Reconstruction of Display and Eyes from a Single Image

Dirk Schnieders, Xingdou Fu and Kwan-Yee K. Wong
Department of Computer Science
University of Hong Kong
{sdirk,xdfu,kykwong}@cs.hku.hk

Abstract

This paper introduces a novel method for reconstructing human eyes and visual display from reflections on the cornea. This problem is difficult because the camera is not directly facing the display, but instead captures the eyes of a person in front of the display. Reconstruction of eyes and display is useful for point-of-gaze estimation, which can be approximated from the 3D positions of the iris and display. It is shown that iris boundaries (limbus) and display reflections in a single intrinsically calibrated image provide enough information for such an estimation. The proposed method assumes a simplified geometric eyeball model with certain anatomical constants which are used to reconstruct the eye. A noise performance analysis shows the sensitivity of the proposed method to imperfect data. Experiments on various subjects show that it is possible to determine the approximate area of gaze on a display from a single image.

1. Introduction

We are a highly visual species, and measuring our eye gaze *(where we are looking)* has a large number of applications in several fields including psychology, industrial engineering and advertising [4]. Despite some recent advances, most systems can only work under a relatively confined and controlled environment. For instance, other than setup-specific calibration in which the relative positions of the cameras, visual display and active illuminations are determined, many systems also require subject-specific calibration. This is typically done by having the subject to fixate on multiple points with known locations in the scene. This might be an obstacle in applications requiring minimal subject cooperation, such as applications with infants [7]. In a commercial state-of-the-art system, the subject-specific calibration will take up to 5 minutes [20]. Some systems require the subject to keep his face considerably still through the use of a bite bar or a forehead support, which might bring physical discomfort. The aforementioned requirements, together with high cost of the hardware (∼US$2,000 - US$45000), greatly hinder the applicability of eye gaze estimation systems and often limit their use to a laboratory setting.

Instead of improving accuracy of existing systems, this paper investigates the minimal information required. By using minimal information, the corresponding hardware setup can be greatly simplified, which in turns results in a simplified and automatic reconstruction.

This work has the following contributions:

1. Estimate the positions of eyes and display from a single image. Specular reflections on the cornea of the eye provide strong constraints on the environment surrounding the subject, and can be exploited to find the positions of objects in front of the subject.

2. No subject-specific parameters for determining where a user is gazing relative to a computer display. Apart from the camera intrinsics, setup-specific parameters do not need to be known.

3. Remove ambiguities by verifying the intersection between optical axis of eyes and display. The elliptical image of the eye's circular limbus can be back-projected into 3D yielding two possible circles. Existing work disambiguates these by making use of anthropomorphic knowledge of the structure of the eyeball [22].

In contrast to commercial technologies, active illumination (such as infrared light) is not employed in this work, and as a result, off-the-shelf equipment can be used.

The rest of this paper is organized as follows. Sect. 2 gives a comparison to existing work. Sect. 3 introduces an approximate eye model, and Sect. 4 reconstructs the eye. Sect. 5 estimates a single edge of the display, and Sect. 6 reconstruct the display through an optimization. Sect. 7 determines the point-of-gaze followed by experimental results on real data in Sect. 8.

978-1-4244-6985-7/10/$26.00 ©2010 IEEE

2. Related work

Early methods for eye movement measurement used contact lenses with embedded coils or rubber rings. Popular in the 1970s, an eye-monitoring technique known as *electroculography* relied on electric measurements from electrodes placed around the eyes for measuring the eye gaze [4]. These crude and highly intrusive techniques of the past have been replaced by refined and relatively user-friendly vision-based methods. Carefully calibrated state-of-the-art methods based on active illumination are capable of recording eye movements with a sampling rate of $2000Hz$ and an accuracy of up to $0.5°$ [20]. Being one of the most important challenges in eye gaze estimation, techniques for reducing the effort of calibration have been studied before.

Multiple views. In the interesting work of Chen et al. [2], a stereo gaze tracking system with two cameras is proposed. In their system, subject-specific parameters are estimated by a simple calibration method involving the subject gazing at only four positions on the display. The work of Guestrin and Eizenman [7] requires an even simpler calibration procedure in which the subject has to fixate on just a single point. The cost for the simple calibration is a custom made setup with two synchronized cameras and four infrared light sources. Shih et al. [18] prove that without information about the cornea and pupil size, at least two cameras and two light sources are needed to recover the eye position and gaze in 3D. Reconstruction of the pupil ellipse in 3D space is done in the work of Kohlbecher et al. [10]. These methods have reduced the subject-specific calibration requirements of early methods significantly by introducing an additional camera. However, having more cameras also means that it is necessary to synchronize the cameras as well as to estimate their relative positions and orientations. The fact that the cameras can usually not see the display further complicates the calibration procedure, and additional hardware and user interaction are needed. Guestrin and Eizenman calibrate their system using a double sided checkerboard pattern and an auxiliary camera, while Chen et al. use a pattern to calibrate the stereo cameras and a planar mirror for the display calibration. Shih et al. additionally require light positions to be known.

Single view. Cross-ratio is an invariant of the projective space and has been used for gaze estimation. In the work of Yoo and Chung [23] and later Kang et al. [9], four infrared LEDs are attached to the corners of a display and another infrared LED is positioned at the center of a zoom lens infrared camera. An additional camera is used to guide the zoom lens camera. Thanks to the use of cross-ratio, their method does not need any setup-specific calibration, and their subject-specific calibration is a simple procedure in which the subject has to fixate on the four LEDs on the display. Other single view gaze-direction estimation methods exist [22]. Since we are interested in the point-of-gaze relative to a display, we will exclude such methods in this discussion.

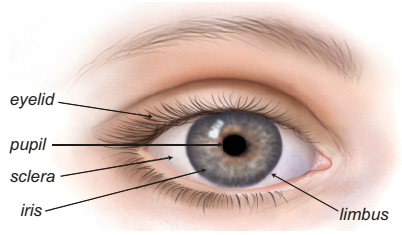

Figure 1. Outer view of human eye. (Illustration taken from [1])

Specular reflection. Reflections can be a powerful cue for pose estimation in computer vision [11]. Francken et al. [6] have used two images and the reflections of gray code illumination patterns on a spherical mirror to calibrate a camera-display setup. Reflections seen on the cornea of the eye are studied by Nishino and Nayar [14]. By relating their corneal imaging system to a catadioptric system, a projection of the environment on the cornea can be computed. Recent advances by Nitschke et al. [15] enable the estimation of display corner positions from multiple images of a user moving around the display. For each subject position, light directions are estimated and finally triangulated to recover the 3D positions of the corners. Motivated by their results, this paper uses reflected curves instead of reflected points, which pose stronger constraints on the 3D edges of the display. Schnieders et al. [17] have shown that line estimation from reflections is ill-conditioned for a single sphere. In this paper, it is shown that two human eyeballs provide enough constraints for accurate single view display estimation.

In summary, it can be said that most existing methods use multiple views, while those using a single view utilize active illumination and tailor made hardware. The financial possibilities of selling such a technology are acknowledged. In contrast, this paper introduces a passive method based on a single view that might not be as accurate as the existing active, multiple view methods, but is readily available to any person with a camera and a computer.

3. Approximate eye model

This section provides a short introduction to the human eye and proposes an approximate geometric model. The iris (see Fig. 1) has a circular aperture known as the pupil, which regulates the amount of light coming into our eyes. Pupil and iris are surrounded by the white sclera. A protective transparent membrane called cornea covers the iris. The boundary between the cornea and the sclera is called the corneal limbus, which has been found to be close to circular [19]. The external surface of the cornea is very smooth and has a thin film of tear fluid on it, which makes

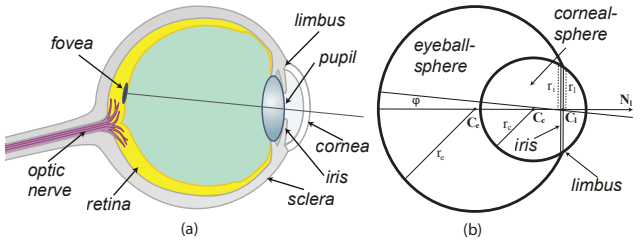

Figure 2. (a) A vertical cross section showing the anatomy of the eye. (b) The eye is modeled by segments of two quasi spheres.

it quite reflective. A simplified anatomy of a human eye is illustrated in a vertical cross section in Fig. 2(a). The optical axis (line-of-gaze) is defined as the line connecting the pupil center with the center of the cornea. The visual axis (line-of-sight) is the line connecting the corneal center with the center of the fovea.

To interpret an image of an eye, a geometric model is required. Although the eye is of an approximate spherical shape, it cannot just be modeled by a single sphere [16]. Instead we model it by segments of two quasi spheres (Fig. 2(b)), namely the eyeball sphere with radius r_{e} and the corneal sphere with radius r_{c}. The circular iris has a radius of approximately r_{i} and its visible part is the circular limbus with a slightly smaller radius r_{l}. It has been found that the difference in eye parameter values among adults is small [19] and it is reasonable to assume that various parameters of an adult human eye are close to certain *anatomical constants* (see Table 1). The angle φ between the optical axis and visual axis is subject dependent, but it is relative small ($\sim 5^\circ$ [13]). The optical axis is assumed to be parallel to the supporting plane normal $\mathbf{N}_{\mathrm{l}}$ of the limbus. Let $\mathbf{C}_{\mathrm{l}}$ and $\mathbf{C}_{\mathrm{c}}$ be the center of the limbus and corneal sphere respectively. The distance $|\mathbf{C}_{\mathrm{l}}\mathbf{C}_{\mathrm{c}}| = \sqrt{r_{\mathrm{c}}^2 - r_{\mathrm{l}}^2}$ is used to obtain the corneal sphere center

$$\mathbf{C}_{\mathrm{c}} = \mathbf{C}_{\mathrm{l}} - |\mathbf{C}_{\mathrm{l}}\mathbf{C}_{\mathrm{c}}|\mathbf{N}_{\mathrm{l}}. \tag{1}$$

4. Limbus reconstruction

In the following, it will be shown that the limbus can be reconstructed (up to a sign ambiguity) from its perspective projection. Sect. 4.2 provides an analysis of the proposed method to imperfect data.

4.1. Closed form solution

Let the camera calibration matrix $\mathbf{K}$ and the radius r_{l} be known. The circular limbus can be represented by a 3×3 symmetric matrix $\mathbf{L}$ given by

$$\mathbf{L} = \begin{bmatrix} \mathbf{I}_2 & -\mathbf{C}_{\mathrm{l}} \\ -\mathbf{C}_{\mathrm{l}}^{\mathrm{T}} & (\mathbf{C}_{\mathrm{l}}^{\mathrm{T}}\mathbf{C}_{\mathrm{l}} - r_{\mathrm{l}}^2) \end{bmatrix}. \tag{2}$$

Parameter	Mean	Reference
r_{c}	7.77mm	[13]
r_{l}	5.55mm	[14],[19]

Table 1. Parameters of the geometric eye model.

Any 2D point X on the limbus plane, with homogeneous representation $\tilde{\mathbf{X}}$, lying on the limbus will satisfy the equation

$$\tilde{\mathbf{X}}^{\mathrm{T}}\mathbf{L}\tilde{\mathbf{X}} = \mathbf{0}. \tag{3}$$

Let the rotation matrix $\mathbf{R}$ and the translation vector $\mathbf{T}$ define the coordinate transformation from the limbus plane to the camera. Without loss of generality, let the limbus be centered at the origin $\mathbf{C}_{\mathrm{l}} = \mathbf{0}$. A planar homography that maps points on the limbus plane to points on the image plane can be defined as

$$\tilde{\mathbf{x}} = \mathbf{H}\tilde{\mathbf{X}} = \mathbf{K}[\ \mathbf{R}_1 \quad \mathbf{R}_2 \quad \mathbf{T}\ ]\tilde{\mathbf{X}}, \tag{4}$$

where $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{X}}$ are the homogeneous representations of points on the image plane and on the limbus plane respectively, and $\mathbf{R}_{\mathrm{i}}$ is the i-th column of $\mathbf{R}$ [8].

It follows from (3), (4) and the equation $\tilde{\mathbf{x}}^{\mathrm{T}}\mathbf{L}_{\mathrm{img}}\tilde{\mathbf{x}} = 0$, where $\mathbf{L}_{\mathrm{img}}$ is the image of the limbus, that $\mathbf{L}_{\mathrm{img}}$ will be a conic given by

$$\mathbf{L}_{\mathrm{img}} = \mathbf{H}^{-\mathrm{T}}\mathbf{L}\mathbf{H}^{-1}. \tag{5}$$

Let us first consider a simplified case where both the camera calibration matrix and the rotation matrix are given by the identity matrix $\mathbf{I}_3$. Under this configuration, the image of the limbus can be obtained using (5) and is given by

$$\mathbf{L'}_{\mathrm{img}} = \begin{bmatrix} 1 & 0 & -\frac{t_1}{t_3} \\ 0 & 1 & -\frac{t_2}{t_3} \\ -\frac{t_1}{t_3} & -\frac{t_2}{t_3} & \frac{t_1^2+t_2^2-r_{\mathrm{l}}^2}{t_3^2} \end{bmatrix}. \tag{6}$$

Note that $\mathbf{L'}_{\mathrm{img}}$ represents a circle centered at $\mathbf{c}' = [\ \frac{t_1}{t_3} \quad \frac{t_2}{t_3}\ ]^{\mathrm{T}}$ with radius $r' = \frac{r_{\mathrm{l}}}{t_3}$. The translation vector can be recovered in terms of r' and $\mathbf{c}'$ as

$$\mathbf{T}' = \frac{r_{\mathrm{l}}}{r'}\begin{bmatrix} \mathbf{c}' \\ 1 \end{bmatrix}. \tag{7}$$

Consider now the general case where the rotation matrix and the camera calibration matrix are given by $\mathbf{R}$ and $\mathbf{K}$ respectively. The effect of $\mathbf{K}$ is first removed by normalizing the image using $\mathbf{K}^{-1}$. The conic $\mathbf{L}_{\mathrm{img}}$ will be transformed to a conic $\hat{\mathbf{L}}_{\mathrm{img}} = \mathbf{K}^{\mathrm{T}}\mathbf{L}_{\mathrm{img}}\mathbf{K}$ in the normalized image, which can be diagonalized into

$$\hat{\mathbf{L}}_{\mathrm{img}} = \mathbf{M}\mathbf{D}\mathbf{M}^{\mathrm{T}} = \mathbf{M}\begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & d \end{bmatrix}\mathbf{M}^{\mathrm{T}}, \tag{8}$$

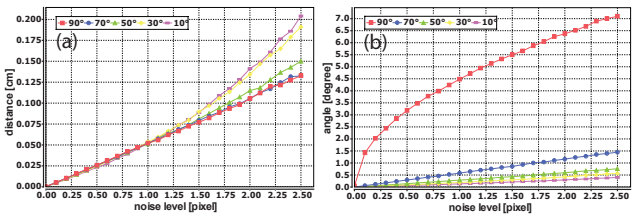

Figure 3. Sensitivity of the limbus reconstruction to imperfect data. RMS error in a) the distance of the limbus center and b) the angle of the limbus normal against the noise level.

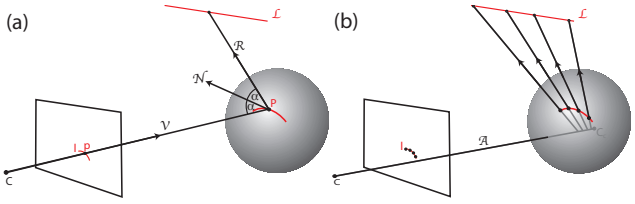

Figure 4. A line $\mathcal{L}$ reflected on a sphere and imaged as l by a camera with center $\mathbf{C}$. (a) Angles of incidence and reflection obey the law of reflection. (b) Reflection lines will intersect the two lines $\mathcal{L}$ and $\mathcal{A}$.

where $\mathbf{M}$ is an orthogonal matrix whose columns are the eigenvectors of $\hat{\mathbf{L}}_{\text{img}}$, and $\mathbf{D}$ is a diagonal matrix consisting of the corresponding eigenvalues. The matrix $\mathbf{M}^{\text{T}}$ defines a rotation that will transform $\hat{\mathbf{L}}_{\text{img}}$ to the ellipse $\mathbf{D}$ which is centered at the origin and whose principal axes are align with the coordinate system.

In order to relate the general case with the previous simplified case, the generalized ellipse $\mathbf{D}$ is transformed to a circle $\mathbf{E}$ by applying the transformation $\mathbf{E} = \mathbf{N}^{\text{T}}\mathbf{D}\mathbf{N}$. WLOG, let $ab > 0$, $ad < 0$ and $|a| > |b|$. The circle $\mathbf{E}$ has the center $\mathbf{c} = \begin{bmatrix} -s_2\frac{\sqrt{(a-b)(b-d)}\cos\alpha}{b} & -s_1 s_2\frac{\sqrt{(a-b)(b-d)}\sin\alpha}{b} \end{bmatrix}^{\text{T}}$ and radius $r = s_3\frac{\sqrt{-ad}}{b}$, where $g = \sqrt{\frac{b-d}{a-d}}$, $h = \sqrt{\frac{a-b}{a-d}}$ and $s_{\text{i}=1\ldots3}$ are undetermined signs. (See [3] for a derivation.)

Note that α represents a rotation angle around the limbus local z-axis. It can be chosen arbitrarily, because the projection of the circular limbus will not be affected by a rotation around its normal. The two unknown signs s_1 and s_3 can be recovered by enforcing the constrains that (a) $\mathbf{C}_{\text{l}}$ is in front of the camera and (b) $\mathbf{N}_{\text{l}}$ faces the camera. The translation vector $\mathbf{T}$ can be recovered from $\mathbf{c}$ and r using (7), and the rotation matrix can be obtained using $\mathbf{R} = \mathbf{M}\mathbf{N}$. We finally recover the center of the limbus and the normal of its supporting plane in the camera coordinate system as

$$\overline{\mathbf{C}}_{\text{l}} = \mathbf{T} \qquad \overline{\mathbf{N}}_{\text{l}} = \mathbf{R}\begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^{\text{T}}. \qquad (9)$$

4.2. Noise analysis

It is difficult to obtain the limbus contour because the iris gradually dissolves into the sclera. The performance of the algorithm is investigated with controlled synthetic data. The experimental setup consists of a circle with radius r_{l} representing the limbus and being viewed by a camera from different viewing directions. The image is obtained analytically using (5). Gaussian noise of different standard deviations is added to the conic where points are sampled and perturbed in a radial direction from the center. A noisy conic is then obtained as a conic fitted to these noisy points using a least squares method [5].

Experiments with noise levels ranging from 0.0 to 2.5 pixels are carried out for distinct viewing directions (optical axis to limbus plane: 90° to 10° in 20° intervals). For each noise level, 200 independent trials are conducted to estimate a) the center and b) the limbus normal. The unknown sign s_2 is determined by the ground truth. Fig. 3 shows a plot of the RMS error in a) the distance of the limbus center and b) the angle of the limbus normal against the noise level. Larger noise levels result in significant increase in the error. The 90°-case should be avoided as it results in an unstable limbus orientation.

The limbus has been reconstructed and it is now possible to determine the corneal sphere center from (1). Sect. 5 will estimate a single display edge from its reflection on the cornea and Sect. 6 will extract a rectangle from four independently estimated display edges.

5. Display edge reconstruction

In this section a closed form solution for a geometric display edge estimation is proposed. A 2D edge l of a display reflection on a single cornea can determine its corresponding 3D edge $\mathcal{L}$. All four 3D display edges are first independently estimated and a subsequent optimization (see Sect. 6) will extract a planar rectangle representing the display.

Let us denote the back-projection of a point $\mathbf{p} \in l$ as the viewing line $\mathcal{V}$ and its reflection on the conreal sphere as the reflection line $\mathcal{R}$. The viewing line will leave the camera center $\mathbf{C}$, pass through the point $\mathbf{p}$ and intersect the corneal sphere at a point $\mathbf{P}$ (see Fig. 4(a)). The corneal sphere center can be determined from (9) and (1). Given the radius r_{c} as an anatomical constant, the point $\mathbf{P}$ can be determined. Let $\mathbf{V}_{\text{p}}$ and $\mathbf{R}_{\text{p}}$ be the unit vectors in the directions of the viewing line and the reflection line respectively. The law of reflection states that the incident angle must be equal to the reflection angle, and the reflection direction is therefore given by $\mathbf{R}_{\text{p}} = (2\mathbf{N}_{\text{p}} \cdot \mathbf{V}_{\text{p}})\mathbf{N}_{\text{p}} - \mathbf{V}_{\text{p}}$, where $\mathbf{N}_{\text{p}}$ is the unit normal vector at point $\mathbf{P}$. The reflection line $\mathcal{R}$ passes through $\mathbf{P}$ in the direction $\mathbf{R}_{\text{p}}$ and will, by construction, intersect the line $\mathcal{L}$ at some point. Note that all the reflection lines constructed in such a way will intersect the line $\mathcal{L}$ and also another line $\mathcal{A}$ from the camera center $\mathbf{C}$ to the corneal

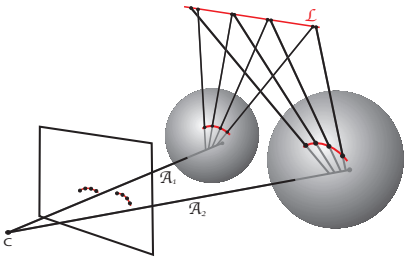

Figure 5. The line $\mathcal{L}$ can be determined as the intersection of reflection lines from two spheres.

sphere center $\mathbf{C}_c$ (see Fig. 4(b)).

The following section formulates line incidence computation through the use of Plücker line coordinatization and Section 5.2 considers the reflection on two eyes.

5.1. Line incidence in Plücker space

In order to formulate line intersections algebraically, we adopt the 6-vector Plücker line coordinates representation for directed lines in P^3[8]. Two points $\mathbf{P} = (p_x, p_y, p_z, 1)$ and $\mathbf{Q} = (q_x, q_y, q_z, 1)$ define a line $\mathcal{L}$ as

$$\mathcal{L} = \begin{bmatrix} l_0 \\ l_1 \\ l_2 \\ l_3 \\ l_4 \\ l_5 \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} p_x q_y - q_x p_y \\ p_x q_z - q_x p_z \\ p_x - q_x \\ p_y q_z - q_y p_z \\ p_z - q_z \\ q_y - p_y \end{bmatrix}^{\mathrm{T}}. \quad (10)$$

With this notation, lines in P^3 are mapped to homogeneous points $\mathcal{L} = (l_0, l_1, l_2, l_3, l_4, l_5)$ or hyperplanes $\overline{\mathcal{L}} = (l_4, l_5, l_3, l_2, l_0, l_1)$ in 5 dimensional Plücker coordinate space.

A major advantage of this representation is the simplicity of the incidence operation. Given two lines $\mathcal{A}$ and $\mathcal{B}$, the incidence operation is the inner product between the homogeneous Plücker representation of line $\mathcal{A}$ and the hyperplane Plücker representation of line $\mathcal{B}$

$$\mathcal{A} \cdot \overline{\mathcal{B}} = a_0 b_4 + a_1 b_5 + a_2 b_3 + a_3 b_2 + a_4 b_0 + a_5 b_1. \quad (11)$$

Since the inner product will be zero for intersecting lines, solving for n lines $\mathcal{I}_1, \mathcal{I}_2, ..., \mathcal{I}_n$ that intersect m given lines $\mathcal{L}_1, \mathcal{L}_2, ..., \mathcal{L}_m$ is equivalent to finding the n-dimensional nullspace of a matrix formed by the Plücker hyperplane representations of the given lines:

$$\mathbf{S}\mathcal{I} = \begin{bmatrix} \overline{\mathcal{L}}_1 \\ \vdots \\ \overline{\mathcal{L}}_m \end{bmatrix} \mathcal{I} = \mathbf{0}. \quad (12)$$

Finding the set of lines $\mathcal{I}$ that map $\mathbf{S}$ to a null vector, implies that for each row i the inner product $\overline{\mathcal{L}}_i \cdot \mathcal{I}$ equals zero. Given

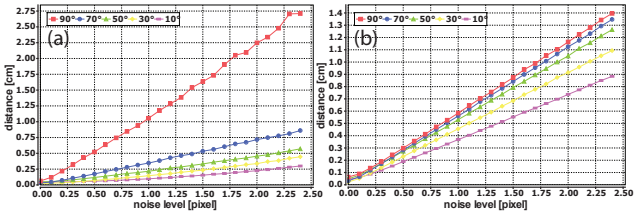

Figure 6. Sensitivity of the display reconstruction to imperfect data. In two experiments, noise is added to (a) the limbic boundary and to (b) the display edge.

the m reflection lines from the previous section and the task of finding the $n = 2$ intersecting lines, we can simply solve for those lines by finding the nullspace of the matrix $\mathbf{S}$ with singular value decomposition $\mathbf{S} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{T}}$, where

$$\mathbf{U} = \begin{bmatrix} u_{11} & \cdots & u_{1\mathrm{m}} \\ \vdots & & \vdots \\ u_{\mathrm{m}1} & \cdots & u_{\mathrm{mm}} \end{bmatrix}, \mathbf{V} = \begin{bmatrix} v_{11} & \cdots & v_{16} \\ \vdots & & \vdots \\ v_{61} & \cdots & v_{66} \end{bmatrix},$$

and

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_6 \\ 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{bmatrix}. \quad (13)$$

For $n = 2$, $\mathbf{S}$ is a rank four matrix, whose nullspace will span a two dimensional subspace that can be parameterized by the two points $\mathbf{d} = (v_{15}, \cdots, v_{65})$ and $\mathbf{e} = (v_{16}, \cdots, v_{66})$, which correspond to the two smallest singular values σ_5 and σ_6. Fortunately not all points on the 5-dimensional line $\mathcal{L}(t) = \mathbf{d}t + \mathbf{e}$ are 3-dimensional lines, but just those lines $\mathcal{B}$ that satisfy

$$\mathcal{B} \cdot \overline{\mathcal{B}} = 0. \quad (14)$$

[21] was the first to formulate and solve this problem by intersecting the line $\mathcal{L}(t)$ with all points that satisfy (14). This produces the quadratic equation $(\mathbf{d} \cdot \overline{\mathbf{d}})t^2 + 2(\mathbf{d} \cdot \overline{\mathbf{e}})t + (\mathbf{e} \cdot \overline{\mathbf{e}}) = 0$, for which the two real roots

$$t_{\pm} = \frac{-(\mathbf{d} \cdot \mathbf{e}') \pm \sqrt{(\mathbf{d} \cdot \mathbf{e}')^2 - (\mathbf{d} \cdot \mathbf{e}')(\mathbf{e} \cdot \mathbf{e}')}}{\mathbf{d} \cdot \mathbf{d}'} \quad (15)$$

correspond to the two intersecting lines.

5.2. Estimation from two eyes

For the reflection on a single corneal sphere, the two intersecting lines ($\mathcal{A}$ and $\mathcal{L}$ in Fig. 4 (b)) can be determined from (15). (12) can simply be extended to include the reflection lines from another corneal sphere. If this is done,

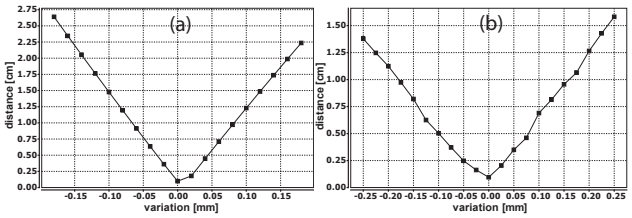

Figure 7. Errors from variations in *anatomical constants*. The error in the average distance between estimated reflection rays and ground truth display edge is estimated and plotted against (a) variations in the limbus radius and (b) variations in the corneal sphere radius.

the line from the camera center to the sphere center ($\mathcal{A}_1$ and $\mathcal{A}_2$ in Fig. 5) will not intersect the reflection lines any more. Instead there will just be a single line intersecting the reflection lines from both spheres ($\mathcal{L}$ in Fig. 5).

5.3. Noise analysis

Similar to the noise analysis in Sect. 4.2, this section analyzes the sensitivity of a single reconstructed display edge to imperfect data. In two experiments, noise is added to (a) the limbic boundary and to (b) the reflected display edge. A program was written to simulate the reflection of a display edge on a sphere and a Bezier curve is automatically fitted to the edges of this synthetic corneal reflection. To add noise to the reflection of the display edge, the Bezier curve is uniformly sampled and individual samples are perturbed in the x and y directions. Noise is first added to the limbic boundary and Fig. 6(a) shows a plot of the RMS error in the average distance between estimated reflection lines and display edge against the noise level. In a second experiment the ground truth limbic boundary is used and noise is just added to the reflected display edge. Fig. 6(b) shows a plot for this experiment. It can be seen that the error increases linearly with the noise level. Increasing the level of noise will result in a relative larger error, because the corneal sphere is estimated from the limbic boundary and small errors in the center of the corneal sphere will result in larger errors in the directions of the reflection lines. This will result in a larger average distance between the reflection lines and ground truth edge.

5.4. Variations in anatomical parameter

The geometric model proposed in Sect. 3 is an approximation and individual parameter could be different from Table 1. This section analysis the error produced by variations from these parameters. The corneal sphere radius of 220 human eyes has been measured with a mean of $7.77mm$. All measurements were within $\pm 0.25mm$ of the mean [13]. There is no data available for variations of the limbus radius and we assume a similar variation.

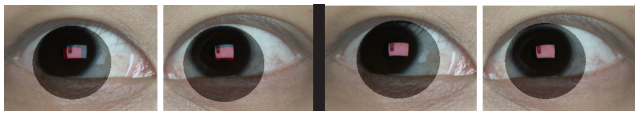

Figure 8. Reflection of reconstructed display (red) on corneal sphere before (left) and after (right) optimization.

Based on this data we have prepared synthetic experiments which estimate the error in the average distance between reflection rays and ground truth display edge against the corneal sphere radius (Fig. 7(a)) and against the limbus radius (Fig. 7(b)). It can be seen that there is a linear relationship between variations in anatomical parameter and quality of the reconstruction. Variations in the limbus radius have a greater impact than variations in the corneal sphere radius.

6. Display reconstruction

Display edges are independently estimated as four lines $\{\mathcal{L}_1, \mathcal{L}_2, \mathcal{L}_3, \mathcal{L}_4\}$ with the closed form solution from Sect. 5. A rectangle is subsequently extracted from the four lines. In general the four lines will not intersect, nor form a perfect rectangle, because:

- The smooth transition from iris to sclera can cause an inaccurate limbus boundary. It has been shown that a noisy limbus boundary will result in large errors (see Sect. 5.3).
- The geometric model in Sect. 3 is an approximation and individual parameter could be different from the a*natomical constants*. It has been shown that slight variations for the anatomical parameters will result in large errors (see Sect. 5.4).

Limbus boundaries and anatomical parameters are adjusted through an optimization. In a simple rectangle with sides $\{\mathcal{L}_1, \mathcal{L}_2, \mathcal{L}_3, \mathcal{L}_4\}$, opposite sides $\{\mathcal{L}_1, \mathcal{L}_3\}$ and $\{\mathcal{L}_2, \mathcal{L}_4\}$ are parallel and let $\angle(\overrightarrow{\mathcal{L}_1}, \overrightarrow{\mathcal{L}_3}) = \angle(\overrightarrow{\mathcal{L}_2}, \overrightarrow{\mathcal{L}_4}) = 0$, where $\overrightarrow{\mathcal{L}}$ is a unit 3D vector of the line $\mathcal{L}$ and $\angle$ represents the positive angle between the two vector. The energy

$$\mathrm{E}_1 \quad = \quad \angle(\overrightarrow{\mathcal{L}_1}, \overrightarrow{\mathcal{L}_3}) + \angle(\overrightarrow{\mathcal{L}_2}, \overrightarrow{\mathcal{L}_4}) \tag{16}$$

will be minimal for a rectangle. Similarly all angles in a rectangle are 90 degree and the function

$$\begin{aligned} \mathrm{E}_2 \quad = \quad & |\angle(\overrightarrow{\mathcal{L}_1}, \overrightarrow{\mathcal{L}_2}) - 90| + |\angle(\overrightarrow{\mathcal{L}_2}, \overrightarrow{\mathcal{L}_3}) - 90| + \\ & |\angle(\overrightarrow{\mathcal{L}_3}, \overrightarrow{\mathcal{L}_4}) - 90| + |\angle(\overrightarrow{\mathcal{L}_4}, \overrightarrow{\mathcal{L}_1}) - 90| \end{aligned} \tag{17}$$

will be minimal. We seek the minimum of the function

$$\mathrm{E}(r_\mathrm{c}, r_\mathrm{l}, \mathbf{L}^1_\mathrm{img}, \mathbf{L}^2_\mathrm{img}) = E_1 + \omega E_2 \tag{18}$$

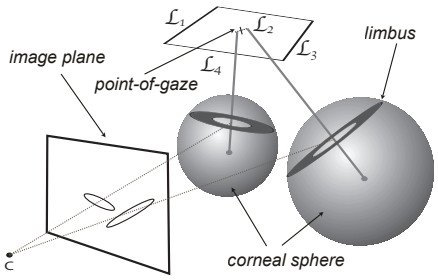

Figure 9. Point-of-gaze is approximated by intersecting the optical axes of both eyes with the display.

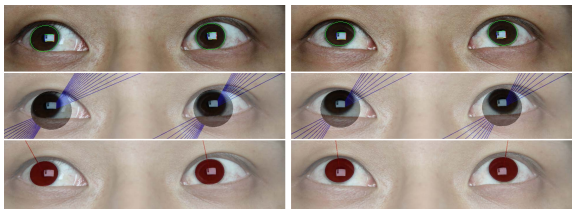

Figure 10. Results overlaid on the image for two distinct point-of-gaze. From first to last row: (1) Input images with segmented limbus and display reflection. (2) Reflection lines for one of the screen edges. (3) Estimated optical axis.

for the corneal sphere radius r_c, the limbus radius r_l and the two limbic projections $\mathbf{L}^1_{\text{img}}$ and $\mathbf{L}^2_{\text{img}}$.

The two limbic projections are initialized with our implementation of a limbus segmentation algorithm [12] and parameterized by their foci points and their major diameter. The corneal sphere and limbus radii are initialized with the values shown in Table 1. Parameters of the ellipse are bounded by ± 1.5 pixel and the radii of corneal sphere and limbus are within ± 0.25 mm and ± 0.2 mm respectively. The optimization is not directly performed on the search space. Instead an initial minimum is found through an exhaustive (but coarse) sampling of the search space. The ellipse parameter are sampled with 0.25 pixel and the anatomical parameter with 0.05 mm. The final optimization is initialized with the initial minimum from this subdivision. This simple procedure avoids an early termination of the optimization in a local minimum. Unfortunately it is time consuming and does not guarantee the global minimum. The reflection of an estimated display on the corneal sphere is shown with and without optimization in Fig. 8.

7. Gaze estimation

The intersection of the optical axes of both eyes with the display can be computed. In general there will be no single intersection, because the optical axis is distinct from the visual axis (see Sect. 3). A single point-of-gaze is approximated as the point on the display that minimizes the sum of the distances to the two lines (see Fig. 9).

To disambiguate and solve for the unknown sign s_2 (from Sect. 4) we reconstruct both gaze directions for each eye and verify if the gaze of both eyes intersect the display.

8. Experimental results

Experiments on different subjects were carried out and results are presented in this section. Given the sensitivity of the approach to noisy data, automatic robust iris and display reflection segmentation is very important. There exists extensive literature on iris segmentation and some methods have been quite successful [12]. Although our implementation of the method proposed by Li et al. has been successful in most cases, a manual segmentation was performed for some of the images. More experiments and further research need to be done to improve existing limbus boundary extraction methods. Specular reflections of the display were extracted with a simple thresholding method and subsequent Bezier curve fitting. To prevent occlusions from the eyelid and nose reflections on the cornea, we place the camera below the display ($\sim 45^\circ$ between display normal and cameras principal axis). We used a 70mm lens and focused manually so that the subjects eyes are sharp at a distance of around 45cm to the display. Results for two distinct viewing directions of a single subject are shown in Fig. 10. In a second experiment, four subjects were asked to gaze at nine points in a small 120mm-120mm area on a computer display. The results of the experiments are shown in Fig. 11. Average errors for the subjects were around 10 per cent. Please note that all experimental results in the paper have been obtained with a single image per gaze direction.

9. Conclusions

This paper introduces a method for reconstructing eyes and display from reflections on the cornea and investigates the minimal information required for gaze estimation. It is shown that iris boundaries and display reflections on a single intrinsically calibrated image provide enough information for this. Quantitative results with real data show that it is possible to determine approximately where a user is focusing his attention relative to a computer screen from just a single image. As expected, the accuracy of the method is worse compared to commercial systems which use multiple images and active illumination. Gaze estimation has potential direct and indirect implications on various fields, including computer science (HCI), psychology (perception), industrial engineering (usability) and advertising. We believe that these areas will still benefit from a system with reduced accuracy. Please note that all experiments were performed using just a single image for the estimation of gaze, display and eyes and we found the main challenge is the accurate estimation of the display.

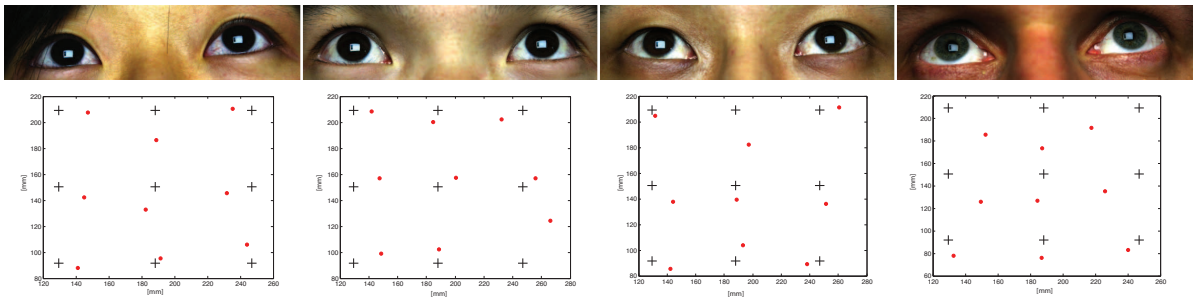

Figure 11. Point-of-gaze estimation results. Four subjects were asked to gaze at 9 points on a display. First row shows one of nine input image used for each subject. The second row plots the point-of-gaze estimation as red with the ground truth shown as a cross. The display was reconstructed independently for each of the 36 images and the point-of-gaze was estimated. Average errors were 15.14mm, 16.90mm, 14.40mm and 25.13mm respectively.

Eye gaze estimation has a history of more than 100 years, yet it is still not widely used because the technology lacks usability requirements that hinder its applicability. One of the main challenges are the calibration requirements for gaze estimation. Minimal setup and calibration requirements of our method enable this technology for everyone and not just for those who buy an commercial gaze tracker and are willing to run through a tedious system calibration process.

In future work, we would like to study a better, more complex approximation of the cornea. In addition, the iris segmentation should be improved to be robust, accurate and able to deal with situations of illumination changes, varying focus and contact lenses.

References

[1] National cancer institutes. http://www.cancer.gov, 2009.

[2] J. Chen and Q. Ji. 3d gaze estimation with a single camera without ir illumination. In *ICPR*, 2008.

[3] Q. Chen, H. Wu, and T. Wada. Camera calibration with two arbitrary coplanar circles. In *ECCV*, pages 521–532, 2004.

[4] A. T. Duchowski. *Eye Tracking Methodology: Theory and Practice*. Springer-Verlag New York, Inc., 2007.

[5] A. Fitzgibbon, M. Pilu, and R. Fisher. Direct least square fitting of ellipses. *PAMI*, 21(5):476–480, 1999.

[6] Y. Francken, C. Hermans, and P. Bekaert. Screen-camera calibration using gray codes. In *Canadian Conference on Computer and Robot Vision*, pages 155–161, 2009.

[7] E. D. Guestrin and M. Eizenman. Remote point-of-gaze estimation requiring a single-point calibration for applications with infants. In *Eye Tracking Research & Applications*, pages 267–274, 2008.

[8] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2004.

[9] J. J. Kang, M. Eizenman, E. D. Guestrin, and E. Eizenman. Investigation of the cross-ratios method for point-of-gaze estimation. *Transactions on Biomedical Engineering*, 55(9):2293–2302, 2008.

[10] S. Kohlbecher, S. Bardinst, K. Bartl, E. Schneider, T. Poitschke, and M. Ablassmeier. Calibration-free eye tracking by reconstruction of the pupil ellipse in 3d space. In *Eye Tracking Research & Applications*, pages 135–138, 2008.

[11] P. Lagger, M. Salzmann, and P. F. V. Lepetit. 3d pose refinement from reflections. In *CVPR*, pages 1–8, 2008.

[12] P. Li, X. Liu, L. Xiao, and Q. Song. Robust and accurate iris segmentation in very noisy iris images. *Image and Vision Computing*, 28(2):246–253, May 2009.

[13] H.-L. Liou and N. A. Brennan. Anatomically accurate, finite model eye for optical modeling. *Journal of the Optical Society of America*, 14(8):1684–1695, 1997.

[14] K. Nishino and S. K. Nayar. Corneal imaging system: Environment from eyes. *IJCV*, 70(1):23–40, 2006.

[15] C. Nitschke, A. Nakazawa, and H. Takemura. Display-camera calibration from eye reflections. In *ICCV*, 2009.

[16] C. Oyster. *The Human Eye: Structure and Function*. Sinauer Associates, 1999.

[17] D. Schnieders, K.-Y. K. Wong, and Z. Dai. Polygonal light source estimation. In *ACCV*, 2009.

[18] S.-W. Shih, Y.-T. Wu, and J. Liu. A calibration-free gaze tracking technique. In *ICPR*, page 4201, 2000.

[19] R. Snell and M. A. Lemp. *Clinical Anatomy of the Eye*. Wiley-Blackwell, 1997.

[20] SR Research Ltd., 5516 Osgoode Main St. Ottawa, Ontario Canada. *EyeLink 1000 Technical specifications*, 2009.

[21] S. Teller and M. Hohmeyer. Determining the lines through four lines. *Journal of Graphics Tools*, 4(3):11–22, 1999.

[22] J.-G. Wang, E. Sung, and R. Venkateswarlu. Estimating the eye gaze from one eye. *CVIU*, 98(1):83–103, 2005.

[23] D. H. Yoo and M. J. Chung. A novel non-intrusive eye gaze estimation using cross-ratio under large head motion. *CVIU*, 98(1):25–51, 2005.