-of-gaze estimation to errors in the eye parameters and some system parameters

In this section, simulations are used to study the sensitivity of the point-of-gaze estimation methodology to errors in the eye parameters and some system parameters. The parameters to be considered in this study are R, K, α_{eye}, β_{eye}, θ_{cam}, φ_{cam}, κ_{cam} and the typical value of the distance between the cornea and the nodal point of the camera (this value is used to calculate λ). The objective is to analyze the variation of the r.m.s. error in the point-of-gaze estimation as a function of each of these parameters considered individually.

For R, K, α_{eye}, β_{eye}, θ_{cam}, φ_{cam} and κ_{cam}, the sensitivity study is done as follows:

(i) For a given position of the center of rotation of the eye, **d**, and a set of points of gaze on the screen $\{\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_N\}$, the coordinates of the pupil and glints centers in the image of the eye are calculated using equation **(2.67)** as $\mathbf{f}_i = \mathbf{f}(\mathbf{g}_i, \mathbf{d}, \boldsymbol{\rho}')$, $i = 1, 2, \ldots, N$, where $\boldsymbol{\rho}'$ is the vector of eye and system parameters with the typical/nominal values given in Appendix B, except for the parameter that we want to study, which varies over a certain interval.

(ii) The point-of-gaze is estimated with equation **(2.38)** as $\hat{\mathbf{g}}_i = \hat{\mathbf{g}}(\mathbf{f}_i, \boldsymbol{\rho}_\circ)$, $i = 1, 2, \ldots, N$, where $\boldsymbol{\rho}_\circ$ is the vector of eye and system parameters with the typical/nominal values given in Appendix B.

(iii) The r.m.s. error in the estimation of the point-of-gaze is evaluated as

$$\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left\|\hat{\mathbf{g}}_i-\mathbf{g}_i\right\|^2}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left\|\hat{\mathbf{g}}(\mathbf{f}(\mathbf{g}_i,\mathbf{d},\boldsymbol{\rho}'),\boldsymbol{\rho}_\circ)-\mathbf{g}_i\right\|^2}\,. \tag{3.3}$$

The position of the center of rotation of the eye **d** = $[0\ 70\ 650]^T$ mm and the 9 points of gaze listed in table 3.3 are considered[4].

i	g_X (mm)	g_Y (mm)	g_Z (mm)
1	-130	100	0
2	0	100	0
3	130	100	0
4	-130	0	0
5	0	0	0
6	130	0	0
7	-130	-100	0
8	0	-100	0
9	130	-100	0

Table 3.3: Coordinates of the points of gaze on the computer monitor used for the simulations.

For the intervals considered, the r.m.s. error, as a function of the magnitude of the deviation of each parameter from its typical/nominal value, increases in a slightly nonlinear manner for R (Fig. 3.2) and almost linearly for K, α_{eye}, β_{eye}, θ_{cam}, φ_{cam} and κ_{cam}. The sensitivity relative to each parameter, considered individually, is summarized in table 3.4. Particularly, this table shows that the sensitivity to errors in φ_{cam} is small.

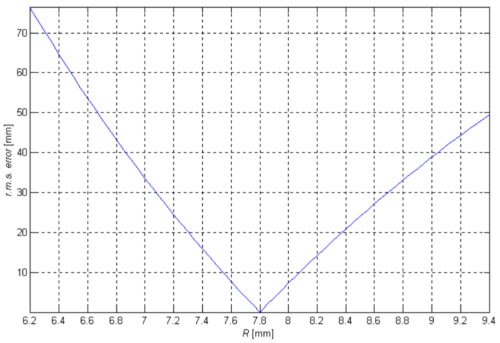


Figure 3.2: R.m.s. error in the estimation of the point-of-gaze as a function of the radius of corneal curvature.

The study of the sensitivity of the point-of-gaze estimation error to changes in the distance between the cornea and the nodal point of the camera (*dcnp* –this value is used to

---

[4] For **d** = $[0\ 70\ 650]^T$ mm the distance between the cornea and the nodal point of the camera is approximately 625 mm.

compute λ) is carried out in a similar way with the exception that the position of the center of rotation of the eye, **d**, varies. More specifically, consider the situation in which the center of rotation of the eye adopts different positions $\mathbf{d}_j$, $j = 1, \ldots, M$ along the optic axis of the camera (the lens is refocused in order to keep a sharp image) but a fixed typical value of the *dcnp* of 625 mm is used to estimate the point-of-gaze. For each point $\mathbf{d}_j$, the *dcnp* and the r.m.s. error in the estimation of the point-of-gaze can be computed and used to describe the r.m.s. error as a function of the deviation of the actual *dcnp* from the typical value. As shown at the bottom of table 3.4, the estimation of the point-of-gaze is robust to error in the estimate of the distance between the cornea and the nodal point of the camera.

Parameter	Sensitivity of the r.m.s. error in the estimation of the point-of-gaze to the deviation from the typical value	Interval considered, Typical / nominal value
R	~53 mm/mm or ~41 mm/10% for 0.8 $R_{typ}<R<0.9\,R_{typ}$, ~42 mm/mm or ~33 mm/10% for 0.9 $R_{typ}<R<R_{typ}$, ~34 mm/mm or ~26.5 mm/10% for $R_{typ}<R<1.1\,R_{typ}$, ~28 mm/mm or ~22 mm/10% for 1.1 $R_{typ}<R<1.2\,R_{typ}$	[6.2,9.4] mm or ~[–20,20] %, R_{typ} = 7.8 mm
K	~59 mm/mm or ~28 mm/10% for $K<K_{typ}$, ~63 mm/mm or ~30 mm/10% for $K>K_{typ}$	[3.8,5.7] mm or [–20,20] %, K_{typ} = 4.75 mm
α_{eye}	~11.6 mm/degree	[–7,–3] °, $\alpha_{eye,\,typ} = -5°$
β_{eye}	~11.7 mm/degree	[–0.5,3.5] °, $\beta_{eye,\,typ} = 1.5°$
θ_{cam}	~3.4 mm/degree	[–3,3] °, $\theta_{cam,\,nom} = 0°$
φ_{cam}	~0.6 mm/degree	[25,29] °, $\varphi_{cam,\,nom} = 27°$
κ_{cam}	~7 mm/degree	[–2,2] °, $\kappa_{cam,\,nom} = 0°$
dcnp	~0.31 mm/10 cm for $dcnp< dcnp_{typ}$, ~0.28 mm/10 cm for $dcnp> dcnp_{typ}$	[500,800] mm, $dcnp_{typ}$ = 625 mm

Table 3.4: Sensitivity of the estimation of the point-of-gaze relative to different parameters.

In order to gain further insight into the effects of errors in the above parameters and the interactions among them, the patterns of the estimation error can be studied through simulations. This is done in a similar way to that described above but in this case the parameter of interest adopts a single value rather than varying over an interval. The parameters considered here are R, K, α_{eye}, β_{eye}, θ_{cam} and κ_{cam}.

- Error pattern relative to *R*

Suppose that the actual value of *R* (radius of corneal curvature) is 5 % larger than the typical value used to estimate the point-of-gaze. The estimates of the point-of-gaze for this case, as well as the actual points of gaze are shown in Fig. 3.3 (the dashed circles have 10 mm of radius). As it can be seen in this figure, *R* has a combination of offset and gain effects. In this example, the offset is approximately –1.4 mm in the horizontal direction and –11.8 mm in the vertical direction, while the gain is very close to 95 % of the correct gain in both directions. In other words, if the offset error were removed, the remaining error is an underestimation error of 5 %.

An intuitive explanation for this reduction in gain can be given as follows. For a fixed position of the center of rotation of the eye, if the radius of corneal curvature increases, the distance between glint centers in the eye image also increases in the same proportion. As the eye rotates, the pupil center moves relative to the glints. Since the distance between glints increased, any given movement of the pupil center appears to be proportionally smaller relative to the distance between glints. This makes eye rotations seem smaller, explaining the reduced gain observed in Fig. 3.3. The opposite occurs if *R* is smaller.

- Error pattern relative to *K*

Suppose that the actual value of *K* (distance between the pupil center and the center of corneal curvature) is 5 % larger than the typical value used to estimate the point-of-gaze. The estimates of the point-of-gaze for this case, as well as the actual points of gaze are shown in Fig. 3.4. As it can be seen in this figure, *K* also has a combination of offset and gain effects. In this example, the offset is approximately 2.7 mm in the horizontal direction and 11.5 mm in the vertical direction, while the gain is close to 105 % of the correct gain in both directions. In other words, if the offset error were removed, the remaining error is an overestimation error of 5 %.

An intuitive explanation for the increment of gain can be given as follows. If *K* increases, so does the radius of rotation of the pupil center (D+K, see Fig. 2.4). This makes the movement of the pupil center larger than what would be expected if *K* had the typical value. In turn, this makes eye rotations seem larger, explaining the increased gain observed in Fig. 3.4. The opposite occurs if *K* is smaller.

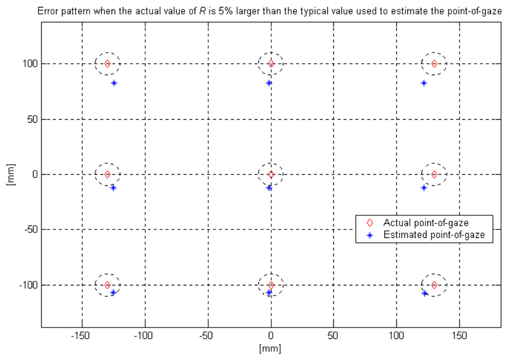


Figure 3.3: Error pattern when the actual value of *R* is 5 % larger than the typical value used to estimate the point of gaze.

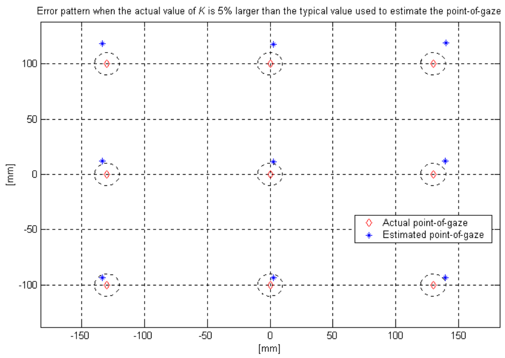


Figure 3.4: Error pattern when the actual value of *K* is 5 % larger than the typical value used to estimate the point of gaze.

Comparing this example with the previous one (Fig. 3.3), it can be observed that the deviations of the estimates of the point-of-gaze are approximately similar but in opposite directions. Fig. 3.5 shows that the r.m.s. error in the estimation of the point-of-gaze as a function of R and K has a valley where the error is small. In fact, if that figure is re-plotted as a function of

$$R_{\circ,norm} = \frac{R - R_{typ}}{R_{typ}} = \frac{R}{R_{typ}} - 1 \tag{3.4}$$

and

$$K_{\circ,norm} = \frac{K - K_{typ}}{K_{typ}} = \frac{K}{K_{typ}} - 1 \ , \tag{3.5}$$

where $R_{\circ,norm}$ and $K_{\circ,norm}$ represent the fractional change of R and K relative to their typical values, the valley is in fact at approximately 45° relative to the $R_{\circ,norm}$-axis. This means that variations of the same proportions and in the same direction in R and K tend to compensate for each other.

- Error patterns relative to α_{eye} and β_{eye}

A deviation in the actual value of α_{eye} (horizontal angle between the visual and optic axes) relative to the value used to estimate the point-of-gaze, produces, essentially, a pure offset in the horizontal direction. A deviation in the actual value of β_{eye} (vertical angle between the visual and optic axes) relative to the value used to estimate the point-of-gaze, produces, essentially, a pure offset in the vertical direction. These two parameters have an offset effect but they are decoupled from each other. Fig. 3.6 shows that the r.m.s. error as a function of α_{eye} and β_{eye} is cone-shaped. This is consistent with the observations made above.

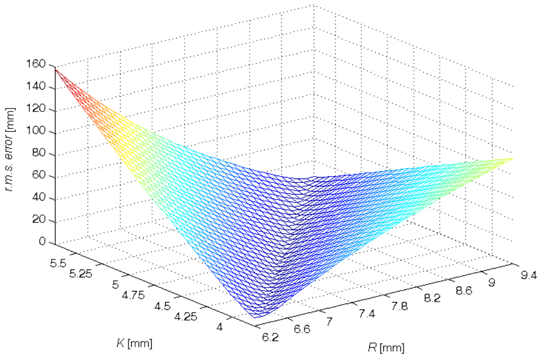


Figure 3.5: R.m.s. error in the estimation of the point-of-gaze as a function of R and K.

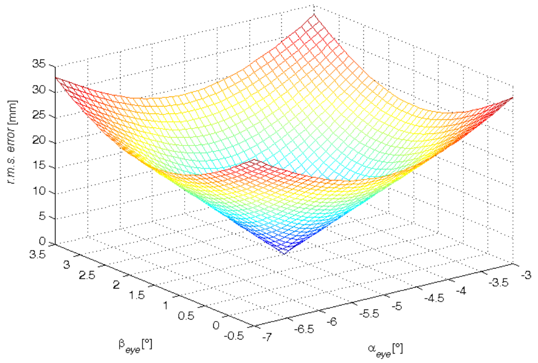


Figure 3.6: R.m.s. error in the estimation of the point-of-gaze as a function of α_{eye} and β_{eye}.

- Error pattern relative to θ_{cam}

Suppose that the actual value of θ_{cam} (pan angle of the camera) is 3° larger than the nominal value used to estimate the point-of-gaze. The estimates of the point-of-gaze for this case, as well as the actual points of gaze are shown in Fig. 3.7. As it can be seen in this figure, θ_{cam} has a mixture of offset and rotation-like effects. The offset is about –10.4 mm in the horizontal direction. The central row of the point-of-gaze estimates forms an angle of –0.762° with the horizontal axis and the central column of the point-of-gaze estimates forms an angle of –0.138° with the vertical axis. This observation indicates that the distortion of the pattern of point-of-gaze estimates would be more adequately modeled as an affine transformation rather than as a simple combination of rotation and translation (Euclidean transformation). Note that an error of 3° was simulated because with smaller errors it is more difficult to appreciate the distortion pattern.

- Error pattern relative to κ_{cam}

Suppose that the actual value of κ_{cam} (angle of rotation of the camera around its optic axis) is 1° larger than the nominal value used to estimate the point-of-gaze. The estimates of the point-of-gaze for this case, as well as the actual points of gaze are shown in Fig. 3.8. This figure shows that κ_{cam} has a mixture of rotation-like and offset effects. The offset is about –5.9 mm in the horizontal direction. The central row of the point-of-gaze estimates forms an angle of –1.264° with the horizontal axis and the central column of the point-of-gaze estimates forms an angle of –1.008° with the vertical axis. As in the previous example, this observation indicates that the distortion of the pattern of the point-of-gaze estimates would be more adequately modeled as an affine transformation rather than as a simple combination of rotation and translation. Note that θ_{cam} and κ_{cam} have effects of similar nature and can be considered somewhat coupled. Fig. 3.9 shows that the r.m.s. error in the estimation of the point-of-gaze as a function of θ_{cam} and κ_{cam} has a valley where the error is small, which means that along that valley the effects of errors in θ_{cam} and κ_{cam} compensate for each other to some extent. The orientation of that valley can be described as $\Delta\kappa_{cam} / \Delta\theta_{cam} \cong -0.47$.

If the calibration were carried out only for the eye parameters, the offset error could be compensated for by α_{eye}, but the rotation-like error introduced by θ_{cam} and κ_{cam} could not be compensated. Fig. 3.10 shows that the r.m.s. error in the estimation of the point-of-gaze as a function of α_{eye} and θ_{cam} also presents a valley where α_{eye} compensates to some extent for the deviation of θ_{cam}. The orientation of that valley can be described as $\Delta\theta_{cam} / \Delta\alpha_{eye} \cong -3.43$.

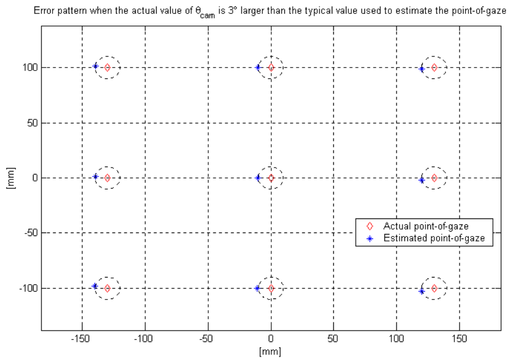


Figure 3.7: Error pattern when the actual value of θ_{cam} is 3° larger than the nominal value used to estimate the point-of-gaze.

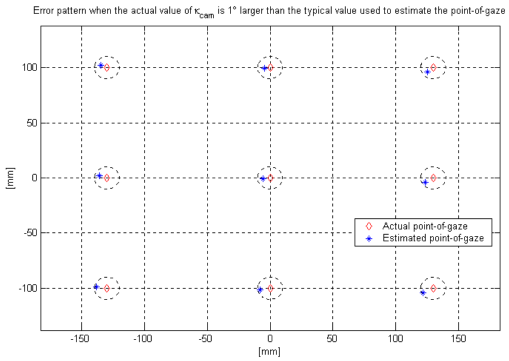


Figure 3.8: Error pattern when the actual value of κ_{cam} is 1° larger than the nominal value used to estimate the point-of-gaze.

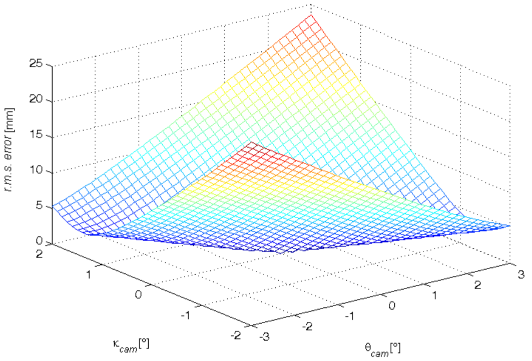


Figure 3.9: R.m.s. error in the estimation of the point-of-gaze as a function of θ_{cam} and κ_{cam}.

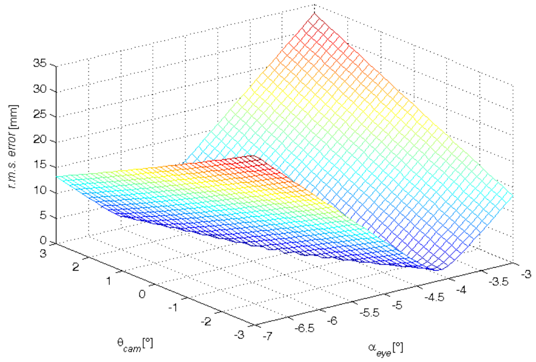


Figure 3.10: R.m.s. error in the estimation of the point-of-gaze as a function of α_{eye} and θ_{cam}.

3.5 Compensation for errors in system parameters that are not calculated by the calibration procedure

The purpose of this section is to show how the calibration procedure can compensate for errors in system parameters that are not calculated by the calibration algorithm. The parameters to be considered are the positions of the light sources and the position of the center of the camera image plane.

This study is done as follows:

(i) For a given position of the center of rotation of the eye, $\mathbf{d}_1$, and a set of points of gaze on the screen $\{\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_N\}$, the coordinates of pupil and glints centers on the image of the eye are calculated using equation **(2.67)** as $\mathbf{f}_i = \mathbf{f}(\mathbf{g}_i, \mathbf{d}_1, \boldsymbol{\rho}')$, $i = 1, 2, \ldots, N$, where $\boldsymbol{\rho}'$ is the vector of eye and system parameters with the typical/nominal values given in Appendix B, except for the parameter that is studied, which takes a different value, and the value of the distance between the cornea and the nodal point of the camera, which is calculated for the particular value of **d**.

(ii) Using the vectors $\mathbf{f}_i$, a calibration (equation **(3.2)**) is performed, obtaining the vector of optimum calibration parameters, $\boldsymbol{\rho}_{cal}$. The elements of the vector of fixed system parameters, $\boldsymbol{\rho}_{fixed}$, used for the calibration, have the nominal values given in Appendix B. In this case, the vector of fixed system parameters is represented by $\boldsymbol{\rho}_{fixed,\circ}$.

(iii) For a given set of positions of the center of rotation of the eye $\{\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_j, \ldots, \mathbf{d}_M\}$ and a set of points of gaze on the screen $\{\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_i, \ldots, \mathbf{g}_N\}$, the coordinates of the pupil and glints centers on the image of the eye are calculated using equation **(2.67)** as $\mathbf{f}_{i,j} = \mathbf{f}(\mathbf{g}_i, \mathbf{d}_j, \boldsymbol{\rho}')$, $i = 1, 2, \ldots, N$, $j = 1, 2, \ldots, M$.

(iv) The point-of-gaze is estimated with equation **(2.38)** as $\hat{\mathbf{g}}_{i,j} = \hat{\mathbf{g}}(\mathbf{f}_{i,j}, [\boldsymbol{\rho}_{cal} \;\; \boldsymbol{\rho}_{fixed,\circ}])$, $i = 1, 2, \ldots, N$, $j = 1, 2, \ldots, M$.

(v) The r.m.s. error in the estimation of the point-of-gaze is evaluated for the position of the center of rotation of the eye adopted for the calibration ($\mathbf{d}_1$) as

$$\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left\|\hat{\mathbf{g}}_{i,1} - \mathbf{g}_i\right\|^2} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left\|\hat{\mathbf{g}}(\mathbf{f}(\mathbf{g}_i, \mathbf{d}_1, \boldsymbol{\rho}'), [\boldsymbol{\rho}_{cal} \;\; \boldsymbol{\rho}_{fixed,\circ}]) - \mathbf{g}_i\right\|^2} \tag{3.6}$$

and for all the tested positions of the center of rotation $\{\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_M\}$ as

$$\sqrt{\frac{1}{MN}\sum_{j=1}^{M}\sum_{i=1}^{N}\left\|\hat{\mathbf{g}}_{i,j} - \mathbf{g}_i\right\|^2} = \sqrt{\frac{1}{MN}\sum_{j=1}^{M}\sum_{i=1}^{N}\left\|\hat{\mathbf{g}}(\mathbf{f}(\mathbf{g}_i, \mathbf{d}_j, \boldsymbol{\rho}'), [\boldsymbol{\rho}_{cal} \;\; \boldsymbol{\rho}_{fixed,\circ}]) - \mathbf{g}_i\right\|^2} \;. \tag{3.7}$$

The values of $\mathbf{g}_i$ used for the simulations are listed in table 3.3, while the values of $\mathbf{d}_j$ are listed in table 3.5. In these simulations, errors of 10 mm in the measurements of the position of one of the light sources (**m**) and errors of 5 mm in the estimate of the position of the center of the camera image plane (**t**) were evaluated. The results obtained are summarized in table 3.6, where the estimation error before calibration (U, uncalibrated) is given for comparison with the estimation error after calibration (C, calibrated). Note that even though the errors considered here are very large compared to the accuracy of the measurements of those parameters, the remaining r.m.s. error in the point-of-gaze estimation after calibration is very small.

The point-of-gaze estimation error due to errors in the measurements of the coordinates of a light source can be large before calibration but can be reduced dramatically with calibration. In the case of errors in the estimate of the coordinates of the center of the camera image plane, the estimation error is relatively small even without calibration but can be made significantly smaller with calibration.

These results and the low sensitivity of the estimation of the point-of-gaze relative to φ_{cam} and the distance between the cornea and the nodal point of the camera (used to calculate

λ), show that the measurement accuracy of the fixed system parameters is not a critical factor for the performance of the methodology developed in Chapter 2.

j	d_X (mm)	d_Y (mm)	d_Z (mm)
1	0	70	650
2	-30	95	650
3	0	95	650
4	30	95	650
5	-30	70	650
6	30	70	650
7	-30	45	650
8	0	45	650
9	30	45	650
10	-25	65	600
11	0	65	600
12	25	65	600
13	-25	45	600
14	0	45	600
15	25	45	600
16	-25	25	600
17	0	25	600
18	25	25	600
19	-35	125	700
20	0	125	700
21	35	125	700
22	-35	100	700
23	0	100	700
24	35	100	700
25	-35	75	700
26	0	75	700
27	35	75	700

Table 3.5: Positions of the center of rotation of the eye used for the simulations.

Situation	C/U	r.m.s. error for $\mathbf{d}_1$ [mm]	r.m.s. error for $\{\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_{27}\}$ [mm]
$m_X = m_{X,nom} + 10\,\text{mm}$	C	0.1589	0.2274
	U	5.9563	5.9563
$m_Y = m_{Y,nom} + 10\,\text{mm}$	C	0.3017	0.8214
	U	9.7012	9.7057
$m_Z = m_{Z,nom} + 10\,\text{mm}$	C	0.1517	0.3233
	U	4.8434	4.8648
$t_X = t_{X,nom} + 5\,\text{mm}$	C	0.1548	0.1865
	U	1.0556	1.0687
$t_Y = t_{Y,nom} + 5\,\text{mm}$	C	0.0546	0.2132
	U	0.7547	0.7804
$t_Z = t_{Z,nom} + 5\,\text{mm}$	C	0.0274	0.1810
	U	0.3678	0.4113

Table 3.6: Estimation error before and after calibration due to deviation of the values of some system parameters that are not adjusted by the calibration procedure.

3.6 Effects of errors in the estimation of the coordinates of pupil and glints centers in the eye images

The purpose of this section is to illustrate the effects that errors in the estimation of the coordinates of pupil and glints centers in the eye images have on the point-of-gaze estimation. In order to do this, the vectors of coordinates of pupil and glints centers $\hat{\mathbf{f}}_i$, i = 1, 2, …, 9, which are used in the calibration procedure, are computed as

$$\begin{gathered}\hat{\mathbf{f}}_i = \frac{1}{100}\sum_{j=1}^{100}\mathbf{f}_{i,j} \ , \\ \mathbf{f}_{i,j} = \mathbf{f}(\mathbf{g}_i, \mathbf{d}, \boldsymbol{\rho}_\circ) + \Delta\mathbf{f}_{i,j}\end{gathered} \tag{3.8}$$

where $\mathbf{d}$=[0 70 650]^T mm, $\mathbf{g}_i$, i = 1, 2, …, 9 are listed in table 3.3, and $\boldsymbol{\rho}_\circ$ is the vector of system and eye parameters with the typical/nominal values given in Appendix B. The noise is defined as

$$\Delta\mathbf{f}_{i,j} = [\Delta r_{left\ glint,i,j} \quad \Delta c_{left\ glint,i,j} \quad \Delta r_{pupil\ center,i,j} \quad \Delta c_{pupil\ center,i,j} \quad \Delta r_{right\ glint,i,j} \quad \Delta c_{right\ glint,i,j}] \tag{3.9}$$

where $\Delta r_{left\ glint,i,j}$, $\Delta r_{pupil\ center,i,j}$ and $\Delta r_{right\ glint,i,j}$ are independent Gaussian random variables with zero-mean and standard deviation of σ = 0.06 pixel, while $\Delta c_{left\ glint,i,j}$, $\Delta c_{pupil\ center,i,j}$ and $\Delta c_{right\ glint,i,j}$ are independent Gaussian random variables with zero-mean and standard deviation of σ = 0.09 pixel. The characteristics of the above random variables are based on experimental data. The image processing algorithms used to estimate the coordinates of pupil and glints centers in the eye images and the associated estimation errors are discussed in [47].

These simulation conditions correspond to a stationary ideal eye with spherical cornea that fixates exactly on the intended points of gaze, but the coordinates of pupil and glints centers that are computed from each eye image suffer from an estimation error. This estimation error is modeled as additive zero-mean Gaussian noise.

Fig. 3.11 shows an example of a simulation carried out as explained above. The calibration was performed for the vectors of coordinates of pupil and glints centers $\hat{\mathbf{f}}_i$, i = 1, …, 9 and the point-of-gaze was estimated for each vector of coordinates of pupil and glints centers $\mathbf{f}_{i,j}$, i = 1, …, 9, j = 1, …, 100 (equation **(3.8)**). The residual r.m.s. error of the calibration was 0.1302 [mm] for this example.

This simulation shows that if the estimation errors of the coordinates of pupil and glints centers have zero-mean, the estimated calibration parameters do not induce biases to the estimation of the point-of-gaze. This is due to the fact that the averaging of 100 estimates greatly reduces the effect of the estimation error of pupil and glints centers. In addition, for noise with the properties used above, all the point-of-gaze estimates fall inside a circle of 10 mm of radius (about 0.9º of visual angle for the simulated eye position) centered on the actual point-of-gaze.

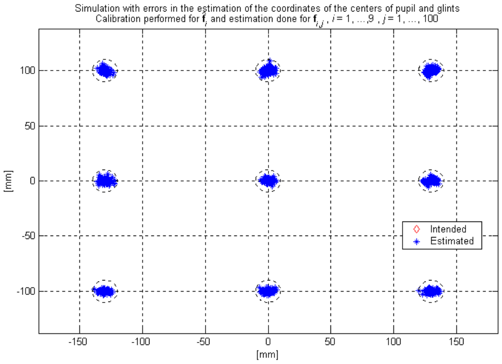


Figure 3.11: Point-of-gaze estimation after calibration, for estimation errors in the coordinates of pupil and glints centers.

3.7 Effects of fixation errors during the calibration procedure

The purpose of this section is to consider the effects of fixation errors during the calibration procedure. The errors in the fixation points can be broadly categorized as large errors and small errors. A large error corresponds to the case in which the subject does not fixate on the point that is presented on the screen during the data collection stage of the calibration procedure, i.e. the subject fixates at a point that is more than 10 mm away from the requested point. Large errors are in general obvious and in practice lead to the repetition of the data collection. Large errors are not considered here.

Small errors are due to small eye movements that are associated with normal fixation. If we assume that these eye movements can result in the fixation target being projected on any part of the fovea, these eye movements are in a range 0.6-1º of visual angle. The projection of this area of position uncertainty onto a computer screen, at 65 cm from the eye, has a diameter that can be as large as 11.3 mm, which means that the error between the actual fixation point and the point presented on the screen could be up to 5.65 mm. In order to simulate this situation, the vectors of coordinates of pupil and glints centers $\hat{\mathbf{f}}_i$, i = 1, 2, …, 9, that are used for the calibration procedure, are computed as

$$\hat{\mathbf{f}}_i = \mathbf{f}(\mathbf{g}_i + \Delta\mathbf{g}_i, \mathbf{d}, \boldsymbol{\rho}_\circ) \ , \qquad \textbf{(3.10)}$$

where **d**=[0 70 650]^T mm, $\mathbf{g}_i$, i = 1, 2, …, 9 are listed in table 3.3, and $\boldsymbol{\rho}_\circ$ is the vector of eye and system parameters with the typical/nominal values given in Appendix B. The errors in the fixation points are modeled as

$$\Delta\mathbf{g}_i = [r_i \cos\phi_i \quad r_i \sin\phi_i \quad 0]^T \qquad \textbf{(3.11)}$$

where r_i is a random variable with uniform distribution over the interval [0,5] mm and ϕ_i is a random variable with uniform distribution over the interval [0,2π) radians.

Using this set of vectors of coordinates of pupil and glints centers $\hat{\mathbf{f}}_i$, i = 1, 2, …, 9, a calibration is carried out and a vector of optimum calibration parameters is obtained. This procedure is repeated 10 times. Fig. 3.12 shows the calibration results for each data set (realization of the simulation) and table 3.7 lists the residual r.m.s. error for each case. Fig. 3.13 shows the estimation results for all ten data sets using the calibration parameters obtained for data set 1. In this case the overall r.m.s. error is ~3.17 mm.

These simulations show that in these conditions the residual r.m.s. error of the calibration for each of the 10 data sets has a relatively small value of less than 3.4 mm or about 0.3º of visual angle. In addition, it shows that if the calibration parameters obtained for a particular data set are used to estimate the point-of-gaze for all data sets, the estimates for the different data sets are consistent with each other and the overall r.m.s. estimation error remains relatively small (~3.17 mm).

In order to study the expected point-of-gaze estimation error for larger errors in the fixation points, the same type of simulation is also carried out when r_i (equation **(3.11)**) is uniformly distributed over the interval [0,10] mm. The calibration results for each data set are illustrated in Fig. 3.14 while the corresponding residual r.m.s. errors are summarized in table 3.8. The estimation results for all ten data sets using the calibration parameters obtained for data set 1 are depicted in Fig. 3.15, where the overall r.m.s. error is ~7.36 mm.

In this case, as expected, the accuracy of the calibration is poorer. The residual r.m.s. error of the calibration for each of the 10 data sets can be up to 7.5 mm or 0.66º of visual angle. When the calibration parameters obtained for a particular data set are used to estimate the point-of-gaze for all data sets, the estimates for the different data sets are less consistent with each other and the overall r.m.s. estimation error increases to ~7.36 mm. This implies that precautions should be adopted to make the fixation errors no larger than 5 mm. For this reason, as mentioned in Section 3.2, a visual stimulus consisting of a circle that reduces its diameter and collapses on itself is used as a calibration target with the purpose of making the fixation as accurate and stable as possible.

	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	Set 7	Set 8	Set 9	Set 10
Rms error [mm]	~3.39	~1.77	~2.63	~2.02	~2.62	~2.11	~1.76	~2.66	~1.58	~2.04

Table 3.7: Residual r.m.s. error of the calibration for each data set when r_i is a uniform random variable over the interval [0,5] mm.

	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	Set 7	Set 8	Set 9	Set 10
Rms error [mm]	~5.02	~5.34	~5.83	~4.77	~4.49	~7.49	~4.80	~5.46	~6.50	~3.77

Table 3.8: Residual r.m.s. error of the calibration for each data set when r_i is a uniform random variable over the interval [0,10] mm.

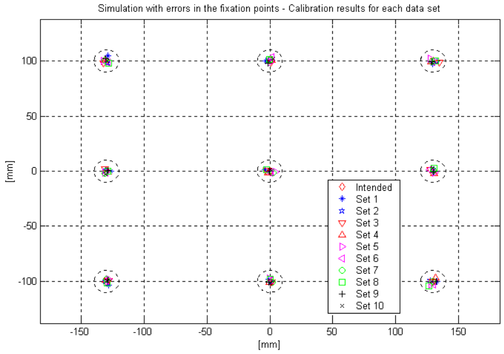


Figure 3.12: Calibration and estimation results for each data set when r_i is a uniform random variable over the interval [0,5] mm.

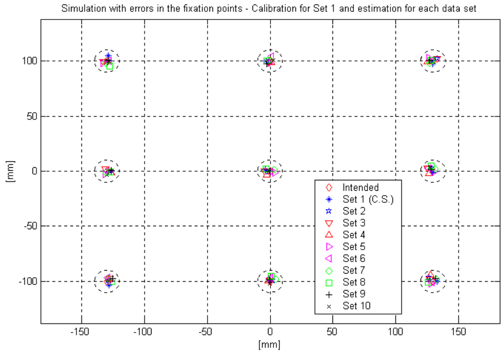


Figure 3.13: Estimation results for each data set using the calibration parameters obtained for data set 1 (C.S.) when r_i is a uniform random variable over the interval [0,5] mm.

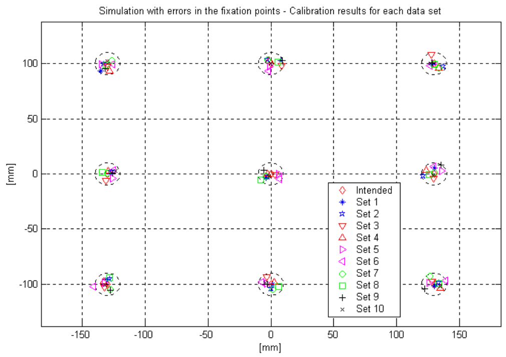


Figure 3.14: Calibration and estimation results for each data set when r_i is a uniform random variable over the interval [0,10] mm.

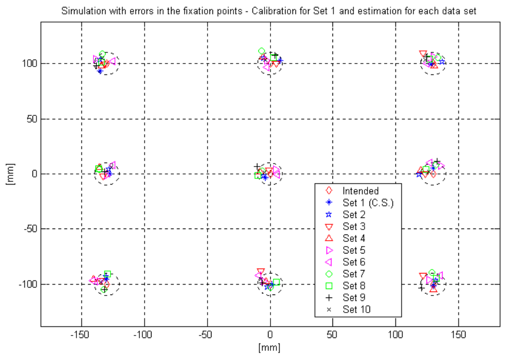


Figure 3.15: Estimation results for each data set using the calibration parameters obtained for data set 1 (C.S.) when r_i is a uniform random variable over the interval [0,10] mm.

The results presented in Sections 3.4 thru 3.7 provide insights into the effects of different types of errors on the estimation of the point-of-gaze. Errors in model parameters were studied in Sections 3.4 and 3.5, showing that accurate measurement of the fixed system parameters is not critical for the performance of the point-of-gaze estimation methodology. In particular, it was seen that the sensitivity to measurement error of the tilt angle of the camera (φ_{cam}) and to errors in the estimation of the distance between the cornea and the nodal point of the camera (used to calculate λ) is very small. It was also shown that the calibration procedure can compensate for measurement errors in the positions of the light sources (**l** and **m**) and the position of the center of the camera image plane (**t**). The actual values of the eye parameters (R, K, α_{eye} and β_{eye}) and the extrinsic camera parameters θ_{cam} and κ_{cam} are estimated accurately by the calibration procedure if other sources of error are absent.

Section 3.6 analyzed the effect of estimation errors in the coordinates of pupil and glints centers in the eye images. It was shown that if the estimation errors have zero-mean, then averaging over a large number (e.g. 100) of estimates for each calibration point can significantly reduce the effect of that estimation error and the estimated calibration parameters do not induce biases to the estimation of the point-of-gaze. This Section gave examples of the effects of fixation errors during the calibration procedure on the point-of-gaze estimation. It was observed that if the fixation errors are reasonable small (less than 5 mm), as expected in a standard calibration procedure, the r.m.s. error of the point-of-gaze estimation on a computer monitor, at 65 cm from the subject, can be expected to be about 0.3º.

Despite the analysis of different sources of errors, none of the cases considered above can explain the error pattern seen in Fig. 3.1. The reason for such a pattern can be categorized as a modeling error and is studied in the next chapter, where solutions to improve the calibration and estimation accuracy are proposed.

4. Estimation of the point-of-gaze for real eyes

4.1 Introduction

An analysis of data of glints coordinates obtained experimentally (see Appendix D for sample eye images) showed that as subjects scan the whole computer screen, the angle given by the inverse tangent of the slope of the line defined by the centers of the two glints (equation **(4.1)**) can vary within a range of 5 or 6º, while for a simulated spherical cornea this range is less than 0.03º. Note that according to Fig. 2.2(c) and equation **(4.1)**, that angle is measured clockwise from the *c*-axis.

$$slope = \frac{r_{right\ glint} - r_{left\ glint}}{c_{right\ glint} - c_{left\ glint}} \ . \tag{4.1}$$

This observation suggested that the spherical approximation to the corneal surface might not be appropriate and might lead to significant errors when the point-of-gaze is calculated. This could be categorized as a modeling error. It is known [48] that the average corneal surface has a radius of curvature of 7.8 mm at the apex increasing to 10 mm at the cornea-sclera boundary. When individual cases are considered, the range of variation of the radius of curvature of the cornea can be greater. This asphericity of the cornea was not taken into account in the model presented in Chapter 2 and it might help to explain the error patterns observed in Fig. 3.1.

In order to verify this hypothesis, a non-spherical corneal model was developed. This model is used to study the effects of corneal asphericity on the accuracy of the point-of-gaze estimation methodology developed in Chapter 2, and to evaluate the performance of algorithms that improve the estimation accuracy.

The development of the non-spherical corneal model and the equations to compute the coordinates of pupil and glints centers for that model are presented in the next Section. In Section 4.3 simulation results using the proposed non-spherical corneal model are shown and compared to experimental results. In Section 4.4, sub-optimal solutions to improve the point-of-gaze estimation accuracy are proposed and their performance is evaluated with simulated and experimental data. Finally, Section 4.5 illustrates the effects of errors in the estimation of the coordinates of pupil and glints centers in the eye images, and gives the overall r.m.s. estimation error of the point-of-gaze estimation system computed from experimental data.

4.2 Non-spherical corneal model

To calculate the coordinates of pupil and glints centers in an eye image for non-spherical corneal models, given a position of the center of rotation of the eye, **d**, and a point-of-gaze, **g**, on the computer screen, the methodology of Section 2.8 needs to be modified, as it was developed specifically for a spherical corneal model. This section describes a generalization for non-spherical corneal models.

4.2.1 Corneal surface model

Towards the development of a non-spherical corneal model, it is necessary to define an eye coordinate system. This coordinate system is a right-handed 3D Cartesian coordinate system whose axes are represented by x_{eye}, y_{eye} and z_{eye}, its origin is at the center of rotation of the eye, the z_{eye}-axis is coincident with the optic axis of the eye, and the x_{eye}-axis is always perpendicular to the Y-axis of the world coordinate system.

In this development, the corneal surface is modeled as a 10th degree polynomial surface of the form

$$z_{eye} = a_{10}\tau^{10} + a_8\tau^8 + a_6\tau^6 + a_4\tau^4 + a_2\tau^2 + a_0 \ , \qquad \textbf{(4.2)}$$

where τ is a function of x_{eye} and y_{eye}. A 10th degree polynomial was chosen as a trade-off between model flexibility and model complexity. If $\tau^2 = x_{eye}^2 + y_{eye}^2$, a surface of revolution is obtained. However, a more general model can be obtained if $\tau^2 = \gamma_1 x_{eye}^2 + \gamma_2 x_{eye} y_{eye} + \gamma_3 y_{eye}^2$ so that the cross-sections of the corneal surface perpendicular to the optic axis are ellipses instead of circles. The coefficients γ_1, γ_2 and γ_3 can be expressed as functions of the ratio of the lengths of the major axis to the minor axis and the orientation of the major axis of those ellipses. Towards this end, consider the following equation of an ellipse whose axes are coincident with some coordinate axes x_o and y_o:

$$\frac{x_o^2}{A^2}+\frac{y_o^2}{B^2}=1 \qquad \textbf{(4.3)}$$

subject to the constraint

$$AB=1 \qquad \textbf{(4.4)}$$

so that the area remains constant regardless of the values of *A* and *B*. Assume that *A* is the half-length of the major axis while *B* is the half-length of the minor axis. The ratio of the lengths of the major axis to the minor axis is represented by

$$\xi=\frac{A}{B}\ ,\ \xi\geq 1\,. \qquad \textbf{(4.5)}$$

Equations **(4.4)** and **(4.5)** can be solved for *A* and *B*, so that equation **(4.3)** is rewritten as

$$\frac{x_o^2}{\xi}+\xi\,y_o^2=1\ . \qquad \textbf{(4.6)}$$

Next, consider a rotation of the x_oy_o coordinate system relative to the $x_{eye}y_{eye}$ coordinate system by an angle δ as shown in Fig. 4.1. This angle δ represents the orientation of the major axis of the ellipse with respect to the x_{eye}-axis. The x_oy_o and the $x_{eye}y_{eye}$ coordinate systems are related by

$$\begin{aligned} x_o &= x_{eye}\cos\delta + y_{eye}\sin\delta \\ y_o &= y_{eye}\cos\delta - x_{eye}\sin\delta \end{aligned}\ . \qquad \textbf{(4.7)}$$

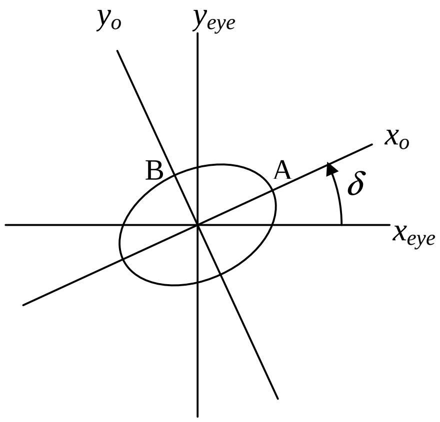


Figure 4.1: Rotation of the x_oy_o coordinate system relative to the $x_{eye}y_{eye}$ coordinate system.

Substituting equation **(4.7)** into equation **(4.6)**, an equation of the form

$$\gamma_1 x_{eye}^2 + \gamma_2 x_{eye} y_{eye} + \gamma_3 y_{eye}^2 = 1 \tag{4.8}$$

is obtained, where

$$\gamma_1 = \frac{\cos^2\delta}{\xi} + \xi\sin^2\delta \ , \tag{4.9}$$

$$\gamma_2 = 2\cos\delta\sin\delta\left(\frac{1}{\xi} - \xi\right) \ , \tag{4.10}$$

$$\gamma_3 = \frac{\sin^2\delta}{\xi} + \xi\cos^2\delta \ . \tag{4.11}$$

In this way, the corneal surface can be described in the eye coordinate system by an implicit equation of the form

$$C_{eye}(x_{eye}, y_{eye}, z_{eye}, \mathbf{a}, \xi, \delta) = 0 \, , \tag{4.12}$$

such that

$$\begin{gathered} C_{eye}(x_{eye}, y_{eye}, z_{eye}, \mathbf{a}, \xi, \delta) = z_{eye} - (a_{10}\tau^{10} + a_8\tau^8 + a_6\tau^6 + a_4\tau^4 + a_2\tau^2 + a_0) \, , \\ \tau^2 = \gamma_1 x_{eye}^2 + \gamma_2 x_{eye} y_{eye} + \gamma_3 y_{eye}^2 \end{gathered} \tag{4.13}$$

where $\mathbf{a} = [a_0 \quad a_2 \quad a_4 \quad a_6 \quad a_8 \quad a_{10}]$ and the coefficients γ_1, γ_2 and γ_3 are given by equations **(4.9)** thru **(4.11)** as a function of the elliptical cross section parameters ξ and δ.

In order to be able to calculate the coordinates of pupil and glints centers in an image of an eye whose corneal surface is described by equation **(4.12)**, it is necessary to describe the corneal surface in the world coordinate system for any position of the center of rotation of the eye, **d**, and any orientation of the optic axis described by the pan angle θ_{eye} and the tilt angle φ_{eye} (Fig. 2.4). In this case, the corneal surface will be described by an equation of the form

$$C_{world}(X, Y, Z, \mathbf{d}, \theta_{eye}, \varphi_{eye}, \mathbf{a}, \xi, \delta) = 0 \ . \tag{4.14}$$

The transformation between the world coordinate system and eye coordinate system can be described by the following sequence of steps:

(i) Translate the world coordinate system *XYZ* by **d** to obtain a coordinate system $X_\mathbf{d} Y_\mathbf{d} Z_\mathbf{d}$. Note that – $Z_\mathbf{d}$ points to the screen.

(ii) Rotate the coordinate system $X_\mathbf{d}Y_\mathbf{d}Z_\mathbf{d}$ around the $Y_\mathbf{d}$-axis by an angle θ_{eye} to obtain a coordinate system $X_\mathbf{d}'Y_\mathbf{d}'Z_\mathbf{d}'$ (Fig. 4.2).

(iii) Rotate the coordinate system $X_\mathbf{d}'Y_\mathbf{d}'Z_\mathbf{d}'$ around the $X_\mathbf{d}'$-axis by an angle φ_{eye} to obtain a coordinate system $X_\mathbf{d}''Y_\mathbf{d}''Z_\mathbf{d}''$ (Fig. 4.3). At this point, the $Z_\mathbf{d}''$-axis is coincident with the optic axis of the eye with $-Z_\mathbf{d}''$ pointing to the screen.

(iv) Obtain the eye coordinate system $x_{eye}y_{eye}z_{eye}$ by taking $x_{eye} = -X_\mathbf{d}''$, $y_{eye} = Y_\mathbf{d}''$ and $z_{eye} = -Z_\mathbf{d}''$ (this is equivalent to a rotation of 180º around the $Y_\mathbf{d}''$-axis). In this way z_{eye} points to the screen.

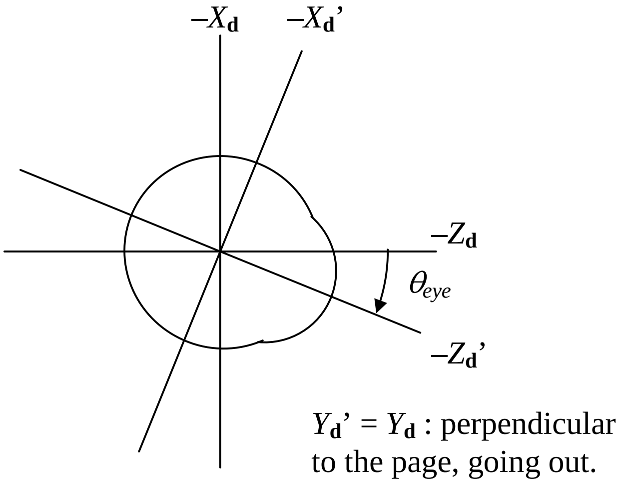


Figure 4.2: Rotation of the coordinate system $X_\mathbf{d}Y_\mathbf{d}Z_\mathbf{d}$ around the $Y_\mathbf{d}$-axis by an angle θ_{eye}.

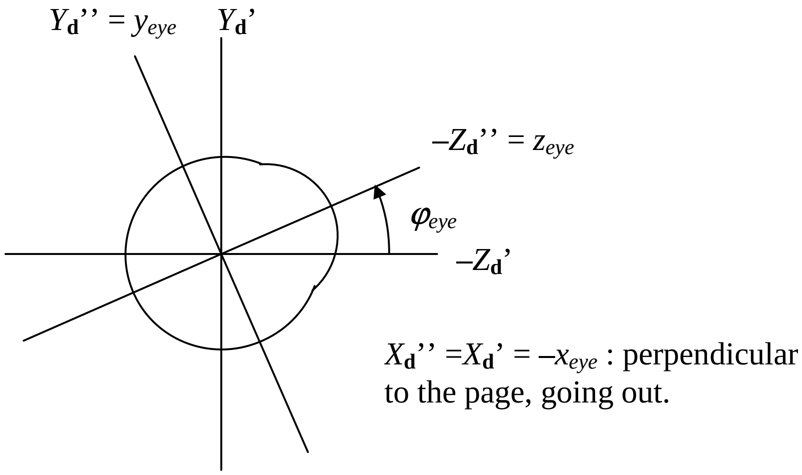


Figure 4.3: Rotation of the coordinate system $X_\mathbf{d}'Y_\mathbf{d}'Z_\mathbf{d}'$ around the $X_\mathbf{d}'$-axis by an angle φ_{eye}.

Step (i) is described as

$$\begin{bmatrix} X_{\mathbf{d}} \\ Y_{\mathbf{d}} \\ Z_{\mathbf{d}} \end{bmatrix} = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} - \mathbf{d} \ . \tag{4.15}$$

Step (ii) is described as

$$\begin{bmatrix} X_{\mathbf{d}}' \\ Y_{\mathbf{d}}' \\ Z_{\mathbf{d}}' \end{bmatrix} = \begin{bmatrix} \cos\theta_{eye} & 0 & \sin\theta_{eye} \\ 0 & 1 & 0 \\ -\sin\theta_{eye} & 0 & \cos\theta_{eye} \end{bmatrix} \begin{bmatrix} X_{\mathbf{d}} \\ Y_{\mathbf{d}} \\ Z_{\mathbf{d}} \end{bmatrix} . \tag{4.16}$$

Step (iii) is described as

$$\begin{bmatrix} X_{\mathbf{d}}'' \\ Y_{\mathbf{d}}'' \\ Z_{\mathbf{d}}'' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi_{eye} & \sin\varphi_{eye} \\ 0 & -\sin\varphi_{eye} & \cos\varphi_{eye} \end{bmatrix} \begin{bmatrix} X_{\mathbf{d}}' \\ Y_{\mathbf{d}}' \\ Z_{\mathbf{d}}' \end{bmatrix} . \tag{4.17}$$

Step (iv) is described as

$$\begin{bmatrix} x_{eye} \\ y_{eye} \\ z_{eye} \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} X_{\mathbf{d}}'' \\ Y_{\mathbf{d}}'' \\ Z_{\mathbf{d}}'' \end{bmatrix} . \tag{4.18}$$

Equations **(4.15)** thru **(4.18)** can be chained to obtain

$$\begin{bmatrix} x_{eye} \\ y_{eye} \\ z_{eye} \end{bmatrix} = \mathbf{J} \left(\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} - \mathbf{d} \right) \tag{4.19}$$

where

$$\mathbf{J} = \begin{bmatrix} -\cos\theta_{eye} & 0 & -\sin\theta_{eye} \\ -\sin\theta_{eye}\sin\varphi_{eye} & \cos\varphi_{eye} & \cos\theta_{eye}\sin\varphi_{eye} \\ \sin\theta_{eye}\cos\varphi_{eye} & \sin\varphi_{eye} & -\cos\theta_{eye}\cos\varphi_{eye} \end{bmatrix} . \tag{4.20}$$

By substituting equations **(4.19)** and **(4.20)** into equation **(4.12)**, equation **(4.14)** is obtained with

$$C_{world}(X, Y, Z, \mathbf{d}, \theta_{eye}, \varphi_{eye}, \mathbf{a}, \xi, \delta) = C_{eye}(x_{eye}, y_{eye}, z_{eye}, \mathbf{a}, \xi, \delta) \ , \ \begin{bmatrix} x_{eye} \\ y_{eye} \\ z_{eye} \end{bmatrix} = \mathbf{J} \left(\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} - \mathbf{d} \right) . \tag{4.21}$$

Having this corneal surface model, the calculation of the coordinates of pupil and glints centers in eye images is considered next.

4.2.2 Calculation of pupil and glints centers in eye images

Using this non-spherical corneal model, it is possible to calculate the coordinates of pupil and glints centers in eye images for any position of the center of rotation of the eye, **d**, and any point-of-gaze on the computer screen, **g**. In order to do this, part of the formulation of Section 2.8 has to be modified. From the values of **d** and **g**, the pan and tilt angles of the optic axis of the eye θ_{eye} and φ_{eye} are calculated using the iterative procedure described by equations **(2.43)** thru **(2.51)**. Note that even though **c** can no longer be considered as the center of corneal curvature, it still has physical meaning as the approximation to the nodal point of the eye (intersection of the optic and visual axes of the eye, Fig. 2.5). Using the computed values of θ_{eye} and φ_{eye}, the points of reflection **q** and **s** on the non-spherical corneal surface are computed with the equations that are developed next. Once **q** and **s** have been computed, the rest of the calculations to obtain the coordinates of pupil and glints centers in the image of the eye are carried out as explained in Section 2.8.

In order to compute the points of reflection **q** and **s**, the same principles of specular reflection considered in Chapter 2 are applied here. For the point of reflection **q**, the fact that it is on the corneal surface can be expressed as

$$C_{world}(q_X, q_Y, q_Z, \mathbf{d}, \theta_{eye}, \varphi_{eye}, \mathbf{a}, \xi, \delta) = C_{world}(\mathbf{q}, \mathbf{d}, \theta_{eye}, \varphi_{eye}, \mathbf{a}, \xi, \delta) = 0 \ . \qquad \mathbf{(4.22)}$$

The direction of the normal at the point of reflection **q** is given by the gradient $\nabla C_{world}(\mathbf{q}, \mathbf{d}, \theta_{eye}, \varphi_{eye}, \mathbf{a}, \xi, \delta)$. Then, the fact that the incident ray, the reflected ray and the normal at the point of reflection are coplanar can be described as

$$(\mathbf{l}-\mathbf{q})\times(\mathbf{o}-\mathbf{q})\bullet\nabla C_{world}(\mathbf{q}, \mathbf{d}, \theta_{eye}, \varphi_{eye}, \mathbf{a}, \xi, \delta) = 0$$

or equivalently as

$$(\mathbf{l}-\mathbf{o})\times(\mathbf{q}-\mathbf{o})\bullet\nabla C_{world}(\mathbf{q}, \mathbf{d}, \theta_{eye}, \varphi_{eye}, \mathbf{a}, \xi, \delta) = 0 \ . \qquad \mathbf{(4.23)}$$

The remaining condition is that the incidence and reflection angles are equal, which can be written, after rearranging terms, as

$$(\mathbf{l}-\mathbf{q})\bullet\nabla C_{world}(\mathbf{q}, \mathbf{d}, \theta_{eye}, \varphi_{eye}, \mathbf{a}, \xi, \delta)\cdot\|\mathbf{o}-\mathbf{q}\| - (\mathbf{o}-\mathbf{q})\bullet\nabla C_{world}(\mathbf{q}, \mathbf{d}, \theta_{eye}, \varphi_{eye}, \mathbf{a}, \xi, \delta)\cdot\|\mathbf{l}-\mathbf{q}\| = 0 \qquad \mathbf{(4.24)}$$

The system of equations **(4.22)** thru **(4.24)** is solved numerically for **q**. For faster convergence, a convenient starting point for this numerical solution can be taken as the intersection of the optic axis with the corneal surface:

$$\mathbf{d} + a_0 \begin{bmatrix} \cos\varphi_{eye} \sin\theta_{eye} \\ \sin\varphi_{eye} \\ -\cos\varphi_{eye} \cos\theta_{eye} \end{bmatrix} . \qquad \mathbf{(4.25)}$$

For the point of reflection **s**, the following equations result:

$$C_{world}(s_X, s_Y, s_Z, \mathbf{d}, \theta_{eye}, \varphi_{eye}, \mathbf{a}, \xi, \delta) = C_{world}(\mathbf{s}, \mathbf{d}, \theta_{eye}, \varphi_{eye}, \mathbf{a}, \xi, \delta) = 0 \ , \qquad \mathbf{(4.26)}$$

$$(\mathbf{m} - \mathbf{o}) \times (\mathbf{s} - \mathbf{o}) \bullet \nabla C_{world}(\mathbf{s}, \mathbf{d}, \theta_{eye}, \varphi_{eye}, \mathbf{a}, \xi, \delta) = 0 \ , \qquad \mathbf{(4.27)}$$

and

$$(\mathbf{m} - \mathbf{s}) \bullet \nabla C_{world}(\mathbf{s}, \mathbf{d}, \theta_{eye}, \varphi_{eye}, \mathbf{a}, \xi, \delta) \cdot \|\mathbf{o} - \mathbf{s}\| - (\mathbf{o} - \mathbf{s}) \bullet \nabla C_{world}(\mathbf{s}, \mathbf{d}, \theta_{eye}, \varphi_{eye}, \mathbf{a}, \xi, \delta) \cdot \|\mathbf{m} - \mathbf{s}\| = 0 \qquad \mathbf{(4.28)}$$

This system of equations is solved numerically for **s**, taking **(4.25)** as a starting point. Once **q** and **s** have been computed, the coordinates of the glints centers in the eye image are calculated as in Section 2.8 using equations **(2.65)** and **(2.66)**. The computation of the pupil center **p** and its image is also done as in Chapter 2 using equations **(2.11)** and **(2.64)**.

When the points of reflection **q** and **s** are computed, it is necessary to confirm that they are in the region where the model of the corneal surface is valid, i.e. they do not fall on the region that corresponds to the sclera of a real eye. A test to validate the results is to compute the distance between each point of reflection and the optic axis of the eye and make sure that it is less that some value τ_{MAX}, which is typically about 5 mm to ensure that the point of reflection is not too close to what corresponds to the cornea-sclera boundary. For the point of reflection **q**, this test can be formally expressed as

$$\sqrt{\|\mathbf{q} - \mathbf{d}\|^2 - \left[(\mathbf{q} - \mathbf{d}) \bullet \mathbf{k}_{eye}\right]^2} \le \tau_{MAX} \qquad \mathbf{(4.29)}$$

or

$$\sqrt{\|\mathbf{q} - \mathbf{c}\|^2 - \left[(\mathbf{q} - \mathbf{c}) \bullet \mathbf{k}_{eye}\right]^2} \le \tau_{MAX} \qquad \mathbf{(4.30)}$$

where

$$\mathbf{k}_{eye} = \begin{bmatrix} \cos\varphi_{eye}\sin\theta_{eye} \\ \sin\varphi_{eye} \\ -\cos\varphi_{eye}\cos\theta_{eye} \end{bmatrix} \tag{4.31}$$

is the unit vector in the direction of the optic axis of the eye (Fig. 2.4). For the point of reflection **s**, equation **(4.29)** or **(4.30)** is used with **q** substituted by **s**.

For the sake of completeness, the computation of the gradient $\nabla C_{world}(X,Y,Z,\mathbf{d},\theta_{eye},\varphi_{eye},\mathbf{a},\xi,\delta)$, that is used in the above equations, is described next. Towards this end, note that

$$C_{world}(X,Y,Z,\mathbf{d},\theta_{eye},\varphi_{eye},\mathbf{a},\xi,\delta) = C_{eye}(x_{eye}(X,Y,Z), y_{eye}(X,Y,Z), z_{eye}(X,Y,Z),\mathbf{a},\xi,\delta) \ . \tag{4.32}$$

Applying the definition of the gradient and the chain rule for differentiation, and writing in matrix form, the following expression can be obtained

$$\underbrace{\begin{bmatrix} \frac{\partial C_{world}}{\partial X} \\ \frac{\partial C_{world}}{\partial Y} \\ \frac{\partial C_{world}}{\partial Z} \end{bmatrix}}_{\nabla C_{world}} = \begin{bmatrix} \frac{\partial x_{eye}}{\partial X} & \frac{\partial y_{eye}}{\partial X} & \frac{\partial z_{eye}}{\partial X} \\ \frac{\partial x_{eye}}{\partial Y} & \frac{\partial y_{eye}}{\partial Y} & \frac{\partial z_{eye}}{\partial Y} \\ \frac{\partial x_{eye}}{\partial Z} & \frac{\partial y_{eye}}{\partial Z} & \frac{\partial z_{eye}}{\partial Z} \end{bmatrix} \underbrace{\begin{bmatrix} \frac{\partial C_{eye}}{\partial x_{eye}} \\ \frac{\partial C_{eye}}{\partial y_{eye}} \\ \frac{\partial C_{eye}}{\partial z_{eye}} \end{bmatrix}}_{\nabla C_{eye}} \ . \tag{4.33}$$

From equation **(4.19)** it can be shown that

$$\begin{bmatrix} \frac{\partial x_{eye}}{\partial X} & \frac{\partial y_{eye}}{\partial X} & \frac{\partial z_{eye}}{\partial X} \\ \frac{\partial x_{eye}}{\partial Y} & \frac{\partial y_{eye}}{\partial Y} & \frac{\partial z_{eye}}{\partial Y} \\ \frac{\partial x_{eye}}{\partial Z} & \frac{\partial y_{eye}}{\partial Z} & \frac{\partial z_{eye}}{\partial Z} \end{bmatrix} = \mathbf{J}^T \ , \tag{4.34}$$

hence

$$\nabla C_{world}(X,Y,Z,\mathbf{d},\theta_{eye},\varphi_{eye},\mathbf{a},\xi,\delta) = \mathbf{J}^T \nabla C_{eye}(x_{eye},y_{eye},z_{eye},\mathbf{a},\xi,\delta) \ , \ \begin{bmatrix} x_{eye} \\ y_{eye} \\ z_{eye} \end{bmatrix} = \mathbf{J}\left(\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} - \mathbf{d}\right) \tag{4.35}$$

Following, the gradient $\nabla C_{eye}(x_{eye},y_{eye},z_{eye},\mathbf{a},\xi,\delta)$ is considered. Note that equation **(4.13)** can be rewritten as

$$\begin{array}{c} C_{eye}(x_{eye}, y_{eye}, z_{eye}, \mathbf{a}, \xi, \delta) = z_{eye} - (a_{10}\tau_s^5 + a_8\tau_s^4 + a_6\tau_s^3 + a_4\tau_s^2 + a_2\tau_s + a_0), \\ \tau_s = \gamma_1 x_{eye}^2 + \gamma_2 x_{eye} y_{eye} + \gamma_3 y_{eye}^2 \end{array} . \qquad \textbf{(4.36)}$$

Applying the chain rule,

$$\nabla C_{eye}(x_{eye}, y_{eye}, z_{eye}, \mathbf{a}, \xi, \delta) = \begin{bmatrix} \dfrac{\partial C_{eye}}{\partial x_{eye}} \\ \dfrac{\partial C_{eye}}{\partial y_{eye}} \\ \dfrac{\partial C_{eye}}{\partial z_{eye}} \end{bmatrix} = \begin{bmatrix} \dfrac{\partial C_{eye}}{\partial \tau_s} \dfrac{\partial \tau_s}{\partial x_{eye}} \\ \dfrac{\partial C_{eye}}{\partial \tau_s} \dfrac{\partial \tau_s}{\partial y_{eye}} \\ 1 \end{bmatrix} \qquad \textbf{(4.37)}$$

where

$$\frac{\partial C_{eye}}{\partial \tau_s} = -(5a_{10}\tau_s^4 + 4a_8\tau_s^3 + 3a_6\tau_s^2 + 2a_4\tau_s + a_2) \ , \qquad \textbf{(4.38)}$$

$$\frac{\partial \tau_s}{\partial x_{eye}} = 2\gamma_1 x_{eye} + \gamma_2 y_{eye} \qquad \textbf{(4.39)}$$

and

$$\frac{\partial \tau_s}{\partial y_{eye}} = \gamma_2 x_{eye} + 2\gamma_3 y_{eye} \ . \qquad \textbf{(4.40)}$$

So, by substituting equations **(4.38)** thru **(4.40)** into equation **(4.37)**, and then substituting equation **(4.37)** into equation **(4.35)**, $\nabla C_{world}(X, Y, Z, \mathbf{d}, \theta_{eye}, \varphi_{eye}, \mathbf{a}, \xi, \delta)$ is obtained.

4.2.3 Model coefficients

Having defined the non-spherical corneal model, different vectors of model coefficients $\mathbf{a} = [a_0 \quad a_2 \quad a_4 \quad a_6 \quad a_8 \quad a_{10}]$ are proposed, as detailed in table 4.1.

	a_{10}	a_8	a_6	a_4	a_2	a_0
Model 1	$-4.1278 \cdot 10^{-10}$	$-1.4064 \cdot 10^{-9}$	$-1.7073 \cdot 10^{-6}$	$-2.2036 \cdot 10^{-4}$	$-6.4103 \cdot 10^{-2}$	13.1
Model 2	0	0	0	$-1.5000 \cdot 10^{-4}$	$-6.4103 \cdot 10^{-2}$	13.1
Model 3	$-5.7374 \cdot 10^{-10}$	$-1.9193 \cdot 10^{-7}$	$-2.4078 \cdot 10^{-6}$	$-4.1837 \cdot 10^{-4}$	$-6.0976 \cdot 10^{-2}$	13.1

Table 4.1: Selected coefficients for the non-spherical corneal models.

Model 1 was obtained from least squares fitting to the ellipsoidal corneal model presented in [48]. Model 2 is a simple fourth order model whose coefficients were adjusted so that the radius of curvature is 7.8 mm at the apex and about 12.5 mm at 6 mm from the optic axis. This model exhibits a larger variation of the radius of curvature than model 1. Model 3 was adjusted by least squares fitting to obtain a desired curvature function inspired on experimental results obtained for a subject wearing contact lenses. Note that in all cases, the value of the coefficient a_0, which represents the distance between the center of rotation of the eye and the apex of the cornea, is $a_0 = D_{typ} + R_{typ}$, where D_{typ} and R_{typ} are given in Appendix B. Comparative plots of the three models and a circular section with $R = R_{typ} = 7.8\,\text{mm}$ are given in Fig. 4.4.

It can be shown that the radius of curvature function for equation **(4.2)** is

$$\rho(\tau) = \frac{\left[1 + (10a_{10}\tau^9 + 8a_8\tau^7 + 6a_6\tau^5 + 4a_4\tau^3 + 2a_2\tau)^2\right]^{3/2}}{\left|90a_{10}\tau^8 + 56a_8\tau^6 + 30a_6\tau^4 + 12a_4\tau^2 + 2a_2\right|} \quad . \qquad \textbf{(4.41)}$$

The radius of curvature functions for the corneal models plotted in Fig. 4.4 are graphed in Fig. 4.5. Note that the corneal models are assumed valid only for $|\tau| \le 6\,\text{mm}$ since the width of the cornea in an adult eye is about 12-13 mm.

It can be observed that even though the longitudinal corneal sections look pretty similar for $|\tau| \le 4\,\text{mm}$, the radius of curvature function differs notably for each model.

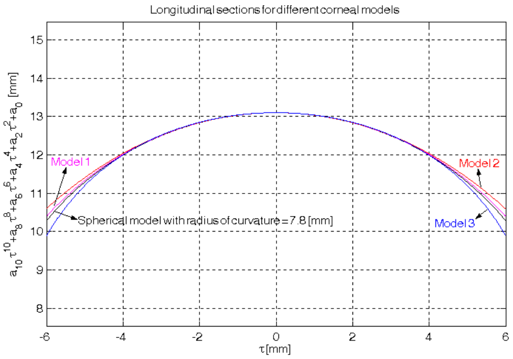


Figure 4.4: Plot of different longitudinal corneal sections.

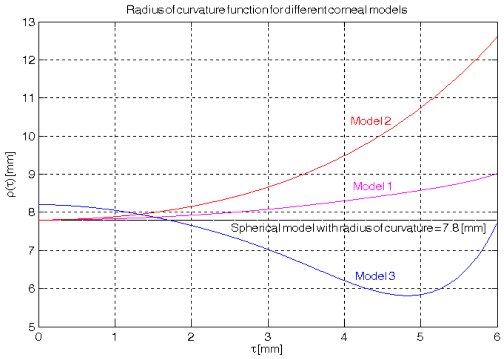


Figure 4.5: Radius of curvature functions for the different corneal models.

4.3 Simulations and experimental results

4.3.1 Simulations

Using the non-spherical corneal models proposed in the previous section, the effects of corneal asphericity on the point-of-gaze estimation accuracy are studied through simulations. The computation of the coordinates of pupil and glints centers in the camera image (**f**) can be described as a function

$$\mathbf{f} = \mathbf{f}_{non-spherical}(\mathbf{g}, \mathbf{d}, [\boldsymbol{\rho}_{eye,\, non-spherical\; cornea} \quad \boldsymbol{\rho}_{system}]) \tag{4.42}$$

where **d** is the position of the center of rotation of the eye, **g** is the point-of-gaze on the computer monitor,

$$\boldsymbol{\rho}_{eye,\, non-spherical\; cornea} = [\mathbf{a} \quad \xi \quad \delta \quad D \quad K \quad \alpha_{eye} \quad \beta_{eye}] \tag{4.43}$$

is the vector of parameters of the eye with non-spherical cornea, and

$$\boldsymbol{\rho}_{system} = [\theta_{cam} \quad \varphi_{cam} \quad \kappa_{cam} \quad \mathbf{l} \quad \mathbf{m} \quad \mathbf{t} \quad focal\_length \quad dcnp] \tag{4.44}$$

is the vector of system parameters. For the simulations of this section, the last four elements of $\boldsymbol{\rho}_{eye,\, non-spherical\; cornea}$ and all the elements of $\boldsymbol{\rho}_{system}$, take the typical / nominal values given in Appendix B, except for the value of the distance between the cornea and the nodal point of the camera (*dcnp*) which is calculated for each particular value of **d**.

The simulations are carried out as follows:

(i) For a given position of the center of rotation of the eye $\mathbf{d}_1$ and a set of points of gaze on the screen $\{\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_N\}$, the coordinates of the centers of pupil and glints on the image of the eye are calculated as $\mathbf{f}_i = \mathbf{f}_{non-spherical}(\mathbf{g}_i, \mathbf{d}_1, [\boldsymbol{\rho}_{eye,\, non-spherical\; cornea} \quad \boldsymbol{\rho}_{system}])$, $i = 1, 2, \ldots, N$.

(ii) Using these vectors $\mathbf{f}_i$, a calibration is performed (equation **(3.2)**), obtaining the vector of optimum values of the calibration parameters $\boldsymbol{\rho}_{cal}$. The elements of the vector of fixed system parameters, $\boldsymbol{\rho}_{fixed}$, used for the calibration, have the nominal values given in Appendix B and is represented here as $\boldsymbol{\rho}_{fixed,\circ}$.

(iii) For a given set of positions of the center of rotation of the eye $\{\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_j, \ldots, \mathbf{d}_M\}$ and a set of points of gaze on the screen $\{\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_i, \ldots, \mathbf{g}_N\}$, the coordinates of pupil and glints centers on the image of the eye are calculated as $\mathbf{f}_{i,j} = \mathbf{f}_{non-spherical}(\mathbf{g}_i, \mathbf{d}_j, [\boldsymbol{\rho}_{eye,\, non-spherical\; cornea} \quad \boldsymbol{\rho}_{system}])$, $i = 1, 2, \ldots, N$, $j = 1, 2, \ldots, M$.

(iv) The point-of-gaze is estimated with equation **(2.38)** for all the $\mathbf{f}_{i,j}$, as $\hat{\mathbf{g}}_{i,j} = \hat{\mathbf{g}}(\mathbf{f}_{i,j},[\boldsymbol{\rho}_{cal} \;\; \boldsymbol{\rho}_{fixed,\circ}])$, i = 1, 2, …, N, j = 1, 2, …, M.

(v) The r.m.s. error in the estimation of the point-of-gaze is evaluated for the position of the center of rotation of the eye adopted for the calibration ($\mathbf{d}_1$) as

$$\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left\|\hat{\mathbf{g}}_{i,1}-\mathbf{g}_i\right\|^2} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left\|\hat{\mathbf{g}}(\mathbf{f}_{non-spherical}(\mathbf{g}_i,\mathbf{d}_1,[\boldsymbol{\rho}_{eye,\,non-spherical\,cornea} \;\; \boldsymbol{\rho}_{system}]),[\boldsymbol{\rho}_{cal} \;\; \boldsymbol{\rho}_{fixed,\circ}])-\mathbf{g}_i\right\|^2} \tag{4.45}$$

and for all the tested positions of the center of rotation {$\mathbf{d}_1$, $\mathbf{d}_2$, …, $\mathbf{d}_M$} as

$$\sqrt{\frac{1}{MN}\sum_{j=1}^{M}\sum_{i=1}^{N}\left\|\hat{\mathbf{g}}_{i,j}-\mathbf{g}_i\right\|^2} = \sqrt{\frac{1}{MN}\sum_{j=1}^{M}\sum_{i=1}^{N}\left\|\hat{\mathbf{g}}(\mathbf{f}_{non-spherical}(\mathbf{g}_i,\mathbf{d}_j,[\boldsymbol{\rho}_{eye,\,non-spherical\,cornea} \;\; \boldsymbol{\rho}_{system}]),[\boldsymbol{\rho}_{cal} \;\; \boldsymbol{\rho}_{fixed,\circ}])-\mathbf{g}_i\right\|^2} \tag{4.46}$$

The values of $\mathbf{g}_i$ used for the simulations are listed in table 3.3, while the values of $\mathbf{d}_j$ are listed in table 3.5. The coefficients of the non-spherical corneal models are given in table 4.1.

Fig. 4.6 corresponds to model 1 with ξ = 1, where the dashed circles centered at the intended points of gaze have a radius of 10 mm. It can be observed that there is a small underestimation error in the vertical direction and the estimates of the point-of-gaze have a pattern that opens upwards/closes downwards: there is overestimation error in the horizontal direction when the gaze is above the screen center and underestimation error below the screen center. This pattern is not very prominent but is clearly visible.

Fig. 4.7 corresponds to model 2 with ξ = 1 and it can be observed that the estimates of the point-of-gaze also have a pattern that opens upwards/closes downwards. This pattern is much more prominent than that seen for model 1 and is consistent with the fact that the variation of corneal curvature is considerably larger for model 2 than for model 1. In addition, a larger dispersion of the estimates of the point-of-gaze for different positions of the center of rotation of the eye is observed.

Fig. 4.8 corresponds to model 2 with ξ = 1.02 and δ = 45º. The purpose of this example is to consider a case with corneal astigmatism, more specifically a case in which the transversal sections of the corneal surface, perpendicular to the optic axis of the eye, are elliptical. The pattern of the estimates is somewhat similar to that shown in Fig. 4.7.

Fig. 4.9 corresponds to model 3 with $\xi = 1$. In this case, there is overestimation error in the vertical direction and the estimates have a pattern that opens downwards/closes upwards: there is underestimation error in the horizontal direction when the gaze is above the screen center and overestimation below the screen center. This pattern, which is inverted relative to the patterns observed in the previous examples, is pretty prominent and is consistent with the corresponding radius of curvature function plotted in Fig. 4.5. The r.m.s. errors of the point-of-gaze estimation for each case are given in table 4.2.

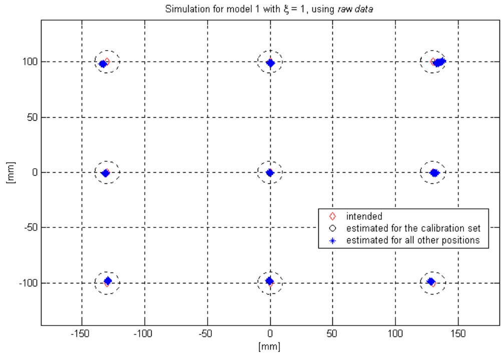


Figure 4.6: Simulation results for model 1 with $\xi = 1$.

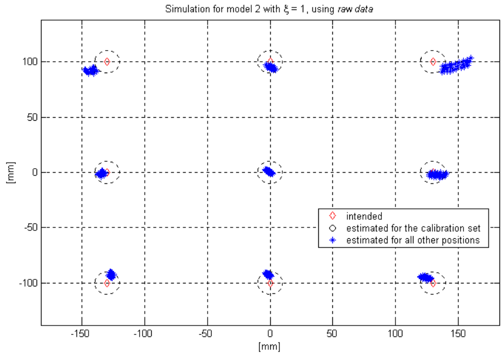


Figure 4.7: Simulation results for model 2 with $\xi = 1$.

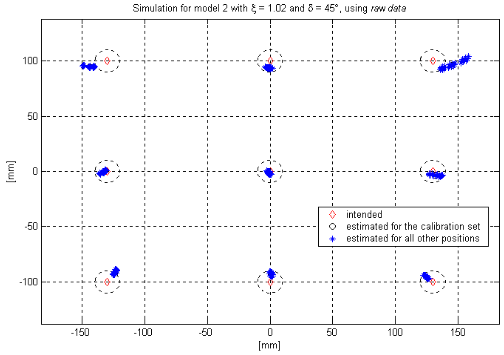


Figure 4.8: Simulation results for model 2 with $\xi = 1.02$ and $\delta = 45°$.

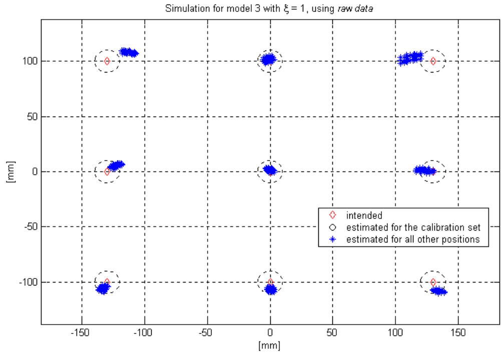


Figure 4.9: Simulation results for model 3 with $\xi = 1$.

Model	Data set(s)	R.m.s. error [mm]
Model 1 $\xi = 1$	Calibration set	2.6320
	All positions	2.8228
Model 2 $\xi = 1$	Calibration set	9.3210
	All positions	10.1439
Model 2 $\xi = 1.02$, $\delta = 45^{\circ}$	Calibration set	9.2449
	All positions	10.0194
Model 3 $\xi = 1$	Calibration set	10.4445
	All positions	10.9931

Table 4.2: R.m.s. error of the estimation of the point-of-gaze for simulations for different non-spherical corneal models.

In order to gain insight regarding the behavior of the glints positions relative to each other for each non-spherical corneal model and to be able to compare this information to that obtained in real experimental situations, the following descriptors are computed:

(i) The slope of the line defined by the centers of the two glints for each intended point-of-gaze, $\mathbf{g}_i$, for the calibration set, as defined in equation **(4.1)**.

(ii) The range (length of the interval) of variation of the slopes computed in (i) expressed in degrees as

$$range\ of\ \ slope\ variation\ in\ angle\ units = \arctan(\max_i slope_i) - \arctan(\min_i slope_i)\ . \qquad \textbf{(4.47)}$$

(iii) The mean distance between glints centers.

(iv) The standard deviation of the distance between glints centers normalized by the mean distance between glints centers (expressed as a percentage).

The slope of the line defined by the centers of the two glints for each intended point-of-gaze $\mathbf{g}_i$ for the calibration set is shown in table 4.3. The range of variation of those slopes, the mean distance between glints centers and the normalized standard deviation, are recorded in table 4.4.

It can be noted that model 1 and model 2, both with $\xi = 1$, present similar patterns of variation of the slope of the line defined by the glints centers since for both models the radius of curvature increases with the distance from the optic axis. However, model 2 shows a greater range of variation of the slope, as well as a greater variation of the distance between glints. This is consistent with the greater variation of the curvature of model 2 relative to model 1.

Model 2 with $\xi = 1.02$ and $\delta = 45^{\mathrm{o}}$ (model with corneal astigmatism) has similar ranges of variation of the slope and distance between glints but the interval of variation of the slope is shifted relative to that of model 2 with $\xi = 1$. This is explained by the ellipticity of the corneal section perpendicular to the optic axis having an orientation given by the angle δ.

Model 3, with $\xi = 1$, shows a pattern of variation of the slope of the line defined by the glints centers that is essentially inverted (opposite sign) relative to that produced by model 2 with $\xi = 1$. This is consistent with the decrease in the radius of curvature of the corneal surface in a neighborhood of the corneal apex for model 3, as opposed to the monotonic increase in the radius of curvature of the corneal surface with the distance from the optic axis for model 2. As mentioned earlier, model 3 was inspired by experimental results observed for a subject wearing contact lenses. Note that as illustrated in Figs. 4.4 and 4.5, small deviations of the corneal shape in the region of the cornea that produces the glints (roughly, less than 5.5 mm from the optic axis) can produce important changes in the curvature of the cornea. Model 3 has ranges of variation of the slope and the distance between glints that are similar to model 2.

$\mathbf{g}_i$ / Model	$\mathbf{g}_1$	$\mathbf{g}_2$	$\mathbf{g}_3$	$\mathbf{g}_4$	$\mathbf{g}_5$	$\mathbf{g}_6$	$\mathbf{g}_7$	$\mathbf{g}_8$	$\mathbf{g}_9$
Model 1, $\xi = 1$	-0.0075	0.0059	0.0182	-0.0047	0.0037	0.0114	-0.0021	0.0016	0.0048
Model 2, $\xi = 1$	-0.0259	0.0203	0.0666	-0.0145	0.0112	0.0369	-0.0059	0.0044	0.0142
Model 2, $\xi = 1.02$, $\delta = 45^{\mathrm{o}}$	-0.0452	0.0010	0.0481	-0.0338	-0.0081	0.0184	-0.0252	-0.0150	-0.0043
Model 3, $\xi = 1$	0.0280	-0.0220	-0.0678	0.0191	-0.0149	-0.0461	0.0089	-0.0068	-0.0196

Table 4.3: Slopes of the line defined by the glints centers for each intended point-of-gaze for the calibration set for the different simulated non-spherical corneal models.

Descriptor Model	Slope range [º]	Mean distance glints [pixels]	Normalized SD [%]
Spherical model, R = 7.8 mm	0.0244	22.1258	0.0568
Model 1, $\xi = 1$	1.4708	22.4951	0.85
Model 2, $\xi = 1$	5.2946	23.3296	3.21
Model 2, $\xi = 1.02$, $\delta = 45º$	5.3372	23.3166	3.11
Model 3, $\xi = 1$	5.4833	21.7410	3.40

Table 4.4: Range of variation of the slopes shown in table 4.3, mean distance between glints centers, and normalized standard deviation of the distance between glints centers, corresponding to the calibration set, for the different simulated non-spherical corneal models. Results for a spherical corneal model are provided for reference.

4.3.2 Experimental results

This subsection describes experimental results and compares them with the simulation results from the previous subsection. In each experiment, the data collection procedure described in Section 3.2, as part of the calibration procedure, was carried out twice for 5 different head positions (i.e. different positions of the center of rotation of the eye). The eye positions for the 5 head positions are described in table 4.5. For each head position, the subject fixated on the 9 points on the computer screen that are given in table 3.2. The head position was stabilized with a chinrest, which means that the head was not perfectly still but its movement was limited to a range that is much smaller than the distance between any two of the 5 different head positions. In this way, 10 sets of 9 vectors of average coordinates of pupil and glints centers were obtained. One of these sets was chosen to be the calibration set (C.S.) and used to compute the calibration parameters (equation **(3.2)**). Using the parameters obtained from this calibration, the point-of-gaze on the computer monitor was estimated for the 10 sets of vectors of pupil and glints centers. It was assumed that due to the averaging process that is applied to the calibration data (i.e. 100 estimates of pupil and glints centers for each fixation point) the estimation error in the coordinates of pupil and glints centers is relatively small and can be neglected.

Position	Description
Pos. 1	Eye at ~65 cm from the screen, approximately centered in the field of view of the camera. One of the sets for this position was picked as the calibration set (C.S.).
Pos. 2	Eye at ~65 cm from the screen, approximately vertically centered in the field of view of the camera, to the right hand side of the image. This corresponds to a lateral shift of ~2.5-3 cm from Pos. 1 to the left of the subject.
Pos. 3	Eye at ~65 cm from the screen, approximately vertically centered in the field of view of the camera, to the left hand side of the image. This corresponds to a lateral shift of ~2.5-3 cm from Pos. 1 to the right of the subject.
Pos. 4	Eye at ~60 cm from the screen, approximately centered in the field of view of the camera.
Pos. 5	Eye at ~70 cm from the screen, approximately centered in the field of view of the camera.

Table 4.5: Different head positions adopted for the experiments.

Fig. 4.10 shows the experimental results for subject B.B. (20/20 vision without optical correction). It can be seen that the estimates have a pattern that opens upwards/closes downwards, similar to that of the simulation run for model 1. Note that the dashed circles centered at the intended points of gaze have a radius of 10 mm. Fig. 4.11 corresponds to subject E.G. (medium myopia) without optical correction. The estimates of the point-of-gaze present a clear pattern that opens upwards/closes downwards, similar to that obtained for the simulations with model 2.

Fig. 4.12 corresponds to subject E.G. with eyeglasses. There is a clear pattern that opens upwards, with a somewhat larger dispersion of the estimates of the point-of-gaze than for the same subject without optical correction. One issue about this experiment is that the eyeglasses alter the path of the rays and movement of the eyeglasses relative to the system affects how the ray paths are altered. The geometrical-optical model of the system does not take into account refractive elements between the light sources and the eye and between the eye and the camera, especially if the refractive elements move relative to the system. This issue will be addressed in the future.

Fig. 4.13 shows the results obtained with subject B.L. (medium myopia) with contact lenses. In this case, the pattern of the estimates of the point-of-gaze is more complicated than that observed for the previous cases. On the left, the pattern closes upwards and, on the right, the pattern opens upwards/closes downwards. In addition, the dispersion of the estimates is considerably larger. It should be noted that when contact lenses are used, the glints are produced by the frontal surface of the contact lens rather than the frontal surface of the cornea. Since the contact lenses move relative to the cornea constantly, the dispersion of the point-of-gaze estimates is increased.

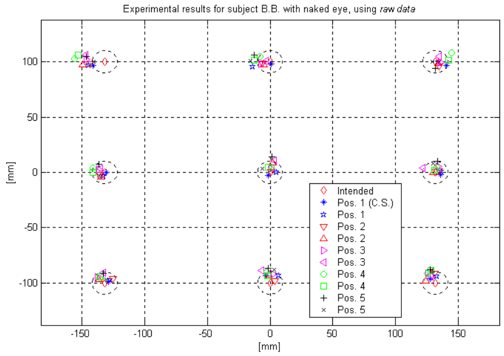


Figure 4.10: Experimental results for subject B.B. (20/20 vision without optical correction).

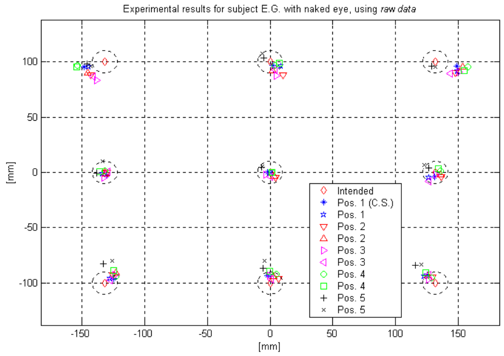


Figure 4.11: Experimental results for subject E.G. (medium myopia) without optical correction.

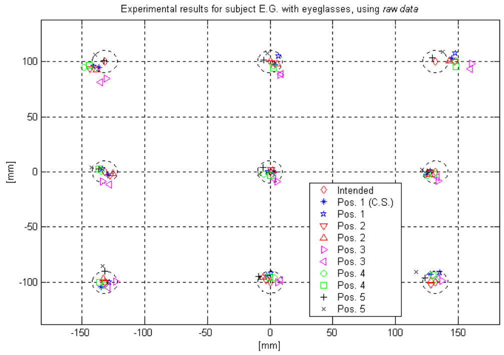


Figure 4.12: Experimental results for subject E.G. (medium myopia) with eyeglasses.

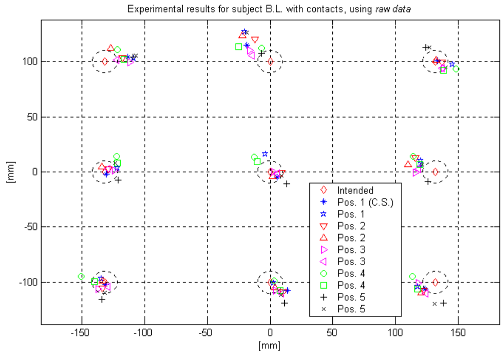


Figure 4.13: Experimental results for subject B.L. (medium myopia) with contact lenses.

The r.m.s. errors of the point-of-gaze estimation for each of these cases are given in table 4.6. The descriptors that were defined in the previous subsection are computed for each experiment. The slope of the line defined by the centers of the two glints is recorded in table 4.7 for each intended point-of-gaze $\mathbf{g}_i$, for the calibration set. The range of variation of those slopes, the mean distance between glints centers and the normalized standard deviation are given in table 4.8. Note that in the case of simulated data the slope of the line defined by the glint centers either increases or decreases monotonically as the gaze moves from left to right (i.e. $\mathbf{g}_1 \to \mathbf{g}_2 \to \mathbf{g}_3$, $\mathbf{g}_4 \to \mathbf{g}_5 \to \mathbf{g}_6$, and $\mathbf{g}_7 \to \mathbf{g}_8 \to \mathbf{g}_9$). However, this does not always happen with the experimental data, as it can be seen in table 4.7, most notably for subject B.L. with contact lenses.

The estimation error patterns, the variation of the slope of the line defined by the glints centers, and the variation of the distance between glints centers observed in the experiments, suggest, when compared to the simulation results, that a non-spherical corneal model is a more adequate representation of a real cornea than a spherical model. In particular, the experimental results of subjects B.B. and E.G. without optical correction are consistent with the simulations for models 1 and 2, respectively. These results suggest that the error pattern of Fig. 3.1 can be explained by corneal asphericity.

Experiment	Data set(s)	R.m.s. error [mm]
Subject B.B. (no optical correction)	Calibration set	6.0147
	All sets	9.3849
Subject E.G. (no optical correction)	Calibration set	9.7290
	All sets	11.8056
Subject E.G. (eyeglasses)	Calibration set	7.1750
	All sets	9.6856
Subject B.L. (contact lenses)	Calibration set	12.9671
	All sets	15.1932

Table 4.6: R.m.s. error of the estimation of the point-of-gaze for the different experiments.

$\mathbf{g}_i$ Experiment	$\mathbf{g}_1$	$\mathbf{g}_2$	$\mathbf{g}_3$	$\mathbf{g}_4$	$\mathbf{g}_5$	$\mathbf{g}_6$	$\mathbf{g}_7$	$\mathbf{g}_8$	$\mathbf{g}_9$
Subject B.B. (no optical correction)	-0.0360	-0.0105	0.0141	-0.0090	-0.0122	0.0041	-0.0116	-0.0132	-0.0143
Subject E.G. (no optical correction)	-0.0430	0.0006	0.0402	-0.0169	-0.0073	0.0066	-0.0037	-0.0104	-0.0051
Subject E.G. (eyeglasses)	-0.0247	-0.0051	0.0200	-0.0086	-0.0078	-0.0124	-0.0121	-0.0124	-0.0087
Subject B.L. (contact lenses)	0.0134	-0.0587	-0.0316	-0.0160	-0.0177	-0.0631	-0.0020	0.0025	-0.0358

Table 4.7: Slopes of the line defined by the glints centers for each intended point-of-gaze for the calibration set for the different experiments.

Descriptor Experiment	Slope range [°]	Mean distance glints [pixels]	Normalized SD [%]
Subject B.B. (no optical correction)	2.8677	21.0756	0.87
Subject E.G. (no optical correction)	4.7640	23.7904	1.76
Subject E.G. (eyeglasses)	2.5587	18.9135	0.60
Subject B.L. (contact lenses)	4.3789	23.5189	3.91

Table 4.8: Range of variation of the slopes shown in table 4.7, mean distance between glints centers, and normalized standard deviation of the distance between glints centers, corresponding to the calibration set, for the different experiments.

4.4 Improvement of the point-of-gaze estimation accuracy

The results from the previous section suggest that the accuracy of the computation of the point-of-gaze could be improved if a non-spherical corneal model were used to estimate the point-of-gaze. However, the number of eye parameters that have to be estimated increases from 4 to 11 for the model developed in Section 4.2, and the number of calibration points on the screen should be increased accordingly. In addition, the time required to estimate the point-of-gaze using the non-spherical model can be more than one order of magnitude larger than that required by the methodology proposed in Chapter 2. For these reasons, sub-optimal solutions that can still use the methodology described in Chapter 2 are proposed. The performance of the sub-optimal solutions is evaluated in subsections 4.4.3 and 4.4.4.

4.4.1 Average slope filter

Inspired by the observations described so far, an approach that reduces to zero the variability of the slope of the line defined by the glints centers is proposed. As explained in the previous chapter, during the calibration procedure the subject is asked to fixate a set of known points $\mathbf{g}_i$, $i = 1, 2, \ldots, N$ on the screen. Typically, for each point $\mathbf{g}_i$, 100 images of the eye are acquired and the positions of the pupil and glints centers are estimated for each image. The results are then averaged to reduce the estimation error, producing the corresponding vectors of estimated coordinates of the pupil and glints centers $\hat{\mathbf{f}}_{raw,i}$, $i = 1, 2, \ldots, N$. The subscript *raw* is used to distinguish the original data from the transformed data that will be obtained later. The vector of coordinates of pupil and glints centers $\hat{\mathbf{f}}_{raw}$ has the form

$$\hat{\mathbf{f}}_{raw} = [r_{left\ glint\ raw} \quad c_{left\ glint\ raw} \quad r_{pupil\ center} \quad c_{pupil\ center} \quad r_{right\ glint\ raw} \quad c_{right\ glint\ raw}]\ . \qquad \textbf{(4.48)}$$

Using the vectors $\hat{\mathbf{f}}_{raw,i}$, i = 1, 2, …, N, the average slope of the lines defined by the glints centers can be computed as

$$avg\_slope = \frac{1}{N}\sum_{i=1}^{N} \frac{r_{right\ glint\ raw,i} - r_{left\ glint\ raw,i}}{c_{right\ glint\ raw,i} - c_{left\ glint\ raw,i}}\ . \qquad \textbf{(4.49)}$$

In this approach, the coordinates of the glints centers are modified in such a way that the slope of the line defined by them has the value of the average slope defined in equation **(4.49)**, while the distance between the glints centers and the position of the middle point between the glints remain unchanged. In this way, the slope variation is reduced to zero. Formally, this transformation of coordinates can be described in general as a function

$$\mathbf{f}_{new\_slope} = \mathcal{S}(\mathbf{f}_{raw}, new\_slope) \qquad \textbf{(4.50)}$$

where the components of $\mathbf{f}_{new\_slope}$ are given by

$$\begin{bmatrix} c_{r.g.new\_slope} \\ r_{r.g.new\_slope} \end{bmatrix} = \frac{1}{2}\left(\begin{bmatrix} c_{r.g.raw} \\ r_{r.g.raw} \end{bmatrix} + \begin{bmatrix} c_{l.g.raw} \\ r_{l.g.raw} \end{bmatrix}\right) + \frac{1}{2}\left\| \begin{bmatrix} c_{r.g.raw} \\ r_{r.g.raw} \end{bmatrix} - \begin{bmatrix} c_{l.g.raw} \\ r_{l.g.raw} \end{bmatrix} \right\| \begin{bmatrix} \cos(\arctan(new\_slope)) \\ \sin(\arctan(new\_slope)) \end{bmatrix} \qquad \textbf{(4.51)}$$

$$\begin{bmatrix} c_{l.g.\mathrm{new}\_slope} \\ r_{l.g.new\_slope} \end{bmatrix} = \frac{1}{2}\left(\begin{bmatrix} c_{r.g.raw} \\ r_{r.g.raw} \end{bmatrix} + \begin{bmatrix} c_{l.g.raw} \\ r_{l.g.raw} \end{bmatrix}\right) - \frac{1}{2}\left\| \begin{bmatrix} c_{r.g.raw} \\ r_{r.g.raw} \end{bmatrix} - \begin{bmatrix} c_{l.g.raw} \\ r_{l.g.raw} \end{bmatrix} \right\| \begin{bmatrix} \cos(\arctan(new\_slope)) \\ \sin(\arctan(new\_slope)) \end{bmatrix} \qquad \textbf{(4.52)}$$

while the coordinates of the pupil center, $r_{pupil\ center}$ and $c_{pupil\ center}$, remain unchanged. The abbreviations *r. g.* and *l. g.* stand for *right glint* and *left glint*, respectively.

In this case, *new_slope* = *avg_slope* as defined by equation **(4.49)** and by applying the above transformation of coordinates, vectors of transformed coordinates of the form

$$\hat{\mathbf{f}}_{avg\_slope} = [r_{l.g.avg\_slope} \quad c_{l.g.avg\_slope} \quad r_{pupil\ center} \quad c_{pupil\ center} \quad r_{r.g.avg\_slope} \quad c_{r.g.avg\_slope}] \qquad \textbf{(4.53)}$$

are obtained. Using this methodology, the original calibration data $\hat{\mathbf{f}}_{raw,i}$, i = 1, 2, …, N, is transformed into new calibration data $\hat{\mathbf{f}}_{avg\_slope,i} = \mathcal{S}(\hat{\mathbf{f}}_{raw,i}, avg\_slope)$, i = 1, 2, …, N. Note that *avg_slope* defined in equation **(4.49)** is regarded as a new model parameter. Using these transformed data, the calibration problem becomes

$$\min_{\boldsymbol{\rho}_{cal,avg\_slope}} \sum_{i=1}^{N} \left\| \hat{\mathbf{g}}(\hat{\mathbf{f}}_{avg\_slope,i}, [\boldsymbol{\rho}_{cal,avg\_slope} \quad \boldsymbol{\rho}_{fixed}]) - \mathbf{g}_i \right\|^2 , \tag{4.54}$$

where $\boldsymbol{\rho}_{fixed}$ is the vector of fixed system parameters and $\boldsymbol{\rho}_{cal,avg\_slope}$ is the vector of parameters that are optimized by the calibration for the transformed data.

In order to estimate the point-of-gaze, the vector of estimated coordinates of pupil and glints centers $\hat{\mathbf{f}}_{raw}$ for each eye image is transformed, as explained above, into a new vector $\hat{\mathbf{f}}_{avg\_slope} = \mathcal{S}(\hat{\mathbf{f}}_{raw}, avg\_slope)$. Using this vector of transformed coordinates of pupil and glints centers $\hat{\mathbf{f}}_{avg\_slope}$, the point-of-gaze is estimated as explained in Chapter 2 (equation **(2.38)**) using the optimum calibration parameters $\boldsymbol{\rho}_{cal,avg\_slope}$ as

$$\hat{\mathbf{g}} = \hat{\mathbf{g}}(\hat{\mathbf{f}}_{avg\_slope}, [\boldsymbol{\rho}_{cal,avg\_slope} \quad \boldsymbol{\rho}_{fixed}]) = \hat{\mathbf{g}}(\mathcal{S}(\hat{\mathbf{f}}_{raw}, avg\_slope), [\boldsymbol{\rho}_{cal,avg\_slope} \quad \boldsymbol{\rho}_{fixed}]) \ . \tag{4.55}$$

By using this methodology, the accuracy of the calibration improves, as it is illustrated later for simulated and experimental data. However, if the center of rotation of the eye (**d**) translates relative to the position assumed during calibration, small errors in the estimation of the point-of-gaze occur. This is especially noticeable in the case of lateral translation. The source of this error is the fact that the slope of the line defined by the glints produced by the spherical cornea assumed to estimate the point-of-gaze does vary over a small interval for different eye positions and different points of gaze on the computer screen, rather than keeping a constant value. A refinement to deal with this problem is described in the following subsection.

4.4.2 Two-stage slope filter

The methodology of the previous subsection can be refined with the addition of a second calibration stage, as well as a second estimation stage. The solution of the calibration problem represented by equation **(4.54)** is a vector of parameters $\boldsymbol{\rho}_{cal,avg\_slope}$. By using this vector of parameters, estimates of the point-of-gaze $\hat{\mathbf{g}}_{avg\_slope}$ and the position of the center of rotation $\hat{\mathbf{d}}_{avg\_slope}$ can be computed with equation **(2.41)** for each vector of transformed coordinates of pupil and glints centers $\hat{\mathbf{f}}_{avg\_slope}$ as

$$[\hat{\mathbf{g}}_{avg\_slope} \quad \hat{\mathbf{d}}_{avg\_slope}] = \mathcal{G}(\hat{\mathbf{f}}_{avg\_slope}, [\boldsymbol{\rho}_{cal,avg\_slope} \quad \boldsymbol{\rho}_{fixed}]) \tag{4.56}$$

Then, the coordinates that the glints centers would have for an ideal eye with spherical cornea, having its center of rotation at $\hat{\mathbf{d}}_{avg\_slope}$ and parameters $\boldsymbol{\rho}_{cal,avg\_slope}$, can be calculated with equation **(2.67)** as

$$\tilde{\mathbf{f}} = \mathbf{f}(\hat{\mathbf{g}}_{avg\_slope}, \hat{\mathbf{d}}_{avg\_slope}, [\boldsymbol{\rho}_{cal,avg\_slope} \quad \boldsymbol{\rho}_{fixed}]) \tag{4.57}$$

in general or as

$$\tilde{\mathbf{f}} = \mathbf{f}(\mathbf{g}, \hat{\mathbf{d}}_{avg\_slope}, [\boldsymbol{\rho}_{cal,avg\_slope} \quad \boldsymbol{\rho}_{fixed}]) \tag{4.58}$$

for the calibration problem given that in that case the actual point-of-gaze $\mathbf{g}$ is known. The vector $\tilde{\mathbf{f}}$ has the form

$$\tilde{\mathbf{f}} = [\tilde{r}_{left\ glint} \quad \tilde{c}_{left\ glint} \quad \tilde{r}_{pupil\ center} \quad \tilde{c}_{pupil\ center} \quad \tilde{r}_{right\ glint} \quad \tilde{c}_{right\ glint}] \tag{4.59}$$

and is used to calculate the corresponding slope

$$\widetilde{slope} = \frac{\tilde{r}_{right\ glint} - \tilde{r}_{left\ glint}}{\tilde{c}_{right\ glint} - \tilde{c}_{left\ glint}} \ . \tag{4.60}$$

This slope value is used to refine the coordinates of the glints centers, obtaining a new vector of coordinates of pupil and glints centers as

$$\hat{\mathbf{f}}_{2nd\_stage} = \mathcal{S}(\hat{\mathbf{f}}_{raw}, \widetilde{slope}) \tag{4.61}$$

where the function $\mathcal{S}(.)$ was defined in equation **(4.50)**.

Through these steps the original calibration data $\hat{\mathbf{f}}_{raw,i}$, i = 1, 2, …, N are transformed into new calibration data $\hat{\mathbf{f}}_{2nd\_stage,i}$ with slope $\widetilde{slope}_i$. The second calibration stage is then formally expressed as

$$\min_{\boldsymbol{\rho}_{cal,2nd\_stage}} \sum_{i=1}^{N} \left\| \hat{\mathbf{g}}(\hat{\mathbf{f}}_{2nd\_stage,i}, [\boldsymbol{\rho}_{cal,2nd\_stage} \quad \boldsymbol{\rho}_{fixed}]) - \mathbf{g}_i \right\|^2 \ . \tag{4.62}$$

The second estimation stage then becomes

$$\hat{\mathbf{g}} = \hat{\mathbf{g}}(\hat{\mathbf{f}}_{2nd\_stage}, [\boldsymbol{\rho}_{cal,2nd\_stage} \quad \boldsymbol{\rho}_{fixed}]) \ . \tag{4.63}$$

By using this second stage, the error in the estimation of the point-of-gaze due to translation of the center of rotation of the eye is reduced (illustrated later). The two sub-optimal approaches described in the previous and present subsections support the simplified model assumption of a spherical corneal surface while improving the estimation accuracy.

4.4.3 Simulations

The purpose of this subsection is to evaluate, through simulations, the performance of the sub-optimal filters proposed above. The simulations are carried out for the data used in subsection 4.3.1.

Fig. 4.14(a) (same as Fig. 4.6) shows the results corresponding to model 1 with $\xi = 1$ using the raw data. It can be observed that the estimates of the point-of-gaze have a pattern that opens upwards/closes downwards. This pattern is not very prominent but is clearly visible. When the average slope filter is used, that pattern disappears, as shown in Fig. 4.14(b). However, for each intended point-of-gaze the estimates form three clearly distinguished clusters. These clusters are due to lateral displacement of the center of rotation of the eye and are an artifact of the average slope filter. The two-stage slope filter overcomes that problem as illustrated in Fig. 4.14(c). The corresponding values of the r.m.s. error of the estimation of the point-of-gaze for the calibration set and for all the positions of the center of rotation of the eye are given in table 4.9. The improvement given by the two-stage slope filter is apparent. Note that the dashed circles centered at the intended points of gaze have a radius of 10 mm.

Fig. 4.15(a) (same as Fig. 4.7) shows the results corresponding to model 2 with $\xi = 1$ using the raw data. It can be observed that the estimates of the point-of-gaze also have a pattern that opens upwards/closes downwards. This pattern is much more prominent than that seen for model 1 and there is a larger dispersion of the estimates of the point-of-gaze for different positions of the center of rotation of the eye. The average slope filter dramatically improves the accuracy of the estimates, as shown in Fig. 4.15(b), and the two-stage slope filter additionally reduces the dispersion of the estimates, as illustrated in Fig. 4.15(c). The benefit of using the two-stage slope filter is clear, as seen in this figures and in table 4.9.

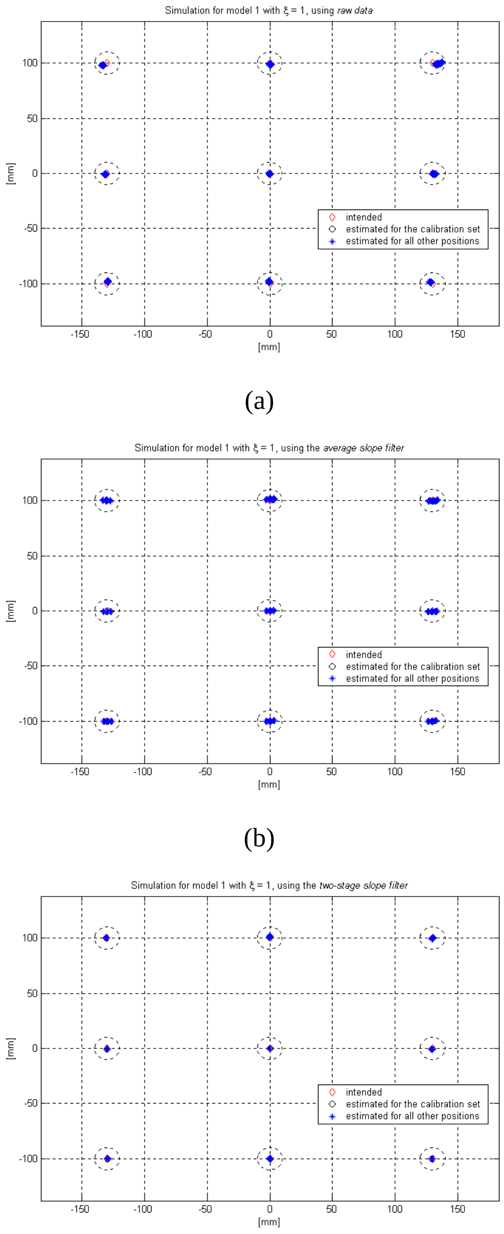
Fig. 4.16(a) (same as Fig. 4.8) shows the results corresponding to model 2 with $\xi = 1.02$ and $\delta = 45^\circ$ using the raw data. The purpose of this example is to consider a case with corneal astigmatism. The pattern of the estimates is somewhat similar to that shown in Fig. 4.15(a). The average slope filter and the two-stage slope filter clearly improve the accuracy of the estimates, as shown in Figs. 4.16(b) and 4.16(c), respectively, and in table 4.9. However, as it can be seen in table 4.9, the accuracy of the estimates in this case (elliptical section) is not as good as for model 2 with $\xi = 1$ (circular section).

Fig 4.17(a) (same as Fig. 4.9) shows the results corresponding to model 3 with $\xi = 1$ using the raw data. In this case, the estimates have a pattern that opens downwards/closes upwards, inverted relative to the previous examples. This pattern is pretty prominent, in consistence with the corresponding curvature function plotted in Fig. 4.5. The average slope filter clearly removes that pattern and improves the accuracy of the estimates, as shown in Fig. 4.17(b), while the two-stage slope filter additionally reduces the dispersion of the estimates, as seen in Fig. 4.17(c). The values of the r.m.s. error given in table 4.9 support the above observations.

Based on the examples studied through simulations, it can be seen that the methodology proposed in this chapter satisfactorily overcomes the problem of asphericity of the corneal surface.

Model	Data set(s)	Raw data	Average slope filter	Two-stage slope filter
Model 1 $\xi = 1$	Calibration set	2.6320	0.5874	0.5977
	All positions	2.8228	2.5305	0.6949
Model 2 $\xi = 1$	Calibration set	9.3210	1.9267	2.0461
	All positions	10.1439	3.3104	2.2887
Model 2 $\xi = 1.02$, $\delta = 45^{\circ}$	Calibration set	9.2449	4.0168	4.1425
	All positions	10.0194	4.9143	4.3495
Model 3 $\xi = 1$	Calibration set	10.4445	2.4610	2.5675
	All positions	10.9931	3.9367	2.9658

Table 4.9: R.m.s. error of the estimation of the point-of-gaze for simulations for different non-spherical corneal models, using the proposed filters vs. raw data. All errors are in mm.

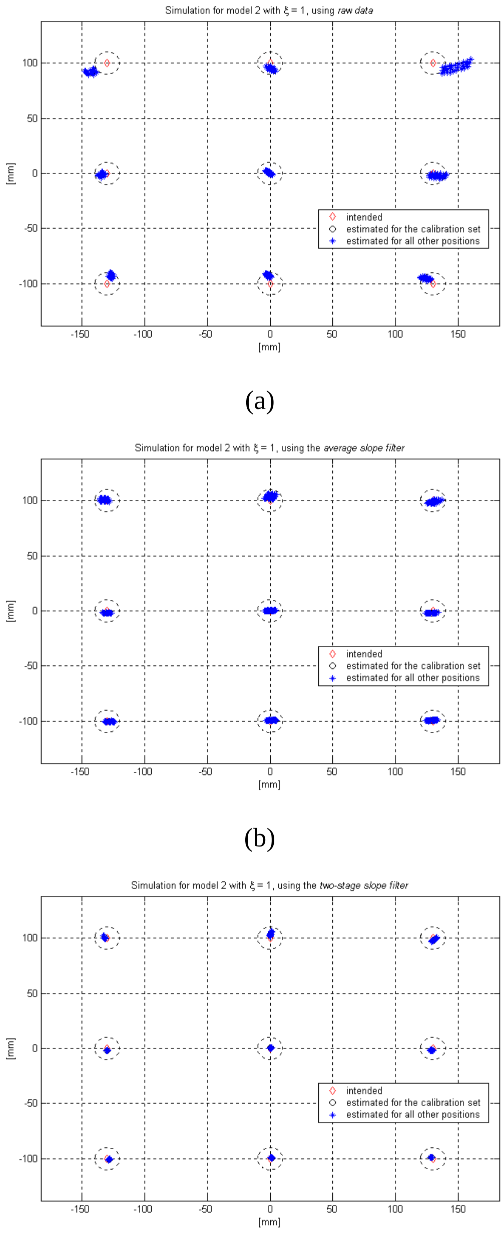


(a)

(b)

(c)

Figure 4.14: Simulation results for model 1 with $\xi = 1$, using (a) the raw data, (b) the average slope filter and (c) the two-stage slope filter.

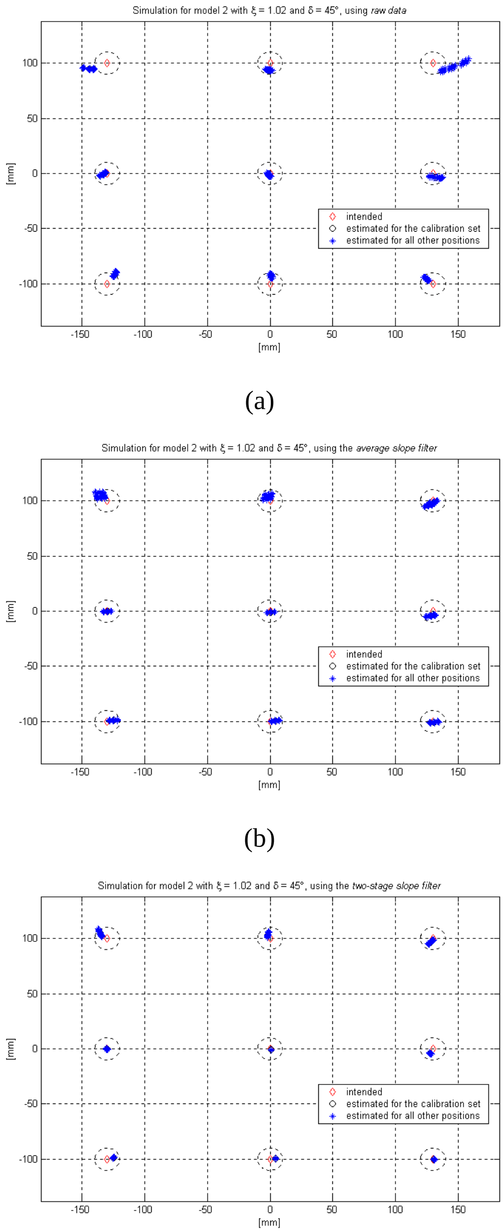


(a)

(b)

(c)

Figure 4.15: Simulation results for model 2 with $\xi = 1$, using (a) the raw data, (b) the average slope filter and (c) the two-stage slope filter.

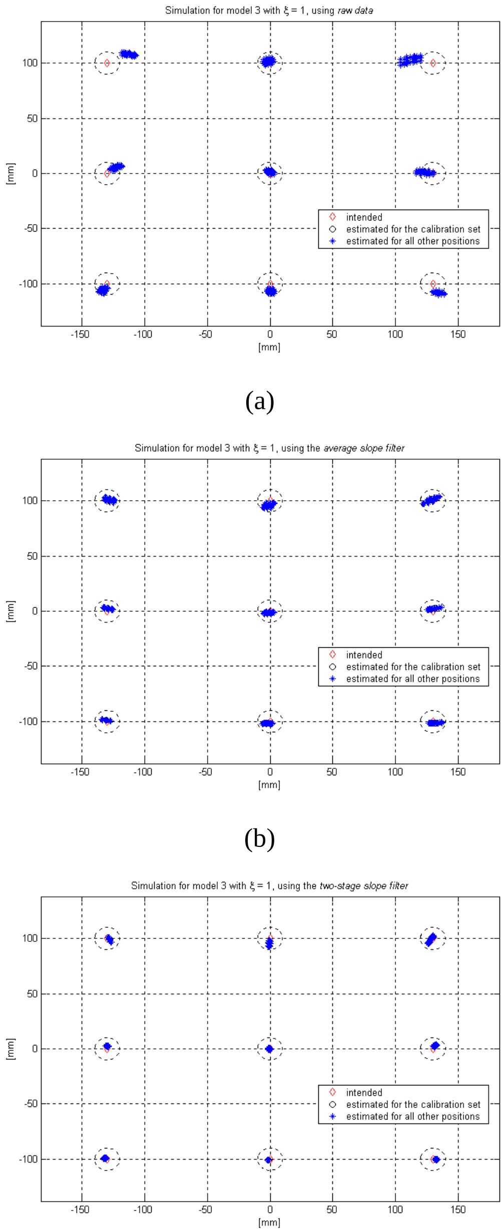


(a)

(b)

(c)

Figure 4.16: Simulation results for model 2 with $\xi = 1.02$ and $\delta = 45^{\circ}$, using (a) the raw data, (b) the average slope filter and (c) the two-stage slope filter.



(a)

(b)

(c)

Figure 4.17: Simulation results for model 3 with $\xi = 1$, using (a) the raw data, (b) the average slope filter and (c) the two-stage slope filter.

4.4.4 Experimental results

This subsection demonstrates the performance of the proposed filters with the experimental data described in subsection 4.3.2. Fig. 4.18(a) (same as Fig. 4.10) shows the experimental results for subject B.B. (20/20 vision without optical correction) using the raw data. It can be seen that the estimates have a pattern that opens upwards/closes downwards. Fig. 4.18(b) shows the results using the average slope filter, where that pattern is no longer noticeable. Fig. 4.18(c) shows the results using the two-stage slope filter and it can be clearly seen that the clustering of the estimates improved. Table 4.10 provides the corresponding r.m.s. errors of the estimation of the point-of-gaze, both for the calibration data set and for all 10 data sets. The two-stage slope filter provides an improvement in the estimation accuracy. Note that the dashed circles centered at the intended points of gaze have a radius of 10 mm.

Fig. 4.19(a) (same as Fig. 4.11) corresponds to subject E.G. (medium myopia) without optical correction, using the raw data. The estimates of the point-of-gaze present a clear pattern that opens upwards/closes downwards. The average slope filter clearly improves this situation, as illustrated in Fig. 4.19(b), while the two-stage slope filter additionally improves the clustering of the estimates, as shown in Figure 4.19(c). The corresponding values of the r.m.s. errors given in table 4.10 support these observations.

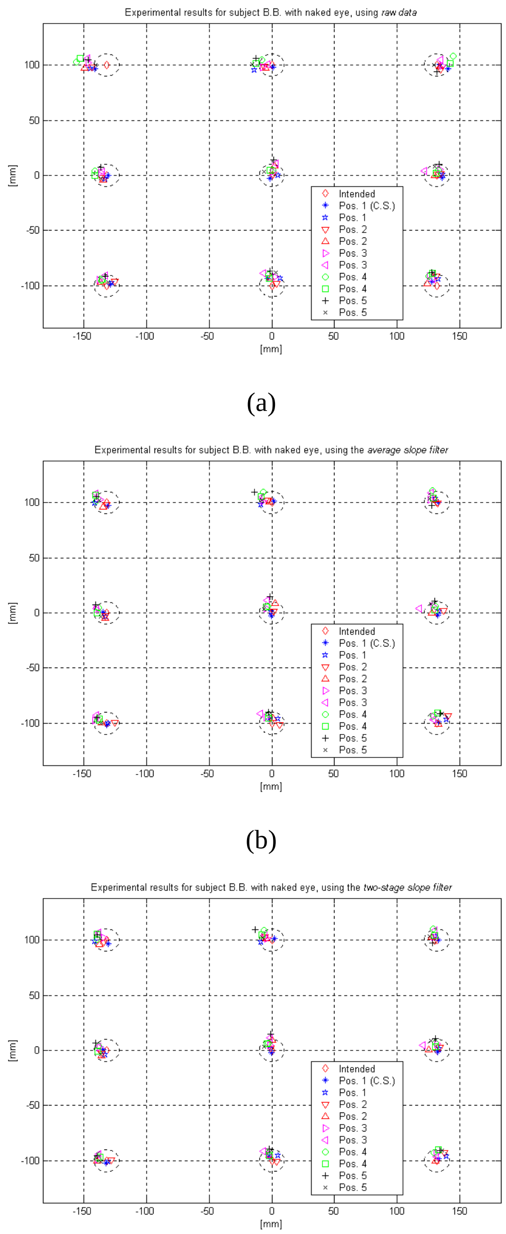
Fig. 4.20(a) (same as Fig. 4.12) corresponds to subject E.G. with eyeglasses, using the raw data. There is a clear pattern that opens upwards, with a somewhat larger dispersion of the estimates of the point-of-gaze than for the same subject without optical correction. As shown in Fig. 4.20(b), the average slope filter clearly improves the situation, however, the two-stage slope filter does not improve the results over the average slope filter for all the data sets, as it can be seen in Fig. 4.20(c) and in table 4.10. Nevertheless, its performance is still reasonable in comparison with the other experiments.

Fig. 4.21(a) (same as Fig. 4.13) shows the results obtained with subject B.L. (medium myopia) with contact lenses, using the raw data. In this case, the pattern of the point-of-gaze estimates is more complicated than in the other cases. On the left, the pattern closes upwards and, on the right, the pattern opens upwards/closes downwards, with a considerably larger dispersion of the estimates. However, despite that complicated pattern and the large dispersion, the average slope filter provides a dramatic improvement, as illustrated in Fig. 4.21(b). The two-stage slope filter gives a small additional improvement in the clustering of the estimates, as seen in Fig. 4.21(c), and in the estimation accuracy, as detailed in table 4.10.

The results obtained with the experimental data show that the filters proposed in this section clearly improve the point-of-gaze estimation accuracy. The r.m.s. estimation error using the two-stage slope filter is between 46 to 77 % of the r.m.s. estimation error obtained when the raw data is used. In particular, for the case that exhibits the largest r.m.s. estimation error with the raw data (subject B.L. wearing contact lenses), the use of the two-stage slope filter reduced the r.m.s. estimation error by a factor of about 2.18, from 15.19 mm to 6.96 mm. For all the experiments carried out, both the average slope filter and the two-stage slope filter keep the r.m.s. estimation error below 8 mm (0.7° of visual angle for an average distance of 65 cm between the eye and the computer screen).

Experiment	Data set(s)	Raw data	Average slope filter	Two-stage slope filter
Subject B.B. (no optical correction)	Calibration set	6.0147	2.5814	2.7697
	All sets	9.3849	7.6185	7.2346
Subject E.G. (no optical correction)	Calibration set	9.7290	2.3644	2.2456
	All sets	11.8056	7.2527	6.7680
Subject E.G. (eyeglasses)	Calibration set	7.1750	3.0693	2.8483
	All sets	9.6856	6.3712	6.8822
Subject B.L. (contact lenses)	Calibration set	12.9671	3.5908	3.4456
	All sets	15.1932	7.2328	6.9609

Table 4.10: R.m.s. error of the estimation of the point-of-gaze for experiments with different subjects, using the proposed filters vs. raw data. All errors are in mm.

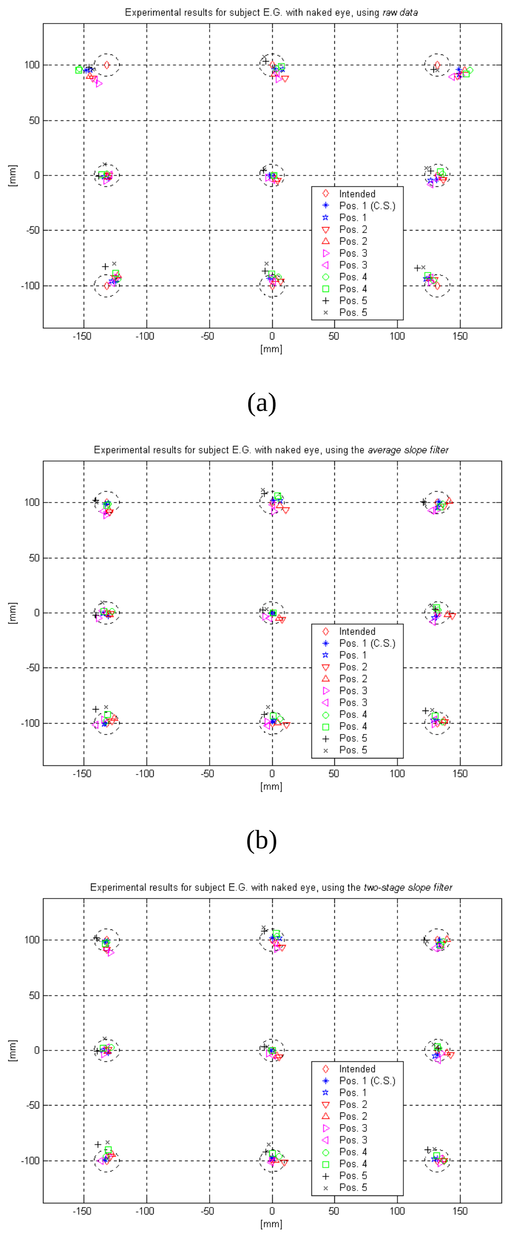


(a)

(b)

(c)

Figure 4.18: Experimental results for subject B.B. (20/20 vision without optical correction), using (a) the raw data, (b) the average slope filter and (c) the two-stage slope filter.

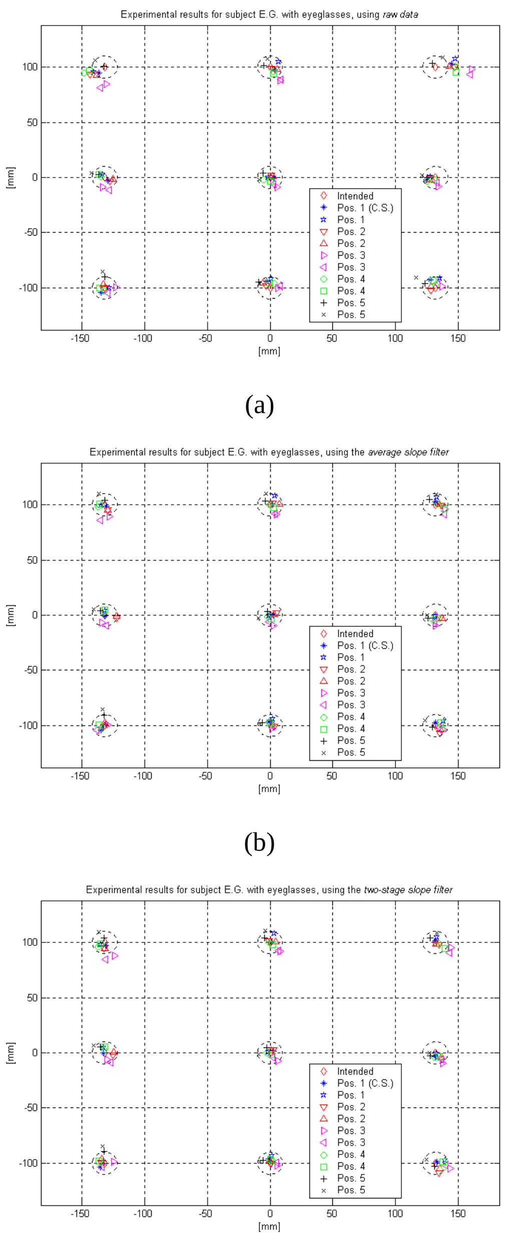


(a)

(b)

(c)

Figure 4.19: Experimental results for subject E.G. (medium myopia) without optical correction, using (a) the raw data, (b) the average slope filter and (c) the two-stage slope filter.

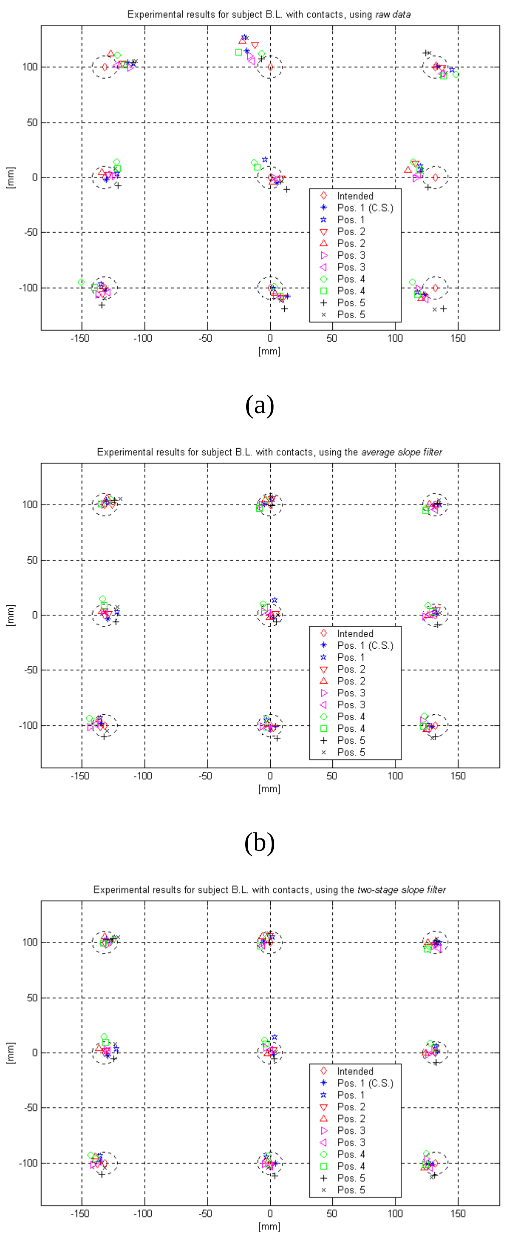


(a)

(b)

(c)

Figure 4.20: Experimental results for subject E.G. (medium myopia) with eyeglasses, using (a) the raw data, (b) the average slope filter and (c) the two-stage slope filter.



(a)

(b)

(c)

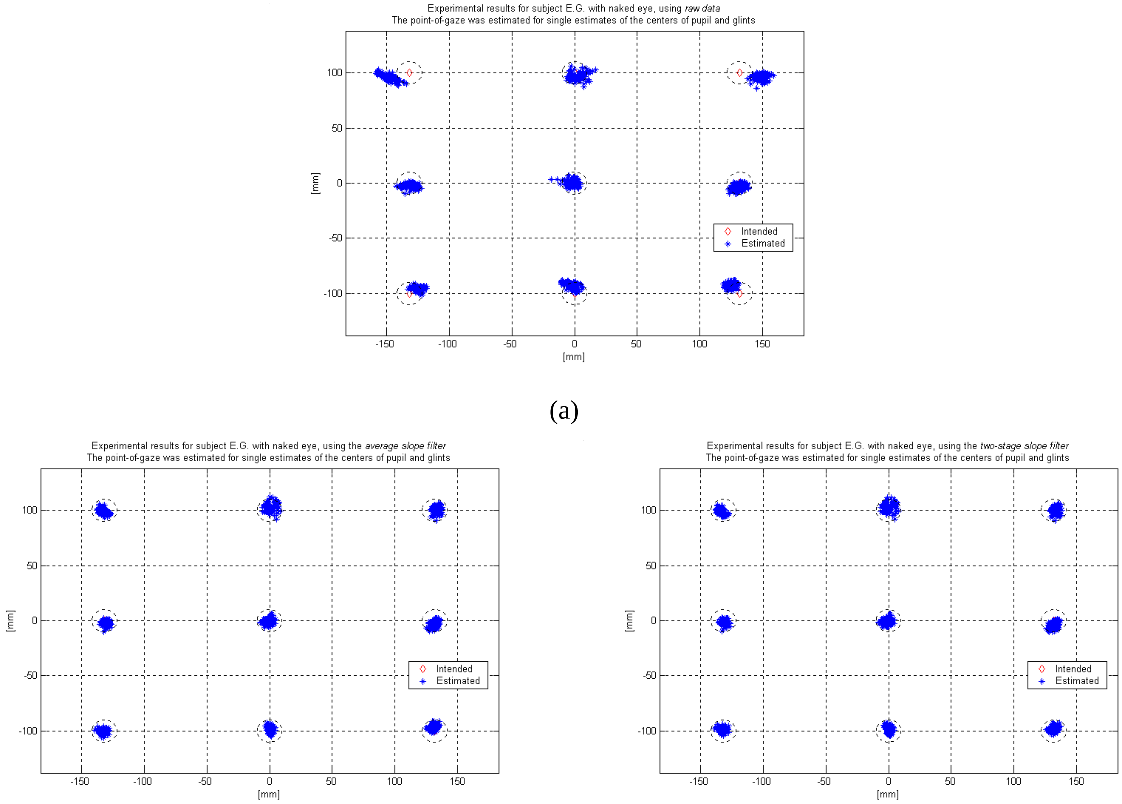
Figure 4.21: Experimental results for subject B.L. (medium myopia) with contact lenses, using (a) the raw data, (b) the average slope filter and (c) the two-stage slope filter.

4.5 Effects of errors in the estimation of the coordinates of pupil and glints centers in the eye images

For the experiments described in this chapter, the coordinates of pupil and glints centers were obtained as the average of 100 estimates, and hence the effect of the estimation error was reduced. However, the estimation of the point-of-gaze in most applications is done for individual estimates of the coordinates of pupil and glints centers.

Due to estimation error caused by image noise, the slope of the line defined by the estimates of the glints centers varies from image to image, even when the eye and gaze remain stationary. In order to illustrate the effect of this phenomenon, Fig. 4.22(a) shows results using raw data corresponding to subject E.G. without optical correction, when the point-of-gaze was computed for each of the 100 estimates of the coordinates of pupil and glints centers that were obtained for each intended point-of-gaze, and from which the calibration data set was obtained by averaging. Figs. 4.22(b) and 4.22(c) illustrate the results obtained by using the average slope filter and the two-stage slope filter, respectively. Clearly, the effect of the estimation error of the coordinates of pupil and glints centers is reduced by the application of the methodology developed in Section 4.4. The accuracy of the estimation of the point-of-gaze improves and the dispersion of the estimates of the point-of-gaze is notably reduced.

The r.m.s. errors in the estimation of the point-of-gaze for all the data collected during the experiments described in subsections 4.3.2 and 4.4.4 were calculated. Table 4.11 presents the r.m.s. errors for each experiment for the data from which the calibration set was obtained and for all the data. Note that a point-of-gaze estimation error of 9 mm is equivalent to 0.8° of visual angle when the eye is at 65 cm from the computer screen. The data in table 4.11 suggest that point-of-gaze estimates with either the average slope filter or the two-stage slope filter satisfy, in r.m.s. sense, the accuracy requirements of this project (1° of visual angle).



(a)

(b) (c)

Figure 4.22: Experimental results for subject E.G. without optical correction, using (a) the raw data, (b) the average slope filter and (c) the two-stage slope filter. The point-of-gaze was estimated for individual estimates of pupil and glints centers for the calibration data set.

Experiment	Data set(s)	Raw data	Average slope filter	Two-stage slope filter
Subject B.B. (no optical correction)	Calibration set	8.1996	5.0101	5.1022
	All sets	11.0958	8.8591	8.5246
Subject E.G. (no optical correction)	Calibration set	10.8812	4.3199	4.2537
	All sets	12.8506	8.2941	7.8704
Subject E.G. (eyeglasses)	Calibration set	9.8203	6.0314	5.8843
	All sets	11.8664	8.1814	8.5703
Subject B.L. (contact lenses)	Calibration set	14.0260	4.9648	4.8500
	All sets	16.2160	8.0938	7.8425

Table 4.11: R.m.s. error in the estimation of the point-of-gaze over individual estimates of the coordinates of pupil and glints centers for experiments with different subjects. All errors are in mm.

5. Pilot study and conclusions

5.1 Introduction

Age-related macular degeneration (ARMD) is the leading cause of legal blindness in North America, affecting about 1.6 million persons over 50 years of age [41]. In macular degeneration, the macula (central 5-7° of the retina), which is used for fine visual tasks such as reading and face recognition, is damaged, resulting in a dramatic reduction in the quality of life. There are different variants of the disease but in most cases it ranges from impaired to complete central vision loss (central scotoma[5]).

A person with a central scotoma unconsciously chooses an eccentric retinal area, a Preferred Retinal Locus (PRL), to perform the visual tasks that the fovea used to perform, such as fixation, reading or tracking. Such persons perform these tasks by moving the eyes so that the image of the objects of interest falls within their PRL [49, 50]. Because it is in the periphery, the visual acuity of a PRL is lower than that of the fovea which means that magnification is required to compensate for the reduced resolution.

It has been shown that with appropriate visual aids, a great proportion of patients with ARMD can be successfully trained to read and perform daily tasks using eccentric viewing. The most common visual aids for near vision are strong positive lenses that by virtue of their short focal length demand that the text is held very close to the eye (typically, 4-5 cm and as close as 1.7 cm in extreme cases). This obviously precludes binocular vision, and is associated with a very small field of view. Nevertheless, the improvement in the ability to read is significant in comparison with no visual aids [51].

---

[5] Scotomas are retinal areas with reduced or null light sensitivity. A dense scotoma (also called absolute scotoma) is insensitive to very bright light, while a relative scotoma has a reduced light sensitivity compared to a healthy retina.

Other type of visual aid comprises closed-circuit TV (CCTV) magnifiers [51, 52] and computer screen magnifiers such as that included in the Microsoft Windows operating system (accessibility tools). A CCTV magnifier [52] consists of a monitor, usually quite large, and a video camera equipped with a zoom lens. The camera and lens look down on a movable platform. Printed material is placed on the platform and the reader views the magnified image on the monitor screen. The reader moves through the text by moving the platform, thereby moving the material through the camera's field of view. Computer screen magnifiers consist of a fixed window where the information from a region around the mouse pointer is displayed with the selected magnification. A key disadvantage of this type of approach is that the amount of information that can be displayed by the magnifier is limited (a few characters) while there is a lack of information regarding the location of the text that is being read. This leads to text navigation problems and it is fairly common for patients that use this type of aid to get lost in the text [52]. These approaches neglect the natural visual navigation skills for reading.

In collaboration with the Vision Science Research Program (VSRP) at the Toronto Western Hospital (Drs. Esther González, Sam Markowitz and Emad Eskander) we are developing a novel visual aid to allow subjects with ARMD to read from a computer monitor. The essence of the concept is to provide a magnifier that can be moved by the subject across the screen. A pair of crosshairs spanning the whole screen is used to provide information regarding the region of the screen that is being explored and to stabilize the gaze. The magnifier itself is a window located with the appropriate eccentricity relative to the center of the crosshairs, where the visual information from a region around the center of the crosshairs is displayed with the required magnification. One of the main objectives of our approach is to preserve natural visual navigation skills for reading.

We are considering two different approaches regarding the control of the magnifier position. One approach consists of moving the magnifier with a pointing device such as a mouse or a trackball, which would make the system inexpensive. The other approach involves moving the magnifier with the gaze ("gaze-slaved") so that the information of interest, which would be seen with the fovea if the eyes were healthy, is now always projected onto the PRL with adequate magnification.

The next section shows some preliminary experimental results of eccentric reading tests with subjects with normal or corrected-to-normal vision and a simulated central scotoma. Section 5.3 summarizes the contributions of this thesis and describes planned future work.

5.2 Preliminary experiments

Three subjects participated in these preliminary experiments. Each subject sat in front of a 19 in. computer monitor (18 in. visible) at a distance of 65 cm using a chinrest to ensure that the right eye was always in the field of view of the camera of the point-of-gaze estimation system. The point-of-gaze estimation system was used to monitor and record the subject's point-of-gaze.

Each subject had to read four previously unread text fragments with the periphery of the retina. The first fragment was used to make the subject familiar with the experiment and performance was not recorded. Each subject read all four text fragments within a single session with brief breaks between fragments. All subjects read the same fragments in the same order. The reading material consisted of passages between 100 and 200 words long excerpted from different books intended for children between 10 and 12 years of age. The normal-sized text displayed on the screen was blurred to prevent the subject from reading it directly.

The central scotoma was simulated as a black disk of 7º of visual angle in diameter. The subject had to keep his gaze at the center of the black disk and read with the periphery of the retina from the eccentrically located magnifier window. The magnifier window was placed below the black disk since the lower visual hemifield is preferred for eccentric viewing by patients with ARMD whenever possible (i.e., if there is no significant lesion in the lower visual hemifield). The lower visual hemifield is preferred over the upper visual hemifield because it has better acuity and is used for most daily tasks such as locomotion. The eccentricity of the center of the magnifier window was about 7º below the center of the black disk. The height of the window subtended 7.3º and the width subtended 19.3º, accommodating 4 to 6 characters. The magnified characters were white on a black background (negative image) as this seemed to favor reading. The font used for these experiments was Helvetica bold, a sans serif of high readability. The magnified characters, shown in the magnifier window, had a height of about 3.5° of visual angle for lowercase characters such as “e”, “o”, etc., and about 4.4° for lowercase characters such as “b”, “p”, etc., and uppercase characters. An example of the visual information presented on the screen is shown in Fig. 5.1. In order to make navigation easier, the vertical position of the center of the black disk was restricted to the lines of text only.

The size of the magnified characters was determined using the results described in [53] which showed that an alphanumeric target presented with an eccentricity of ε degrees has to be magnified by a factor of $(1+1.844\ \varepsilon)$ relative to a target presented in the fovea in order to attain equal identification ability. If the eccentricity of the center of the magnifier window is about 7°, the magnification factor should be around 13.9. Now, considering that for a normal sighted subject there is no difficulty in reading characters that have a height of 0.25° (15 min of arc), the magnified characters should then have a height of about 3.5° of visual angle. Making the magnified characters much larger is not recommended because the number of characters that can be perceived per fixation reduces in such a way that reading is impaired.

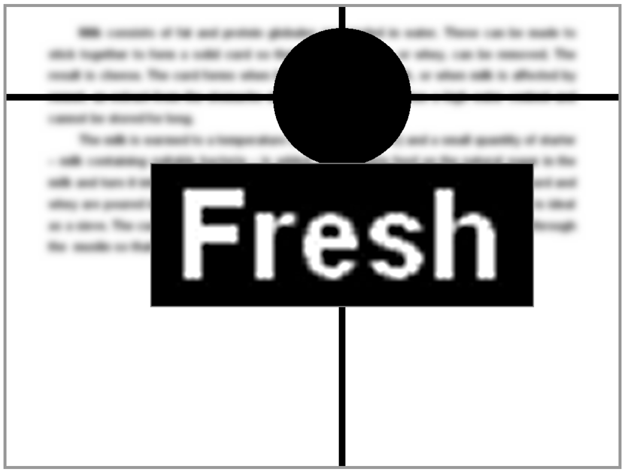


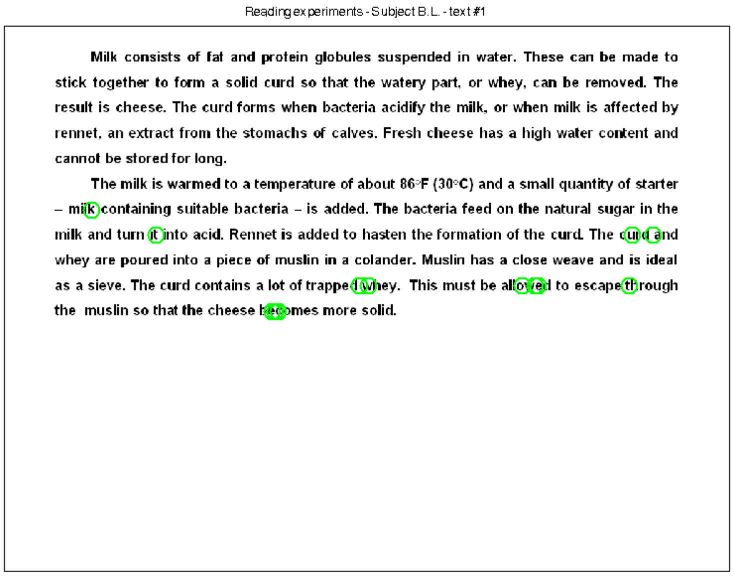
Figure 5.1: Example of the information presented on the screen (blurred text, crosshair, black disk that simulates the central scotoma and magnifier window).

The subject was instructed to keep his gaze at the center of the black disk for the duration of the experiment and read aloud with the periphery of the retina from the eccentrically located magnifier window. In the case where the magnifier was moved using a mouse, the point-of-gaze estimation system was used to monitor and record the point-of-gaze in real time along with the position of the center of the black disk. If the estimate of the point-of-gaze was below the center of the black disk by more than a certain threshold (about 1.5°), it was considered as an attempt to read with the fovea and the magnifier window was blanked. Immediately after the experiment, using the same criteria on the recorded data, an image of the normal-sized text (without blurring) could be displayed with markers superimposed on the words where an attempt to read with the fovea was suspected (Figs. 5.2 and 5.3). The reading speed was computed for each text fragment as the number of words correctly read with the periphery of the retina divided by the time required to read the fragment of text. Table 5.1 summarizes the results.

The results of these reading tests showed that the reading performance improved with practice for two of the three subjects tested. As mentioned in [51], training is fundamental for patients with ARMD to be able to use eccentric viewing effectively. Another observation from this experiment is that although the vertical position of the center of the black disk was restricted to the lines of text only, one of the subjects showed some text navigation problems. He frequently skipped a line or went back to an already read line and, even though he realized that and eventually found and read the correct line, this impacted on the reading performance. This observation suggested that it would be useful to be able to control the position of the magnifier using the arrow keys of the computer keyboard in addition to the pointing device.

Subject	Text #	Total number of words	Number of words read incorrectly or missed	Number of words suspected to have been read with the fovea	Number of words read correctly with the periphery of the retina	Effective reading speed [words per minute]
B.B.	1	168	0	37	131	18.11
	2	184	0	4	180	30.17
	3	180	3	2	175	33.76
B.L.	1	168	0	8	160	16.19
	2	184	0	25	159	14.72
	3	180	0	15	165	21.20
M.E.	1	168	1	34	133	18.47
	2	184	0	65	119	15.56
	3	180	1	70	109	13.00

Table 5.1: Results of the reading tests.

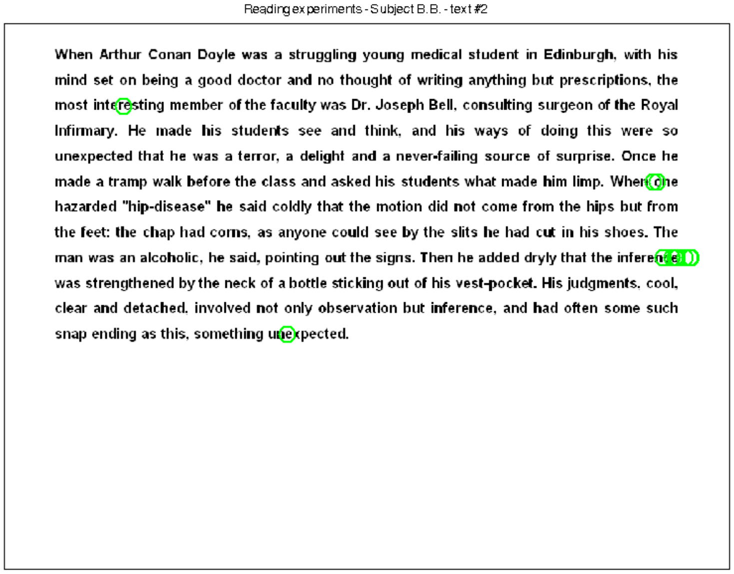


Reading experiments - Subject B.L. - text #1

Milk consists of fat and protein globules suspended in water. These can be made to stick together to form a solid curd so that the watery part, or whey, can be removed. The result is cheese. The curd forms when bacteria acidify the milk, or when milk is affected by rennet, an extract from the stomachs of calves. Fresh cheese has a high water content and cannot be stored for long.

The milk is warmed to a temperature of about 86°F (30°C) and a small quantity of starter – milk containing suitable bacteria – is added. The bacteria feed on the natural sugar in the milk and turn it into acid. Rennet is added to hasten the formation of the curd. The curd and whey are poured into a piece of muslin in a colander. Muslin has a close weave and is ideal as a sieve. The curd contains a lot of trapped whey. This must be allowed to escape through the muslin so that the cheese becomes more solid.

Figure 5.2: Passage with markers superimposed on the words where attempts to read with the fovea were suspected (Text no. 1, subject B.L.).



Reading experiments - Subject B.B. - text #2

When Arthur Conan Doyle was a struggling young medical student in Edinburgh, with his mind set on being a good doctor and no thought of writing anything but prescriptions, the most interesting member of the faculty was Dr. Joseph Bell, consulting surgeon of the Royal Infirmary. He made his students see and think, and his ways of doing this were so unexpected that he was a terror, a delight and a never-failing source of surprise. Once he made a tramp walk before the class and asked his students what made him limp. When one hazarded "hip-disease" he said coldly that the motion did not come from the hips but from the feet: the chap had corns, as anyone could see by the slits he had cut in his shoes. The man was an alcoholic, he said, pointing out the signs. Then he added dryly that the inference was strengthened by the neck of a bottle sticking out of his vest-pocket. His judgments, cool, clear and detached, involved not only observation but inference, and had often some such snap ending as this, something unexpected.

Figure 5.3: Passage with markers superimposed on the words where attempts to read with the fovea were suspected (Text no. 2, subject B.B.).

The gaze-slaved magnifier, which uses the point-of-gaze estimation system to control the position of the magnifier, was also tried but it was found that reading was more difficult.

The main observation was that normal visual scanning patterns for reading were ineffective in this case. Normal reading consists of a succession of fixation-saccade[6] patterns. Each fixation is used to perceive a string of 4 or 5 characters and is followed by a saccade to move the gaze to the next 4 or 5 characters [54]. In our experiment, the normal-sized text is not only blurred but it is also occluded by the black disk that simulates the central scotoma. It seems that since the text is occluded, there is no clear reference for the saccadic eye movements performed in normal reading. In addition, the magnifier window showed only 4 to 6 characters (it was not possible to show more characters because of the size of the screen), which means that if due to lack of visual reference the amplitude of the saccades is larger than the length of the character string showed in the magnifier window, some characters are skipped, making reading difficult. However, when subjects forced themselves to move the gaze slowly (keeping the eyes stationary relative to the head and rotating the head slowly, for example), it was possible to read using the gaze-controlled magnifier. Another possibility would be to incorporate a filter that transforms the step-like saccades into smooth ramp-like movements of the magnifier. In any case, training in the use of this visual aid is a key issue.

The visual aid has been recently tested with an 87 year-old patient (retired professor) who suffers from ARMD. The patient had a small lesion in the right eye and a large lesion in the left eye. Using the right eye, the patient could read large print and newspaper headlines without any visual aid other than his eyeglasses (visual acuity of 20/64 as tested with the ETDRS chart at 1 m), while the use of his left eye was very limited (considered legally blind). Using the proposed visual aid, the subject was able to read from the computer monitor using the eye that is considered legally blind (the other eye was covered by an eye-patch). This result is certainly encouraging. The experience with patients with ARMD will provide valuable information to improve the implementation of the visual aid. The point-of-gaze estimation system may be used in the first stage of the experiments to evaluate fixation stability and visual scanning patterns in patients with ARMD.

---

[6] A saccade is a fast eye movement that shifts the point-of-gaze.

5.3 Conclusions

5.3.1 Contributions

The main contributions of this thesis are the following:

- Development of a novel model-based free-head point-of-gaze estimation system.

A system that estimates the point-of-gaze in real time using a geometrical-optical model was developed. The system does not require head attachments and allows for limited head movement. The current implementation of the system estimates the point-of-gaze at a rate of up to 12 Hz while allowing for head movements in a volume of 6 cm laterally by 4 cm vertically by 8 cm longitudinally. The range of head movements is limited only by the field of view of the camera. The development of the point-of-gaze estimation system comprised the following milestones:

- Development of a geometrical-optical model of the system and the eye that relates the coordinates of pupil and glints centers in an eye image with the point-of-gaze on a computer monitor. This model is used to estimate the point-of-gaze from the estimated coordinates of pupil and glints centers, and to compute the coordinates of pupil and glints centers for a given eye position in space and a point-of-gaze on the screen. This model overcomes the problem of estimation error due to head displacement.
- Development of a calibration procedure and evaluation of the sensitivity of the point-of-gaze estimation method to errors in model parameters, noise in the estimation of pupil and glints centers, and fixation errors during calibration.
- Development of a non-spherical corneal model to study the robustness of the proposed point-of-gaze estimation method.
- Development of sub-optimal solutions to improve the point-of-gaze estimation accuracy.
- Experimental evaluation of the point-of-gaze estimation system. These experiments demonstrated that the root mean square error of the estimation of the point-of-gaze on the computer monitor is less than 1° of visual angle.

- Proposal of a novel visual aid for patients with age-related macular degeneration (ARMD).

  A novel visual aid to allow patients with ARMD to read from a computer monitor was proposed. The concept was evaluated by simulating the disease in subjects with healthy retina. The visual aid has been recently tested with a patient with ARMD with encouraging results.

5.3.2 Future work

The current implementation of the system allows for head movements in a volume of 6 cm laterally by 4 cm vertically by 8 cm longitudinally. The range of head movements is limited only by the field of view of the camera. This figures are more than enough if a chinrest is used to stabilize the head or if the system is used by motor disabled persons who cannot move the head. However, for applications involving able-bodied persons, the possibility of allowing for natural head movement is in general desired and, hence, a larger volume of allowed head movement is required. This can be accomplished with a combination of a higher resolution image sensor and a moving camera. An image sensor with a higher resolution will allow a larger field of view of the camera while keeping the effective spatial resolution of the camera approximately constant.

At this stage, the point-of-gaze can be estimated at a rate of up to 12 Hz. This is adequate for applications such as a gaze-slaved mouse or magnifier but other applications require higher estimation rates. One example is the study of changes in the visual scanning patterns in patients with macular degeneration for which an estimation rate of 30 Hz and even 60 Hz is desirable if saccadic eye movements are to be analyzed. This may be accomplished by the convergence of the optimization of the estimation algorithms and the evolution of computer technology.

Experimental results showed that the root mean square error in the estimation of the point-of-gaze on the computer monitor was slightly less than 1° of visual angle. This is sufficient for applications such as those described in [3, 4, 5] where the subject is presented with slides containing four pictures of different themes and the objective is to determine the visual preference (dwell time, number of visits, etc.) and correlate it with the level of depression. On the other hand, for applications such as a gaze-slaved pointing device for motor disabled persons or a gaze-slaved magnifier for persons with macular degeneration, better accuracy in the estimation of the point-of-gaze is desired, especially if the amount of information presented on the screen is not to be compromised (i.e. the size and separation of icons or the line spacing of text can be increased but at the cost of reducing the amount of information presented simultaneously).

The results obtained in this research showed that a spherical corneal model is not, in general, a good model for real corneas and hence, a more complex corneal model, such as an ellipsoid or a higher order model could be used to improve the accuracy of the point-of-gaze estimation. The use of a more complex corneal model to estimate the point-of-gaze in real time will require increased computing power, which may be available in the near future with the fast evolution of computer technology. In addition, some preliminary simulation results showed that the robustness of the point-of-gaze estimation system can be improved by locating the light sources higher relative to the camera. However, such changes in the physical configuration of the system increase the probability of interference from the upper eyelid (shadow in the iris) that affects the performance of the image processing algorithms. It is planned to consider a more elaborate illumination configuration in which the light sources that produce the corneal reflections are located higher and a diffuse IR light source is used to overcome the eyelid interference problem. Another issue that needs further exploration is the calibration algorithm. This involves, possibly, more calibration target points, as well as reviewing which parameters should be included in the calibration procedure and the formulation of more elaborate calibration constraints.

The current geometrical-optical model does not account for eyeglasses. Eyeglasses are refractive elements that alter the path of the light rays both from the light source to the cornea and back from the cornea to the camera. If the eyeglasses remain stationary relative to the system, which in general implies that the head does not move relative to the system, the estimation of the point-of-gaze is hardly affected by the eyeglasses. If the eyeglasses move relative to the system, the extent by which the optical path of the light rays is affected can change significantly and produce estimation errors of up to 2-3º (in the conditions tested). This problem will be addressed in the future. Another aspect of eyeglasses is that each lens and the frame of the eyeglasses can produce several reflections than can interfere with the image of the eye. These bright reflections can completely mask the pupil, rendering the image of the eye unusable. This issue is not as dramatic if the eyeglasses have an antireflective coating. Nevertheless, it is an issue to deal with and the location of the light sources could be a factor.

The development of the novel visual aid for patients with age-related macular degeneration (ARMD) might benefit from an improved point-of-gaze estimation system that can be used to control the eye-slaved magnifier. In addition, the point-of-gaze estimation system can be used to study fixation stability and visual scanning patterns in patients with ARMD.

Bibliography

[1] L. R. Young and D. Sheena, “Methods and designs - Survey of eye movement recording methods,” *Behavior Research Methods & Instrumentation*, vol. 7, no. 5, pp. 397-429, 1975.

[2] R. M. Berne and M. N. Levy (Eds.), *Principles of Physiology*, 3rd ed. St. Louis, MO: Mosby, 2000.

[3] L. H. Yu, “Point-of-gaze reconstruction in head-mounted eye tracking systems,” M.A.Sc. Thesis, Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Canada, 2002.

[4] L. H. Yu and M. Eizenman, “A new methodology for determining point-of-gaze in head-mounted eye tracking systems,” *IEEE Transactions on Biomedical Engineering*, to be published.

[5] M. Eizenman, L. H. Yu, L. Grupp, E. Eizenman, M. Ellenbogen, M. Gemar, and R. D. Levitan, “A naturalistic visual scanning approach to assess selective attention in major depressive disorder,” *Psychiatry Research*, to be published.

[6] C. Karatekin and R. F. Asarnow, “Exploratory eye movements to pictures in childhood-onset schizophrenia and attention-deficit/hyperactivity disorder (ADHD),” *Journal of Abnormal Child Psychology*, vol. 27, no. 1, pp. 35-49, Feb. 1999.

[7] M. De Luca, E. Di Pace, A. Judica, D. Spinelli, and P. Zoccolotti, “Eye movement patterns in linguistic and non-linguistic tasks in developmental surface dyslexia,” *Neuropsychologia*, vol. 37, no. 12, pp. 1407-1420, Nov. 1999.

[8] M. Eizenman, T. Jares, and A. Smiley, “A new methodology for the analysis of eye movements and visual scanning in drivers,” in *Proc. of the 31st Annual Conference on Ergonomics and Safety, Hall, Quebec, Canada*, 1999.

[9] A. Zelinsky, "Visual human-machine interaction," in *Advanced Topics in Artificial Intelligence, Lecture Notes in Artificial Intelligence 1747*, N. Foo, Ed. Berlin: Springer-Verlag, 1999, pp. 440-452.

[10] J-C. Popieul, P. Simon, R. Leroux, and J-C. Angué, "Automatic processing of a car driver eye scanning movements on a simulated highway driving context," in *Proc. of the IEEE Intelligent Vehicles Symposium 2000*, 2000, pp. 13-18.

[11] J. L. Harbluk, I. Y. Noy, and M. Eizenman, "The impact of cognitive distraction on driver visual and vehicle control," in *Proc. of the Transportation Research Board 81st Annual Meeting, Washington, DC, USA*, Jan. 2002.

[12] D. Cleveland, "Unobtrusive eyelid closure and visual point of regard measurement system," in *Proc. of the Technical Conference on Ocular Measures of Driver Alertness, sponsored by The Federal Highway Administration – Office of Motor Carrier and Highway Safety and The National Highway Traffic Safety Administration – Office of Vehicle Safety Research, Herndon, VA, USA*, 1999, pp. 57-74.

[13] P. Smith, M. Shah, and N. da Vitoria Lobo, "Monitoring head/eye motion for driver alertness with one camera," in *Proc. of the 15th International Conference on Pattern Recognition*, 2000, vol. 4, pp. 636-642.

[14] P. A. Wetzel, G. Krueger-Anderson, C. Poprik, and P. Bascom, "An eye tracking system for analysis of pilots' scan paths," United States Air Force Armstrong Laboratory, Tech. Rep. AL/HR-TR-1996-0145, Apr. 1997.

[15] J. H. Goldberg and X. P. Kotval, "Computer interface evaluation using eye movements: methods and constructs," *International Journal of Industrial Ergonomics*, vol. 24, no. 6, pp. 631-645, Oct. 1999.

[16] R. Sharma, V. I. Pavlović, and T. S. Huang, "Toward multimodal human-computer interface," *Proc. of the IEEE*, vol. 86, no. 5, pp. 853-869, May 1998.

[17] T. E. Hutchinson, K. P. White, W. N. Martin, K. C. Reichert, and L. A. Frey, "Human-computer interaction using eye-gaze input," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 19, no. 6, pp. 1527-1534, Nov./Dec. 1989.

[18] L. A. Frey, K. P. White, and T. E. Hutchinson, "Eye-gaze word processing," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 20, no. 4, pp. 944-950, Jul./Aug. 1990.

[19] R. S. Allison, M. Eizenman, and B. S. K. Cheung, “Combined head and eye tracking system for dynamic testing of the vestibular system,” *IEEE Transactions on Biomedical Engineering*, vol. 43, no. 11, pp. 1073-1082, Nov. 1996.

[20] J. Heinzmann and A. Zelinsky, “3-D facial pose and gaze point estimation using a robust real-time tracking paradigm,” in *Proc. of the Third IEEE International Conference on Automatic Face and Gesture Recognition*, 1998, pp. 142-147.

[21] R. Newman, Y. Matsumoto, S. Rougeaux, and A. Zelinsky, “Real-time stereo tracking for head pose and gaze estimation,” in *Proc. of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, 2000, pp. 122-128.

[22] Y. Matsumoto and A. Zelinsky, “An algorithm for real-time stereo vision implementation of head pose and gaze direction measurement,” in *Proc. of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, 2000, pp. 499-504.

[23] J. G. Wang and E. Sung, “Gaze determination via images of irises,” *Image and Vision Computing*, vol. 19, no. 12, pp. 891-911, Oct. 2001.

[24] J. G. Wang and E. Sung, “Study on eye gaze estimation,” *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics,* vol. 32, no. 3, pp. 332-350, Jun. 2002.

[25] A. Cowey, “The basis of a method of perimetry with monkeys,” *The Quarterly Journal of Experimental Psychology*, vol. 15, part 2, pp. 81-90, 1963.

[26] J. Merchant, R. Morrissette, and J. L. Porterfield, “Remote measurement of eye direction allowing subject motion over one cubic foot of space,” *IEEE Transactions on Biomedical Engineering,* vol. BME-21, no. 4, pp. 309-317, Jul. 1974.

[27] K. P. White, T. E. Hutchinson, and J. N. Carley, “Spatially dynamic calibration of an eye-tracking system,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 23, no. 4, pp. 1162-1168, Jul./Aug. 1993.

[28] Y. Ebisawa, K. Kaneko, S. Kojima, T. Ushikubo, and T. Miyakawa, “Non-invasive eye-gaze position detecting method used on man/machine interface for the disabled,” in *Proc. of the Fourth Annual IEEE Symposium on Computer-Based Medical Systems*, 1991, pp. 374-380.

[29] Y. Ebisawa, “A noncontact eye-gaze point detection method used on support system for the disabled,” in *Proc. of the 14th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 1992, vol. 3, pp. 1168-1170.

[30] A. Sugioka, Y. Ebisawa, and M. Ohtani, “Noncontact video-based eye-gaze detection method allowing large head displacements,” in *Proc. of the 18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 1996, vol. 2, pp. 526-528.

[31] Y. Ebisawa, M. Ohtani, A. Sugioka, and S. Esaki, “Single mirror tracking system for free-head video-based eye-gaze detection method,” in *Proc. of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 1997, vol. 4, pp. 1448-1451.

[32] The EyeGaze development system. A tool for eyetracking applications. Product brochure. (2002, June). LC Technologies, Inc., Fairfax, VA. [Online]. Available: http://www.eyegaze.com/doc/eds.htm

[33] Applied Science Laboratories, Bedford, MA. [Online]. Available: http://www.a-s-l.com

[34] J. Liu, “Determination of the point of fixation in a head-fixed coordinate system,” in *Proc. of the 14th International Conference on Pattern Recognition*, 1998, vol. 1, pp. 501-504.

[35] D. H. Yoo, J. H. Kim, B. R. Lee, and M. J. Chung, “Non-contact eye gaze tracking system by mapping of corneal reflections,” in *Proc. of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, 2002, pp. 94-99.

[36] R. Grace, V. E. Byrne, D. M. Bierman, J. M. Legrand, D. Gricourt, R. K. Davis, J. J. Staszewski, and B. Carnahan, “A drowsy driver detection system for heavy vehicles,” in *Proc. of the Technical Conference on Ocular Measures of Driver Alertness, sponsored by The Federal Highway Administration – Office of Motor Carrier and Highway Safety and The National Highway Traffic Safety Administration – Office of Vehicle Safety Research, Herndon, VA, USA*, 1999, pp. 75-86.

[37] Y. Ebisawa, “Improved video-based eye-gaze detection method,” in *Proc. of the 10th Anniversary IEEE Instrumentation and Measurement Technology Conference*, 1994, vol. 2, pp. 963-966.

[38] M. Ohtani and Y. Ebisawa, "Eye-gaze detection based on the pupil detection technique using two light sources and the image difference method," in *Proc. of the 17th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 1995, vol. 2, pp. 1623-1624.

[39] Y. Ebisawa, "Unconstrained pupil detection technique using two light sources and the image difference method," in *Visualization and Intelligent Design in Engineering and Architecture II*, S. Hernández and C. A. Brebbia, Eds. Southamptom, UK: Computational Mechanics Publications, 1995, pp. 79-89.

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

[41] Vision problems in the U.S. Prevalence of adult vision impairment and age-related eye disease in America (2002). Prevent Blindness America and the National Eye Institute, USA. [Online]. Available: http://www.nei.nih.gov/eyedata/pdf/VPUS.pdf

[42] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 2nd ed. Reading, MA: Addison-Wesley, 1992.

[43] A. M. Slater and J. M. Findlay, "The measurement of fixation position in the newborn baby," *Journal of Experimental Child Psychology*, vol. 14, pp. 349-364, 1972.

[44] A. G. Gale, "A note on the remote oculometer technique for recording eye movements," *Vision Research*, vol. 22, no. 1, pp. 201-202, 1982.

[45] R. Fletcher, *Practical Methods of Optimization*. Chichester, UK: John Wiley & Sons, 1987.

[46] *Optimization Toolbox User's Guide – Version 2*. Natick, MA: The MathWorks, Jun. 2001.

[47] B. J. M. Lui, "A point-of-gaze estimation system for studies of visual attention," M.A.Sc. Thesis, Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Canada, 2003.

[48] A. I. Tew, "Simulation results for an innovative point-of-regard sensor using neural networks," *Neural Computing & Applications*, vol. 5, no. 4, pp. 230-237, 1997.

[49] D. C. Fletcher and R. A. Schuchard, "Preferred retinal loci relationship to macular scotomas in a low-vision population," *Ophthalmology*, vol. 104, no. 4, pp. 632-638, Apr. 1997.

[50] D. C. Fletcher, R. A. Schuchard, and G. Watson, “Relative locations of macular scotomas near the PRL: Effect on low vision reading,” *Journal of Rehabilitation Research and Development*, vol. 36, no. 4, pp. 356-364, Oct. 1999.

[51] U. L. Nilsson and S. E. G. Nilsson, “Rehabilitation of the visually handicapped with advanced macular degeneration,” *Documenta Ophthalmologica*, vol. 62, pp. 345-367, 1986.

[52] P. J. Beckmann and G. E. Legge, “Psychophysics of reading – XIV. The page navigation problem in using magnifiers,” *Vision Research*, vol. 36, no. 22, pp. 3723-3733, 1996.

[53] J. E. Farrell and M. Desmarais, “Equating character-identification performance across the visual field,” *Journal of the Optical Society of America A*, vol. 7, no. 1, pp. 152-159, Jan. 1990.

[54] G. E. Legge, D. G. Pelli, G. S. Rubin, and M. M. Schleske, “Psychophysics of reading – I. Normal vision,” *Vision Research*, vol. 25, no. 2, pp. 239-252, 1985.

Appendix A: Transformation between world and camera coordinate systems

This appendix presents a more traditional way in which the transformation from the world coordinate system to the camera coordinate system is considered as a sequence of steps:

(i) translate the world coordinate system XYZ by $\mathbf{t}$ to obtain the coordinate system $X_\mathbf{t}Y_\mathbf{t}Z_\mathbf{t}$ (Fig. A.1),

(ii) rotate the coordinate system $X_\mathbf{t}Y_\mathbf{t}Z_\mathbf{t}$ around the $Y_\mathbf{t}$-axis by an angle θ_{cam} to obtain the coordinate system $X_\mathbf{t}'Y_\mathbf{t}'Z_\mathbf{t}'$ (Fig. A.2),

(iii) rotate the coordinate system $X_\mathbf{t}'Y_\mathbf{t}'Z_\mathbf{t}'$ around the $X_\mathbf{t}'$-axis by an angle φ_{cam} to obtain the coordinate system $x_0y_0z_0$ (Fig. A.3),

(iv) rotate the coordinate system $x_0y_0z_0$ around the z_0-axis by an angle κ_{cam} to obtain the camera coordinate system xyz (Fig. A.4).

Step (i) is described by

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} X_\mathbf{t} \\ Y_\mathbf{t} \\ Z_\mathbf{t} \end{bmatrix} + \mathbf{t} \ . \tag{A.1}$$

Step (ii) is modeled as

$$\begin{bmatrix} X_\mathbf{t} \\ Y_\mathbf{t} \\ Z_\mathbf{t} \end{bmatrix} = \mathbf{A} \begin{bmatrix} X_\mathbf{t}' \\ Y_\mathbf{t}' \\ Z_\mathbf{t}' \end{bmatrix} \tag{A.2}$$

where

$$\mathbf{A} = \begin{bmatrix} \cos\theta_{cam} & 0 & \sin\theta_{cam} \\ 0 & 1 & 0 \\ -\sin\theta_{cam} & 0 & \cos\theta_{cam} \end{bmatrix} . \tag{A.3}$$

Step (iii) is given by

$$\begin{bmatrix} X_{\mathbf{t}}' \\ Y_{\mathbf{t}}' \\ Z_{\mathbf{t}}' \end{bmatrix} = \mathbf{B} \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} \qquad \textbf{(A.4)}$$

where

$$\mathbf{B} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi_{cam} & \sin\varphi_{cam} \\ 0 & -\sin\varphi_{cam} & \cos\varphi_{cam} \end{bmatrix} . \qquad \textbf{(A.5)}$$

And step (iv) can be written as

$$\begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} = \mathbf{C} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \qquad \textbf{(A.6)}$$

where

$$\mathbf{C} = \begin{bmatrix} \cos\kappa_{cam} & -\sin\kappa_{cam} & 0 \\ \sin\kappa_{cam} & \cos\kappa_{cam} & 0 \\ 0 & 0 & 1 \end{bmatrix} . \qquad \textbf{(A.7)}$$

Equations **(A.1)**, **(A.2)**, **(A.4)** and **(A.6)** can then be chained to get

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \mathbf{ABC} \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \mathbf{t} \; . \qquad \textbf{(A.8)}$$

Noting that the transformation matrices **A**, **B** and **C** are orthonormal, which implies that $\mathbf{A}^{-1} = \mathbf{A}^T$ and so on, we can write the inverse equation as

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \mathbf{C}^T \mathbf{B}^T \mathbf{A}^T \left(\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} - \mathbf{t} \right) = (\mathbf{ABC})^T \left(\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} - \mathbf{t} \right) . \qquad \textbf{(A.9)}$$

Note that the comparison between equations **(2.21)** and **(A.8)** yields

$$\mathbf{ABC} = \begin{bmatrix} \mathbf{i}_{cam} & \mathbf{j}_{cam} & \mathbf{k}_{cam} \end{bmatrix} . \qquad \textbf{(A.10)}$$

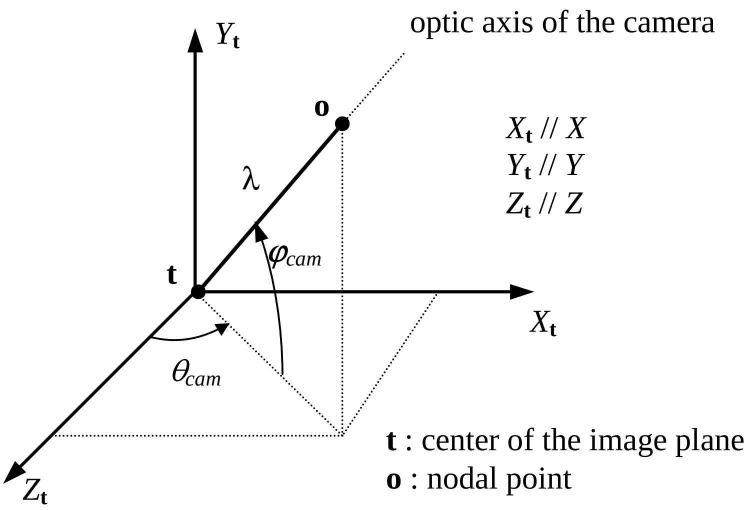


Figure A.1: Translation by **t**.

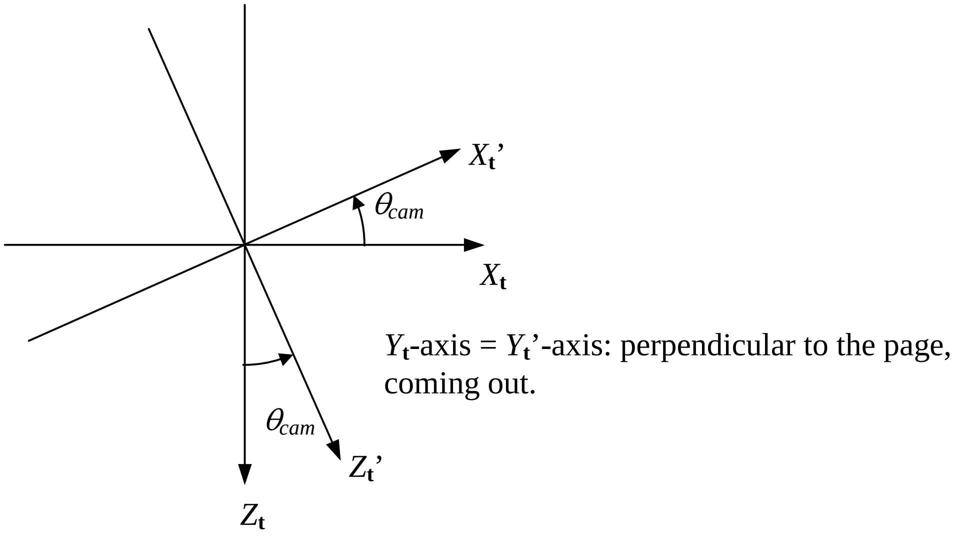


Figure A.2: Rotation of the camera around the $Y_\mathbf{t}$-axis by an angle θ_{cam}.

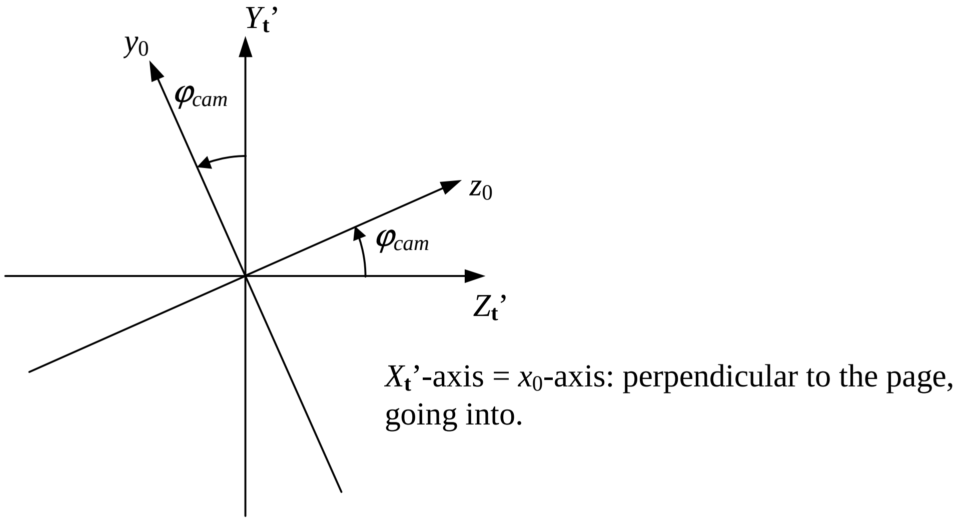


Figure A.3: Rotation of the camera around the $X_{\mathbf{t}}$'-axis by an angle φ_{cam}.

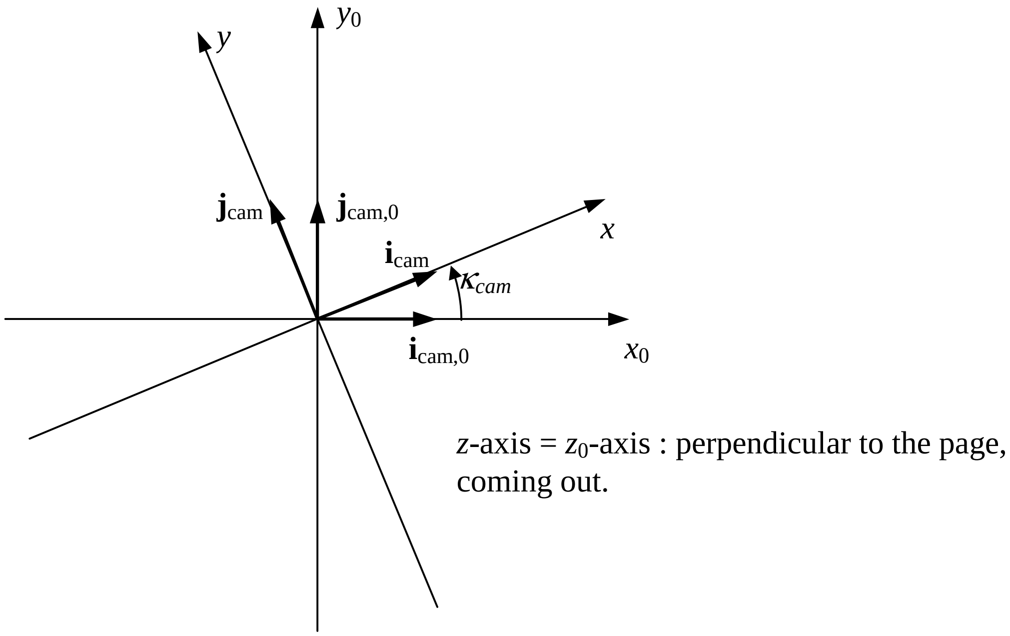


Figure A.4: Rotation of the camera around its optic axis by an angle κ_{cam}.

Appendix B: Typical values of the eye parameters and nominal values of the system parameters

Parameter	Description	Typical / Nominal value	Reference / Comments
D	Distance between the center of corneal curvature and the center of rotation of the eye	5.3 mm	[1]
R	Radius of corneal curvature	7.8 mm	[1, 44]
K	Distance between the center of the pupil and the center of corneal curvature	4.75 mm	[44]
α_{eye}	Horizontal angle between visual and optic axes of the eye	–5º for the right eye, 5º for the left eye	[43]
β_{eye}	Vertical angle between visual and optic axes of the eye	1.5º	[43]
θ_{cam}	Pan angle of the camera	0º	
φ_{cam}	Tilt angle of the camera	27º	
κ_{cam}	Angle of rotation of the camera around its optic axis	0º	
$\mathbf{l}$	Light source 1	$[-249.1\ -142.2\ 5.4]^T$ mm	
$\mathbf{m}$	Light source 2	$[249.1\ -142.2\ 5.4]^T$ mm	
$\mathbf{t}$	Center of the camera image plane	$[0\ -232.16\ 48.38]^T$ mm*	
focal length	Focal length of the camera lens	35 mm	
dcnp	Distance between the cornea and the nodal point of the camera	625 mm	
c_{center}	Column coordinate of the image center	320.5 pixels	
r_{center}	Row coordinate of the image center	240.5 pixels	
pixel pitch	Distance between adjacent pixels in the CCD	7.4 µm	

Table B.1: Typical values of the eye parameters and nominal values of the system parameters.

*: The value of **t** is computed as

$$\mathbf{t} = \begin{bmatrix} 0 \\ -263.9 \\ 20.45 \end{bmatrix} + 42.28 \begin{bmatrix} 0 \\ \sin(21.65^\circ + \varphi_{cam}) \\ \cos(21.65^\circ + \varphi_{cam}) \end{bmatrix} \qquad \textbf{(B.1)}$$

Appendix C: Change of variables for the calibration problem

In Section 3.4, the root mean squared error of the estimation of the point-of-gaze was studied as a function of the deviation of the actual value of each calibration parameter from the typical/nominal value used to estimate the point-of-gaze. Each parameter was individually considered in a neighborhood of its typical/nominal value and the root mean squared error showed to be a linear or approximately linear function of the magnitude of the deviation of the parameter value. This implies that the mean squared error of the estimation of the point-of-gaze as a function of the deviation of each calibration parameter from the typical/nominal value can be described by a quadratic or approximately quadratic function. Hence, the objective function of the calibration problem

$$f(R, K, \alpha_{eye}, \beta_{eye}, \theta_{cam}, \kappa_{cam}) = \sum_{i=1}^{N} \left\| \hat{\mathbf{g}}(\hat{\mathbf{f}}_i, [R \;\; K \;\; \alpha_{eye} \;\; \beta_{eye} \;\; \theta_{cam} \;\; \kappa_{cam} \;\; \boldsymbol{\rho}_{fixed}]) - \mathbf{g}_i \right\|^2 \qquad \textbf{(C.1)}$$

can be approximated by a quadratic function of six variables, at least in a neighborhood of the optimum point. The vectors $\hat{\mathbf{f}}_i$ are the vectors of coordinates of pupil and glints centers obtained for each calibration point $\mathbf{g}_i$, $\boldsymbol{\rho}_{fixed}$ is the vector of fixed system parameters, and the calibration parameters are given explicitly.

It was also observed that for each parameter the slope of the r.m.s. error as a function of the deviation of the parameter from the typical/nominal value is different and that there are interactions among parameters. This implies then that the level hyper-surfaces of the objective function are approximately hyper-ellipsoids whose axes are, in general, not parallel to the axes corresponding to the design variables (i.e. the calibration parameters). Conditioning and, hence, convergence of the calibration problem could be improved if a change of variables, namely an affine transformation, were applied to the objective function so that the level hyper-surfaces of the objective function become approximately hyper-spherical. However, since the objective function has six variables, this problem is difficult and computationally expensive to analyze. For this reason, only a simplified analysis for a particular simulated situation and taking the variables in pairs according to their interactions is presented in this appendix. This analysis is carried out under the assumption that the actual optimum point of the calibration problem is very close to the point corresponding to the typical/nominal values of the calibration parameters.

For this analysis, the vectors $\hat{\mathbf{f}}_i$ of coordinates of pupil and glints centers, that are used in the evaluation of the objective function, are computed with equation **(2.67)** as $\hat{\mathbf{f}}_i = \mathbf{f}(\mathbf{g}_i, \mathbf{d}, \boldsymbol{\rho}_\circ)$, where $\mathbf{d}$ = $[0\ 70\ 650]^T$ mm, the calibration points $\mathbf{g}_i$ are listed in table 3.3 (N = 9) and $\boldsymbol{\rho}_\circ$ is the vector of eye and system parameters with the typical/nominal values given in Appendix B.

One of the observations made in Section 3.4 is that the parameters R and K have both offset and gain effects and that they can be considered coupled to each other as changes of the same proportions and in the same direction in both parameters have opposite effects that cancel each other out to a certain degree. Fig. C.1 shows a plot of the objective function as a function of R and K with all other parameters fixed at their typical / nominal values, that is, $f(R, K, \alpha_{eye,typ}, \beta_{eye,typ}, \theta_{cam,nom}, \kappa_{cam,nom})$. Clearly, there is a valley where the value of the objective function is small. Moreover, if that surface is re-plotted as a function of

$$R_{\circ,norm} = \frac{R - R_{typ}}{R_{typ}} = \frac{R}{R_{typ}} - 1 \tag{C.2}$$

and

$$K_{\circ,norm} = \frac{K - K_{typ}}{K_{typ}} = \frac{K}{K_{typ}} - 1\ , \tag{C.3}$$

where $R_{\circ,norm}$ and $K_{\circ,norm}$ represent the fractional change of R and K relative to their typical values, the valley is in fact at approximately 45º relative to the $R_{\circ,norm}$-axis, in concordance with the observation stated above.

The presence of the valley suggested that convergence could be improved if a change of variables, that makes the surface of Fig. C.1 approximate a circular paraboloid, was introduced. Towards this end, two new variables X_{RK} and Y_{RK} are defined such that the X_{RK}-

axis is coincident with the valley and the Y_{RK}-axis is perpendicular to the valley. Formally, the variables $R_{\circ,norm}$ and $K_{\circ,norm}$ are related to the new variables by

$$\begin{bmatrix} R_{\circ,norm} \\ K_{\circ,norm} \end{bmatrix} = \frac{\sqrt{2}}{2} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} X_{RK} \\ Y_{RK} \end{bmatrix} . \tag{C.4}$$

The next step is to scale the new variables so that the level curves are approximately circular rather than elliptical. The X_{RK}-axis can then be compressed by a factor k_{RK} such that the level curves are approximately circular. Formally, this is done by introducing other two new variables x_1 and x_2, such that

$$X_{RK} = k_{RK} x_1 \ , \tag{C.5}$$

$$Y_{RK} = x_2 \ . \tag{C.6}$$

For example, the level curve for $f = 1$ is approximately elliptical with a ratio between the lengths of the major axis (direction of the valley) and the minor axis (perpendicular to the valley) of about 19. Thus, a value of $k_{RK} = 19$ was adopted.

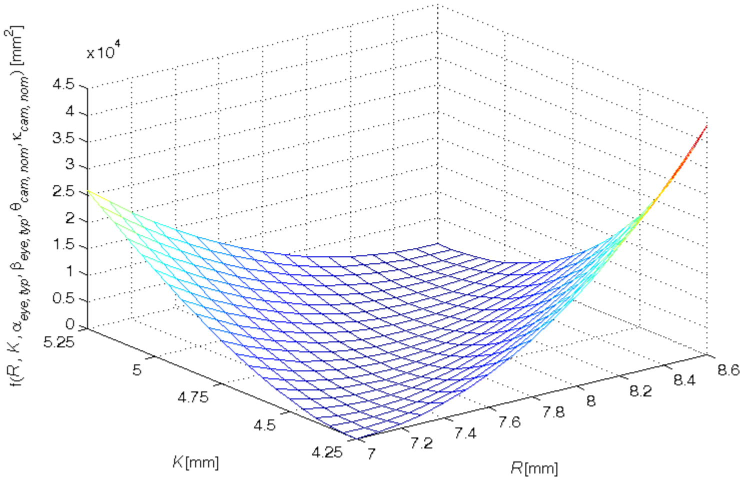


Figure C.1: Objective function plotted as a function of *R* and *K*.

Another observation that was made in Section 3.4 is that the parameters α_{eye} and β_{eye} have an offset effect but their effects are independent of each other. This is in fact what can be seen from Fig. C.2, where the objective function has been plotted as a function of α_{eye} and β_{eye} while all other parameters remained fixed at their typical / nominal values, that is, $f(R_{typ}, K_{typ}, \alpha_{eye}, \beta_{eye}, \theta_{cam,nom}, \kappa_{cam,nom})$.

In this case, the level curves are already circular. However, the variables should be scaled such that the level curves have approximately the same diameters as for the objective function as a function of x_1 and x_2. Formally, this is done, by translating to the origin and introducing two new variables x_3 and x_4, as

$$\alpha_{eye} - \alpha_{eye,typ} = k_{\alpha\beta} x_3 \ , \tag{C.7}$$

$$\beta_{eye} - \beta_{eye,typ} = k_{\alpha\beta} x_4 \ . \tag{C.8}$$

Comparing the level curves for $f = 1$, it is seen that a compression by a factor of about 34 is required. Hence, $k_{\alpha\beta} = 34$ was adopted.

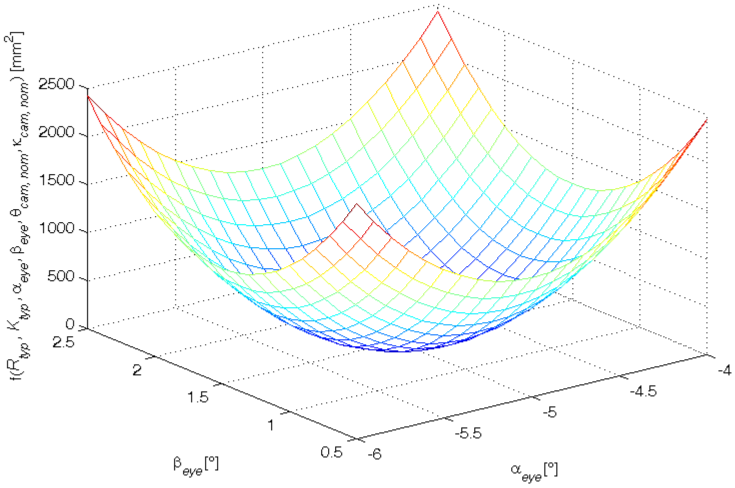


Figure C.2: Objective function plotted as a function of α_{eye} and β_{eye}.

It was also seen in Section 3.4 that θ_{cam} and κ_{cam} have rotation-like effects and that they are coupled to each other to a certain extent. Fig. C.3 shows a plot of the objective function as a function of θ_{cam} and κ_{cam} with all other parameters fixed at their typical values, that is, $f(R_{typ}, K_{typ}, \alpha_{eye,typ}, \beta_{eye,typ}, \theta_{cam}, \kappa_{cam})$. This case also presents a valley where the value of the objective function is small, evidence of some mutual compensation between θ_{cam} and κ_{cam}.

The presence of this valley also suggested that convergence could be improved if a change of variables such that the level curves become approximately circular, with a scaling consistent with that of the other variables, was introduced. With this purpose, two new variables $X_{\theta\kappa}$ and $Y_{\theta\kappa}$ are defined such that the $X_{\theta\kappa}$-axis is coincident with the valley while the

$Y_{\theta\kappa}$-axis is perpendicular to the valley. If the orientation of the valley in the θ_{cam}-κ_{cam} coordinate system relative to the θ_{cam}-axis is represented by ψ, the relation between the old variables and the new variables can be formally expressed as

$$\begin{bmatrix} \theta_{cam} \\ \kappa_{cam} \end{bmatrix} = \begin{bmatrix} \cos\psi & -\sin\psi \\ \sin\psi & \cos\psi \end{bmatrix} \begin{bmatrix} X_{\theta\kappa} \\ Y_{\theta\kappa} \end{bmatrix} \tag{C.9}$$

where $\psi \cong -25.32°$.

The next step is to compress along the valley by a factor $k_{\theta\kappa,1}$ in order to obtain circular level curves, followed by a scaling in both axes by a factor $k_{\theta\kappa,2}$ for consistence with the other variables. This is formally expressed, after introducing two new variables x_5 and x_6, as

$$X_{\theta\kappa} = k_{\theta\kappa,2} k_{\theta\kappa,1} x_5 \ , \tag{C.10}$$

$$Y_{\theta\kappa} = k_{\theta\kappa,2} x_6 \ . \tag{C.11}$$

Analyzing and comparing the level curves for f = 1, it is found that $k_{\theta\kappa,1} \approx 7.4$ and $k_{\theta\kappa,2} \approx 53$, hence these values were adopted.

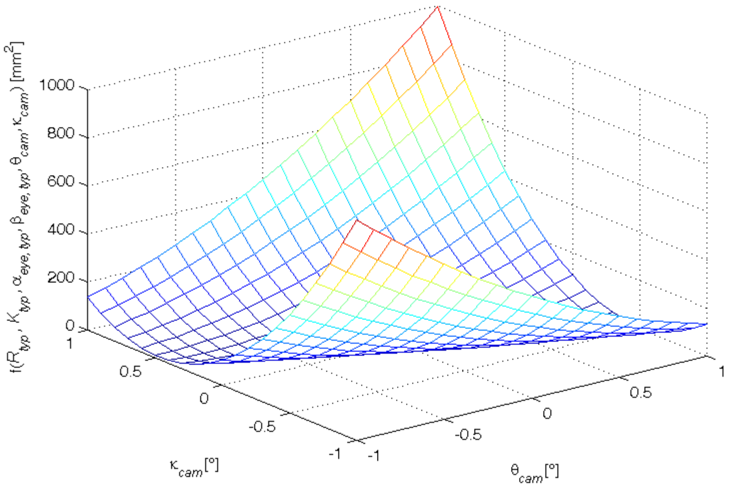


Figure C.3: Objective function plotted as a function of θ_{cam} and κ_{cam}.

The change of variables can be summarized as follows

$$\begin{bmatrix} R \\ K \\ \alpha_{eye} \\ \beta_{eye} \\ \theta_{cam} \\ \kappa_{cam} \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{2}}{2} R_{typ} k_{RK} & -\frac{\sqrt{2}}{2} R_{typ} & 0 & 0 & 0 & 0 \\ \frac{\sqrt{2}}{2} K_{typ} k_{RK} & \frac{\sqrt{2}}{2} K_{typ} & 0 & 0 & 0 & 0 \\ 0 & 0 & k_{\alpha\beta} & 0 & 0 & 0 \\ 0 & 0 & 0 & k_{\alpha\beta} & 0 & 0 \\ 0 & 0 & 0 & 0 & k_{\theta\kappa,1} k_{\theta\kappa,2} \cos\psi & -k_{\theta\kappa,2} \sin\psi \\ 0 & 0 & 0 & 0 & k_{\theta\kappa,1} k_{\theta\kappa,2} \sin\psi & k_{\theta\kappa,2} \cos\psi \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} + \begin{bmatrix} R_{typ} \\ K_{typ} \\ \alpha_{eye,typ} \\ \beta_{eye,typ} \\ \theta_{cam,nom} \\ \kappa_{cam,nom} \end{bmatrix}$$

(C.12)

Then, the objective function in the new variables can be expressed as $f^*(\mathbf{x}) = f^*(x_1, x_2, x_3, x_4, x_5, x_6)$. Figs. C.4, C.5 and C.6 show plots of the new objective function as a function of the new variables, corresponding to Figs. C.1, C.2 and C.3, respectively. The goal of the transformation of variables for this simplified approach has been clearly achieved.

Although this change of variables is not fully justified in general, it shows good behavior even if the solution point is not close to the typical values of the parameters. Increased convergence speed cannot be guaranteed in all cases but the scaling associated with the change of variables is favorable for the computation of the finite difference gradients and the adoption of the search direction in the optimization algorithms, thus making it attractive in practice.

Using the above affine transformation, the calibration problem becomes

$$\min_{\mathbf{x} \in \Re^6} f^*(\mathbf{x}) = \min_{x_1, x_2, x_3, x_4, x_5, x_6} f^*(x_1, x_2, x_3, x_4, x_5, x_6) \ , \qquad \textbf{(C.13)}$$

where the starting point is usually taken as $\mathbf{x} = [0\ 0\ 0\ 0\ 0\ 0]^T$. This problem is solved as a general constrained optimization problem, with the same constraints as in Section 3.2, using a Sequential Quadratic Programming (SQP) algorithm [46]. In this case, a minimum step size on the order of 10^{-7} to 10^{-8} and a maximum step size of 10^{-1} lead to a stable numerical calculation of the gradient.

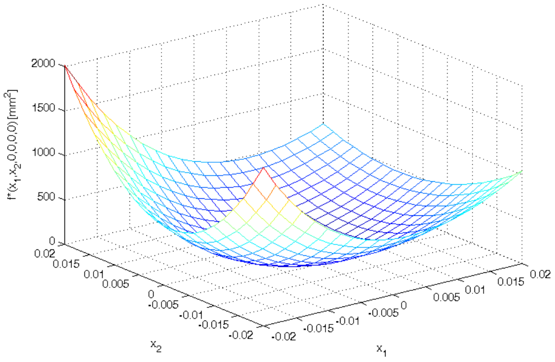


Figure C.4: New objective function as a function of x_1 and x_2.

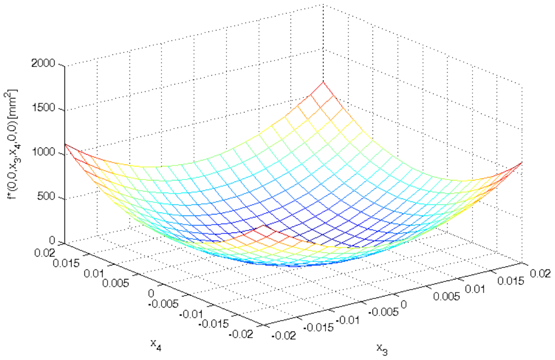


Figure C.5: New objective function as a function of x_3 and x_4.

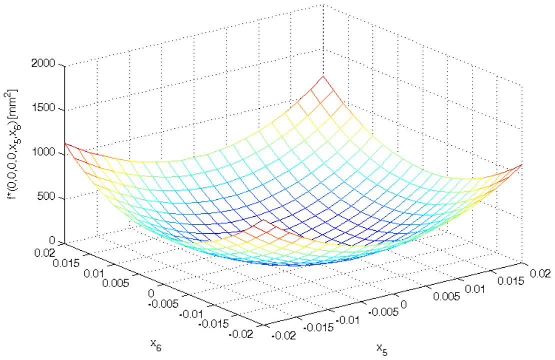


Figure C.6: New objective function as a function of x_5 and x_6.

Finally, another approach is described briefly. This approach is also based on the observations made in Section 3.4 and under the assumption that the actual optimum point of the calibration problem is very close to the point corresponding to the typical/nominal values of the calibration parameters.

It was mentioned above that the objective function is an approximately quadratic function in a neighborhood of the optimum point. With this idea in mind, it can be said that

$$f(\boldsymbol{\rho}_{cal}) \approx \frac{1}{2}(\boldsymbol{\rho}_{cal} - \boldsymbol{\rho}_{cal,\circ})^T \mathbf{H}_\circ (\boldsymbol{\rho}_{cal} - \boldsymbol{\rho}_{cal,\circ}) \qquad \textbf{(C.14)}$$

in a neighborhood of $\boldsymbol{\rho}_{cal,\circ}$, the point that minimizes the objective function. The matrix $\mathbf{H}_\circ$ is the Hessian matrix (matrix of second order derivatives) of the quadratic approximation at $\boldsymbol{\rho}_{cal} = \boldsymbol{\rho}_{cal,\circ}$, and is positive definite since $\boldsymbol{\rho}_{cal,\circ}$ is the minimizer.

Given that this matrix $\mathbf{H}_\circ$ determines the shapes and orientation of the level hyper-surfaces, and hence the conditioning of the problem, the idea is to introduce a change of variable

$$\boldsymbol{\rho}_{cal} - \boldsymbol{\rho}_{cal,\circ} = \mathbf{A}\mathbf{x} \qquad \textbf{(C.15)}$$

such that equation **(C.14)** becomes

$$\tilde{f}^*(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T \mathbf{A}^T \mathbf{H}_\circ \mathbf{A}\mathbf{x} \qquad \textbf{(C.16)}$$

and has level hyper-surfaces that are approximately hyper-spherical. To accomplish this, it is required that

$$\mathbf{A}^T \mathbf{H}_\circ \mathbf{A} = k\mathbf{I} \ , \qquad \textbf{(C.17)}$$

where k is a constant and **I** is a 6 by 6 identity matrix. This equation can be rewritten as

$$\mathbf{H}_\circ = k\mathbf{A}^{-T}\mathbf{A}^{-1} \ . \qquad \textbf{(C.18)}$$

By defining

$$\mathbf{R} = \sqrt{k}\,\mathbf{A}^{-1} \qquad \textbf{(C.19)}$$

the previous equation can be rewritten as

$$\mathbf{H}_\circ = \mathbf{R}^T \mathbf{R} \ , \qquad \textbf{(C.20)}$$

which is the Cholesky factorization of $\mathbf{H}_\circ$. If $\mathbf{H}_\circ$ is computed and the Cholesky factors are obtained, the transformation of variables can be rewritten as

$$\boldsymbol{\rho}_{cal} - \boldsymbol{\rho}_{cal,\circ} = \frac{1}{\sqrt{k}}\mathbf{R}^{-1}\mathbf{x} \ , \qquad \textbf{(C.21)}$$

where the constant k is adopted in consistence with the minimum step size used for the numerical approximation to the gradient used by the optimization algorithm. Note, however, that this procedure requires the numerical computation of the Hessian matrix $\mathbf{H}_\circ$, which is computationally expensive as it requires $3^6 = 729$ function evaluations. In fact, this procedure was tried with the same simulated example and it was found that the numerical computation of $\mathbf{H}_{\circ}$ is sensitive to the step size used, as it can produce an indefinite Hessian matrix even though it is known that $\mathbf{H}_{\circ}$ has to be positive definite. Using a step size of 10^{-5} for R and K and a step size of 10^{-4} for α_{eye}, β_{eye}, θ_{cam} and κ_{cam}, a positive definite Hessian matrix was obtained and the desired goal was attained in a small neighborhood of $\boldsymbol{\rho}_{cal,\circ}$. However, this approach is only justified if the actual solution is in a small neighborhood of $\boldsymbol{\rho}_{cal,\circ}$. In a general calibration situation, this approach may not provide any advantage over the approach discussed at the beginning of this appendix. For this reason, a transformation of variables as described in the first part of this appendix is used for the calibration of the point-of-gaze estimation system. The calibration problem is open to refinements as part of future work.

Appendix D: Samples of eye images

This appendix shows samples of eye images obtained when subjects fixated on the 9 points on the computer screen that are given in table 3.2.

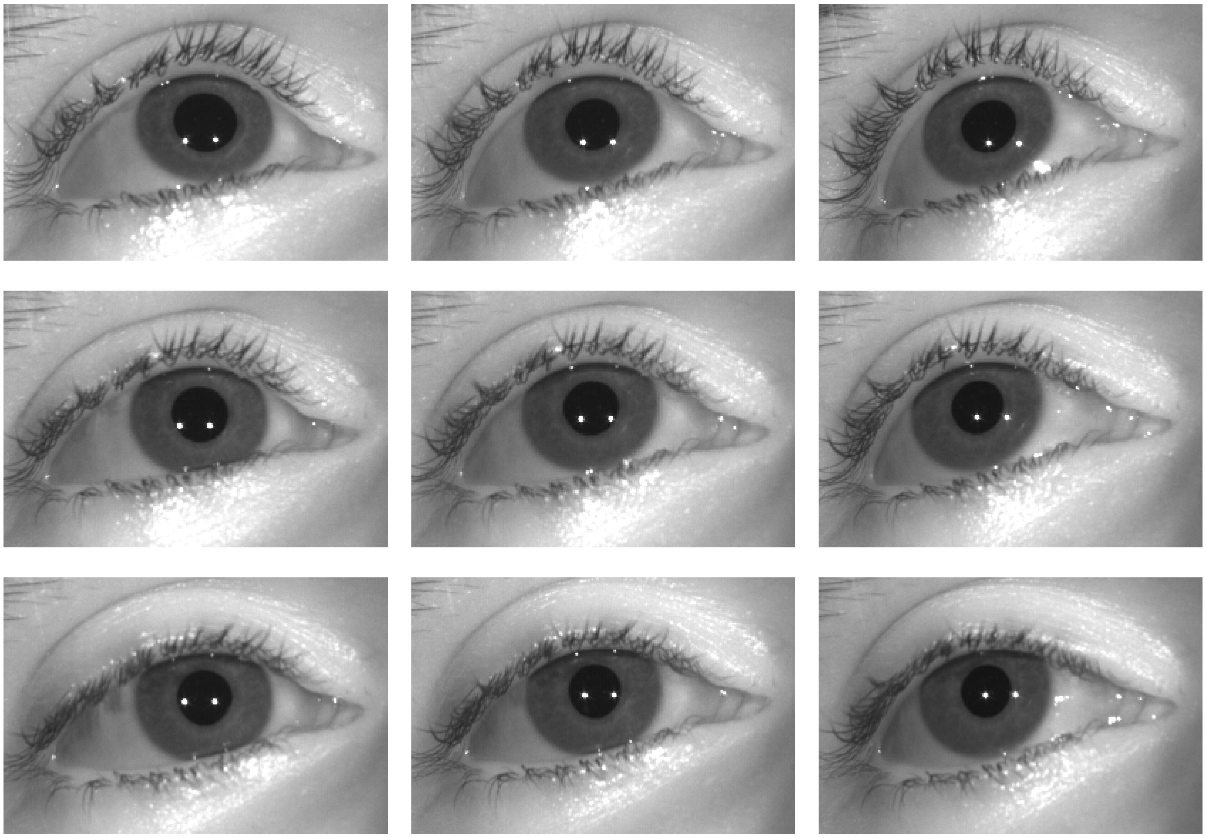


Figure D.1: Sample eye images for subject B.B. (20/20 vision without optical correction).

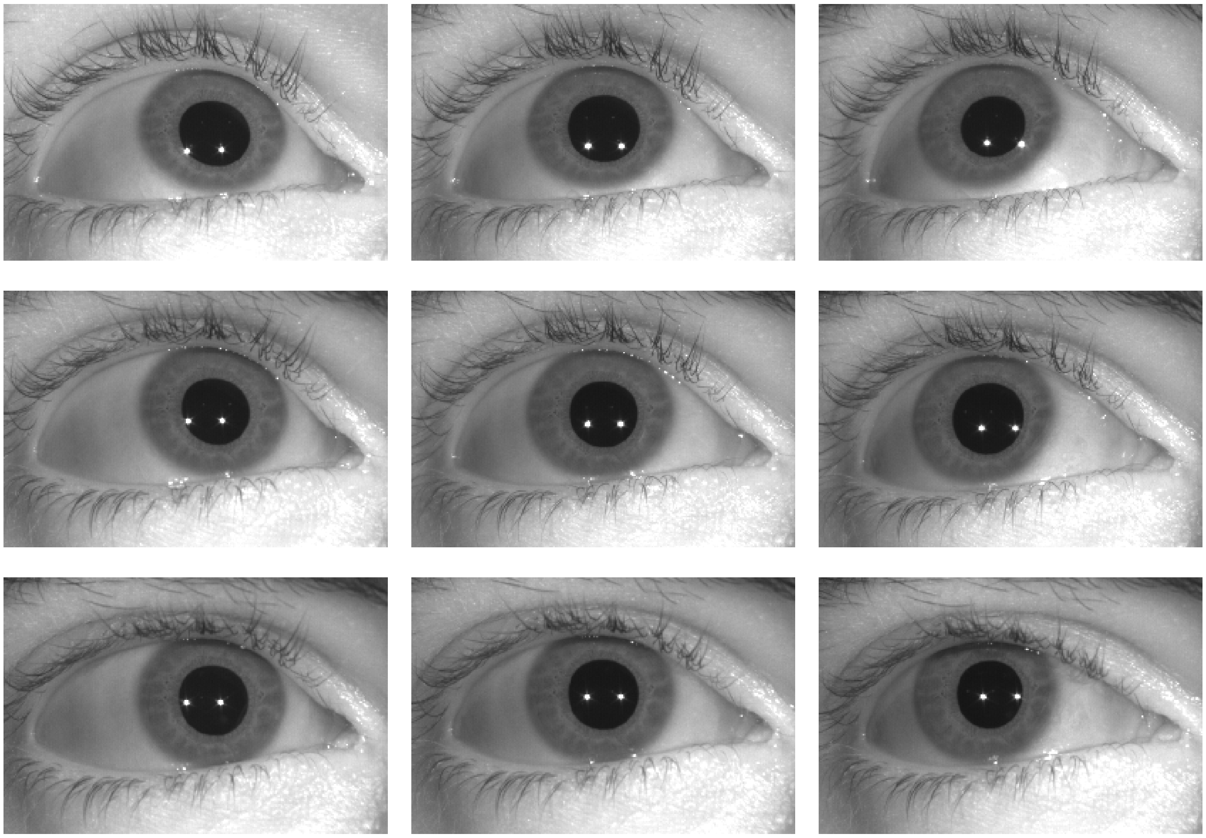


Figure D.2: Sample eye images for subject E.G. (medium myopia) without optical correction.

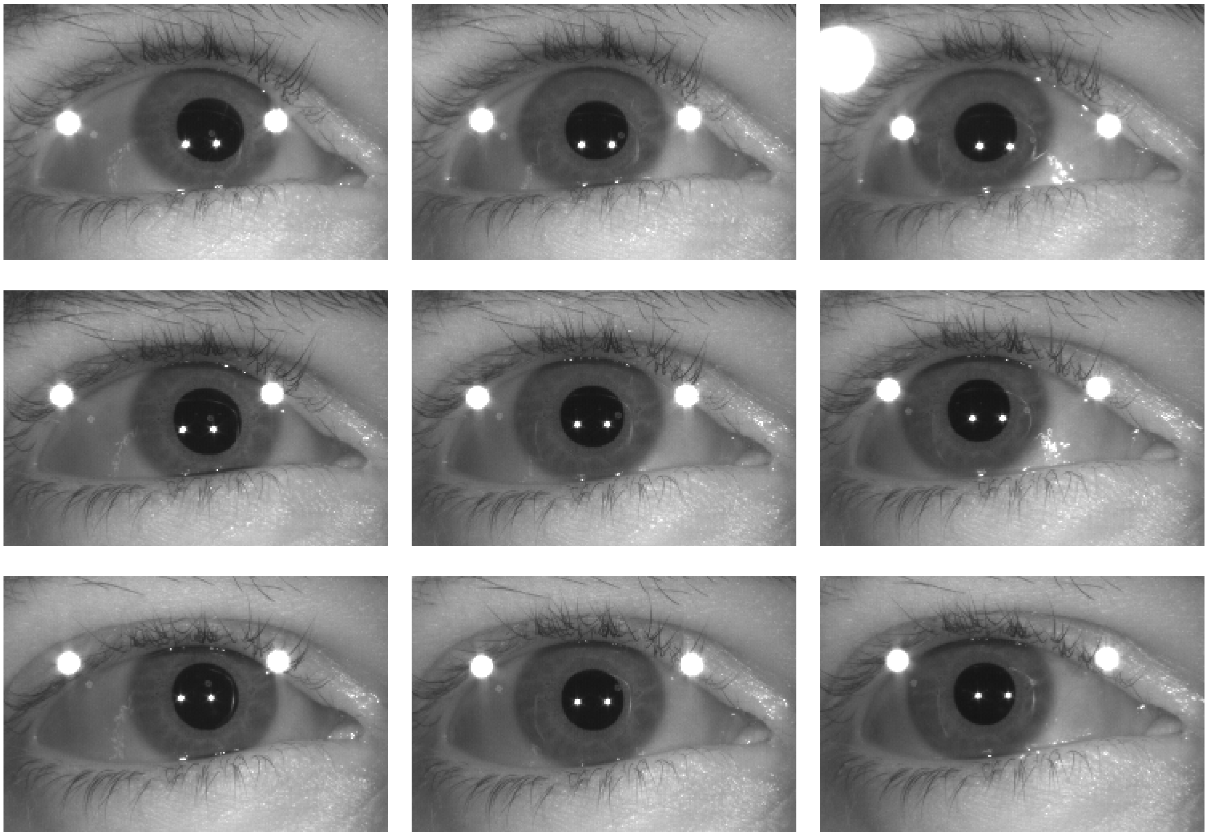


Figure D.3: Sample eye images for subject E.G. (medium myopia) with eyeglasses.

(Note the large reflections produced by the eyeglasses and the reflection of the frame on the cornea).

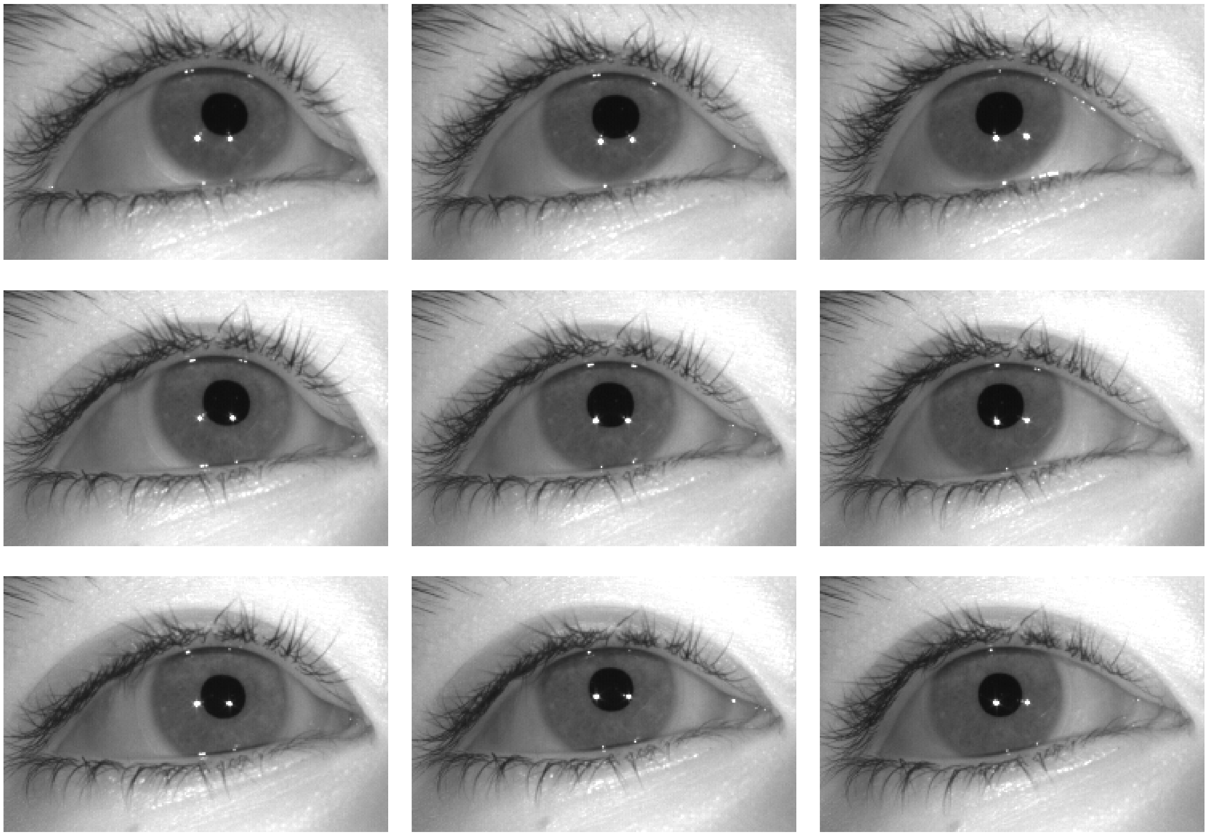


Figure D.4: Sample eye images for subject B.L. (medium myopia) with contact lenses.