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

Craig Hennessey
University of British Columbia

Borna Noureddin
University of British Columbia

Peter Lawrence
University of British Columbia

Abstract

Eye-gaze as a form of human machine interface holds great promise for improving the way we interact with machines. Eye-gaze tracking devices that are non-contact, non-restrictive, accurate and easy to use will increase the appeal for including eye-gaze information in future applications. The system we have developed and which we describe in this paper achieves these goals using a single high resolution camera with a fixed field of view. The single camera system has no moving parts which results in rapid reacquisition of the eye after loss of tracking. Free head motion is achieved using multiple glints and 3D modeling techniques. Accuracies of under 1° of visual angle are achieved over a field of view of 14x12x20 cm and over various hardware configurations, camera resolutions and frame rates.

Keywords: Single camera, free head motion, eye-gaze tracking, eye model, fast reacquisition, human computer interface, human machine interface

1. Introduction

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

Eye-gaze information has proven useful in a diverse number of applications such as psychological studies [Rayner 1998], usability studies in driving and aviation [Petersson et al. 2004; Hanson 2004], and analysis of layout effectiveness in advertising [Lohse 1997]. In particular it is well suited to human computer interfaces for mouse augmentation and control [Zhai et al 1999] and eye typing for the physically disabled [Majaranta and Raiha 2002].

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

ETRA 2006, San Diego, California, 27–29 March 2006.
© 2006 ACM 1-59593-305-0/06/0003 $5.00

The system described in this paper meets these key requirements as follows. A single high resolution camera with a fixed field of view is used which does not make any contact with the user. A 3D method based on multiple glints is used to allow free head motion within the field of view of the camera. The high resolution images permit a larger field of view while still possessing accurate image features, resulting in accurate eye gaze estimation. Using a single camera with no moving parts simplifies the system geometry and calibration and leads to short reacquisition times. These advantages make the system easy to set up and simple to use.

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

2. Related Works

There have been many methods developed for tracking the eye-gaze of a subject including sensors attached to the face and eye, restrictive video systems requiring a fixed head location, head mounted video systems, and non-contact and non-restrictive video based systems. We feel that the non-contact, non-restrictive video based methods hold the greatest promise for a widely acceptable eye-gaze tracking interface and as such will focus on research in this area. For an overview of alternative methods for eye-gaze tracking see the review by Young and Sheena [1975], and more recently Morimoto and Mimica [2005].

Video based systems require high resolution images of the eye to accurately estimate the point of gaze (POG). Ohno et al. developed a single camera system which achieved accuracies of under 1° of visual angle [Ohno et al. 2002]. The system verified the ability of their methods to determine the POG, however it had a relatively small field of view of 4x4 cm at 60 cm distance. Morimoto et al. proposed a single camera method for estimating the POG which achieved an average accuracy of 2.5° in simulations [Morimoto et al. 2002]. To date there is no reported system implementation based on this proposal. Shih and Liu proposed a method which used only a single camera [Shih and Liu 2004]. The system they implemented utilized two stereo cameras however, which were required to provide additional constraints for their algorithms. The fixed field of view was restricted to 4x4 cm. Their system operated at 30 Hz with an accuracy of approximately 1° of visual angle.

The main difficulty with the above fixed single camera systems is the limited field of view required to capture sufficiently high resolution images.

To allow for free head motion a large field of view is required. Many systems utilize multiple cameras to achieve these goals, with wide angle (WA) lens cameras used to direct a movable narrow angle (NA) lens camera. Yoo and Chung developed a free

head system which utilizes a WA camera to direct a NA camera mounted on a pan-tilt mechanism [Yoo and Chung 2005]. Their system operates at 15 Hz and achieves an accuracy of 0.98° of visual angle in the horizontal direction and 0.82° in the vertical direction. Noureddin et al. developed a two camera system where the fixed WA camera uses a rotating mirror to direct the orientation of the NA camera [Noureddin et al. 2005]. The rotating mirror can achieve faster slew rates when compared with pan-tilt mechanisms. Their system operates at 9 Hz with an accuracy of 2.9°. The latest reported system by Ohno and Mukawa utilizes 3 cameras, two fixed stereo WA cameras and a NA camera mounted on a pan-tilt mechanism [Ohno and Mukawa 2004]. Their system uses two computers and achieves an accuracy of about 1° of visual angle while operating at 30 Hz. Beymer and Flickner developed a 4 camera system which uses 2 stereo WA cameras and 2 stereo NA cameras [Beymer and Flickner 2003]. The WA cameras direct galvanometer motors to orient the NA cameras. The calibration task is considerable due to the multiple stereo cameras and the variable focal lengths of the NA cameras. Their system operates at 10 Hz and has a reported accuracy of 0.6°.

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

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

3. Methods

The methods we have developed for estimating the POG are based on 3D models of the camera, system and eye. The camera is modeled using the pin-hole camera model and the eye is modeled using a simplified version of the Gullstrand schematic eye [Goss and West 2002]. Population averages compiled by Gullstrand are used for the model parameters of interest. Subject deviations from the population averages are compensated by a one-time per user calibration.

Shown in Figure 1 is an example of the simplified eye model with the parameters of interest, r, r_d, and n, and the points $\boldsymbol{p_c}$ and $\boldsymbol{c}$, which are required to compute the optical axis (OA) vector $\boldsymbol{L}$. The OA is defined as the vector from the center of the cornea $\boldsymbol{c}$ to the center of the pupil $\boldsymbol{p_c}$. The OA is different from the line of sight (LOS) which is the vector that traces from the fovea (high acuity portion of the retina) through the center of the pupil and ultimately to the real POG. The location of the fovea varies from person to person, and can be located up to 5° from the OA [Goss and West 2002]. The offset between the estimated POG and the real POG due to the difference between the OA and the LOS is fixed for each user and is compensated for by the calibration technique described in Section 3.4.

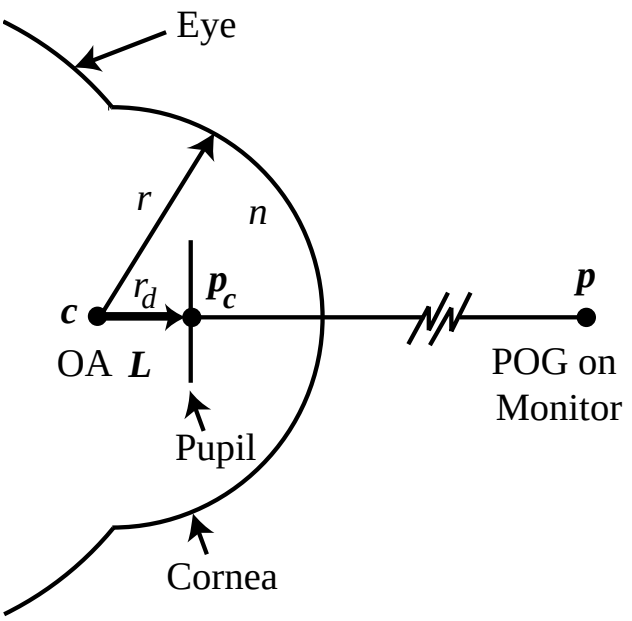


Figure 1 – Eye model used to calculate the POG. The parameters of interest taken from population averages are: radius of the cornea, r, distance from the center of the cornea to the center of the pupil r_d and the index of refraction of the aqueous humor, n. The center of the cornea is located at point $\boldsymbol{c}$ and the center of the pupil is located at point $\boldsymbol{p_c}$. The OA $\boldsymbol{L}$ is the vector formed from $\boldsymbol{c}$ to $\boldsymbol{p_c}$, and the POG $\boldsymbol{p}$ is the intersection of the OA with the monitor plane.

The following outline provides an overview of the steps required to determine the POG $\boldsymbol{p}$:

1. The POG is found by intersecting the OA vector $\boldsymbol{L}$ with the monitor plane (where $\boldsymbol{L} = \boldsymbol{p_c} - \boldsymbol{c}$).
2. To determine the cornea center $\boldsymbol{c}$, the eye model is used along with the image locations of two glints off the surface of the cornea. Using multiple glints provides a method for triangulating the 3D cornea center.
3. The pupil center $\boldsymbol{p_c}$ is determined by using the eye model, the cornea center $\boldsymbol{c}$ and the perimeter points of the pupil image.
4. The estimated POG is corrected for possible errors by a one-time per user calibration.
5. The image locations of the glints and pupil contour used in steps 2 and 3 above are extracted from images of the eye using image processing techniques.

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

3.1. POG Estimation

The POG $\boldsymbol{p}$ is the intersection of the vector $\boldsymbol{L}$ with the surface of the computer monitor. The monitor surface is modeled with a plane equation given the measured locations of three of the screen corners. The 3D parametric equation of a line defined by

$$\boldsymbol{p} = \boldsymbol{c} + t \cdot \boldsymbol{L} \qquad (1)$$

is used to determine the POG. This 3D vector equation has 4 unknowns (p_x, p_y, p_z, t). Adding the constraint that the POG must lie on the plane defined by the monitor provides the additional constraint required to solve for the POG explicitly, assuming that $\boldsymbol{c}$ and $\boldsymbol{p_c}$ are known.

The methods for determining the location of the cornea center $\boldsymbol{c}$ and the pupil center $\boldsymbol{p_c}$ required to compute the OA are given in the following sections.

3.2. Cornea Center Estimation

We have implemented an extension of Shih and Liu's proposed single camera method for estimating the location of the cornea center in 3D space [Shih and Liu 2004].

A ray can be traced from each glint light source $\boldsymbol{q}_i$ through the points $\boldsymbol{g_i}$, $\boldsymbol{o}$, and $\boldsymbol{i_i}$ as shown in Figure 2.

Shih and Liu noted that the set of points $\{\boldsymbol{q}_i\, \boldsymbol{g}_i\, \boldsymbol{c}\, \boldsymbol{o}\, \boldsymbol{i}_i\}$ are co-planar. An auxiliary coordinate system can be defined for each glint light source such that all these points lie in a plane defined by two axes of the coordinate system, thus reducing the solution space from 3 degrees of freedom to 2. A rotation matrix $\boldsymbol{R}_i$ and its inverse can be formulated for each glint to transform points between the auxiliary coordinate systems and the world coordinate system.

Using the geometry illustrated in Figure 3 it is possible to define the center of the cornea $\hat{\boldsymbol{c}}_{\boldsymbol{i}}$ in the auxiliary coordinate system as a function of a single unknown parameter $\hat{g}_{ix}$ for each glint as follows:

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

where

$$\hat{\alpha}_i = \cos^{-1}\left(\frac{-\hat{\boldsymbol{I}}_i \cdot \hat{\boldsymbol{Q}}_i}{\left\|-\hat{\boldsymbol{I}}_i\right\| \cdot \left\|\hat{\boldsymbol{Q}}_i\right\|}\right) \tag{3}$$

$$\hat{\beta}_i = \tan^{-1}\left(\frac{\hat{g}_{ix} \cdot \tan(\hat{\alpha}_i)}{\hat{l}_i - \hat{g}_{ix}}\right) \tag{4}$$

When the auxiliary cornea center $\hat{\boldsymbol{c}}_{\boldsymbol{i}}$ is transformed back to the world coordinate system using

$$\boldsymbol{c_i} = \boldsymbol{R_i^{-1}}\, \hat{\boldsymbol{c}}_{\boldsymbol{i}} \tag{5}$$

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

$$\boldsymbol{c_1} = \boldsymbol{c_2} \tag{6}$$

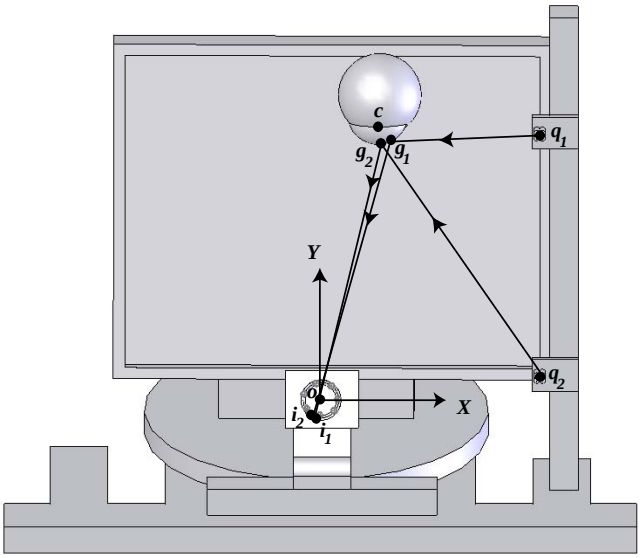


Figure 2 – Rays traced from multiple glint light sources to the surface of the camera sensor. The glint light source is located at point $\boldsymbol{q}_i$. The glints on the surface of the spherical cornea (center $\boldsymbol{c}$, radius r), are located at point $\boldsymbol{g}_i$. The focal point of the pin-hole camera model is located at point $\boldsymbol{o}$ and the image of the glint on the surface of the CCD sensor is located at point $\boldsymbol{i}_i$.

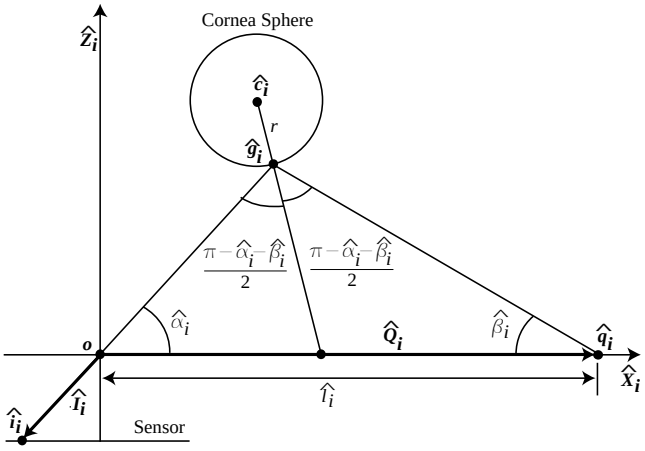


Figure 3 - Auxiliary coordinate system geometry. Each auxiliary coordinate system is defined with the origin at point $\boldsymbol{o}$, the $\hat{\boldsymbol{X}}_{\boldsymbol{i}}$-axis defined along $\hat{\boldsymbol{Q}}_{\boldsymbol{i}}$, the $\hat{\boldsymbol{Z}}_{\boldsymbol{i}}$-axis such that the vector $\hat{\boldsymbol{I}}_{\boldsymbol{i}}$ lies in the $\hat{\boldsymbol{X}}_{\boldsymbol{i}}$ - $\hat{\boldsymbol{Z}}_{\boldsymbol{i}}$ plane and the $\hat{\boldsymbol{Y}}_{\boldsymbol{i}}$-axis orthonormal to the $\hat{\boldsymbol{X}}_{\boldsymbol{i}}$ and $\hat{\boldsymbol{Z}}_{\boldsymbol{i}}$-axis. The vectors $\hat{\boldsymbol{I}}_{\boldsymbol{i}}$ and $\hat{\boldsymbol{Q}}_{\boldsymbol{i}}$ are the vectors from points $\boldsymbol{o}$ to $\hat{\boldsymbol{i}}_{\boldsymbol{i}}$ and from $\boldsymbol{o}$ to $\hat{\boldsymbol{q}}_{\boldsymbol{i}}$ respectively. The scalar $\hat{l}_i$ is the distance from points $\boldsymbol{o}$ to $\hat{\boldsymbol{q}}_{\boldsymbol{i}}$.

The over defined set of equations then consist of 9 equations with 8 unknowns which are solved numerically for $\boldsymbol{c}$ using a gradient descent algorithm.

3.3. Pupil Center Estimation

The second point required for the OA is the center of the pupil $\boldsymbol{p_c}$. The OA requires the center of the real pupil and not its refracted image recorded by the camera. The center of the real pupil can be found by computing the average of at least 2 opposing points on the real pupil perimeter, although in practice we found using 6 perimeter points provided a more robust estimate.

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

$$\boldsymbol{u_i} = \boldsymbol{k_i} + s_i \cdot \boldsymbol{K_i} \tag{7}$$

is traced from the pupil perimeter point $\boldsymbol{k_i}$ on the surface of the camera sensor to the surface of the cornea through the focal point of the pin-hole camera, as illustrated in Figure 4.

Adding the constraint that the point $\boldsymbol{u_i}$ must lie on the surface of the spherical cornea with center $\boldsymbol{c}$ and radius r

$$(u_{ix} - c_x)^2 + (u_{iy} - c_y)^2 + (u_{iz} - c_z)^2 = r^2 \tag{8}$$

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

$$\hat{\boldsymbol{u}}_i = \boldsymbol{u_i} + w_i \cdot \hat{\boldsymbol{K}}_i \tag{9}$$

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

$$\left\| \hat{\boldsymbol{u}}_i - \boldsymbol{c} \right\| = r_{ps} \tag{10}$$

where r_{ps} is defined as

$$r_{ps} = \sqrt{r_d{}^2 + r_p{}^2} \tag{11}$$

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

The pupil center $\boldsymbol{p_c}$ is computed by averaging the pupil perimeter values $\hat{\boldsymbol{u}}_i$. The OA can thus be computed with the estimated pupil and cornea centers, and ultimately used to estimate the POG as per equation 1.

3.4. Calibration Method

There are a number of simplifications employed in the models above which may result in POG inaccuracies. Such simplifications include the pin-hole camera model used to approximate the real camera and lens, the simplified eye model and the use of population averages for the parameters of the eye. A one-time calibration is performed on a per-user basis to correct

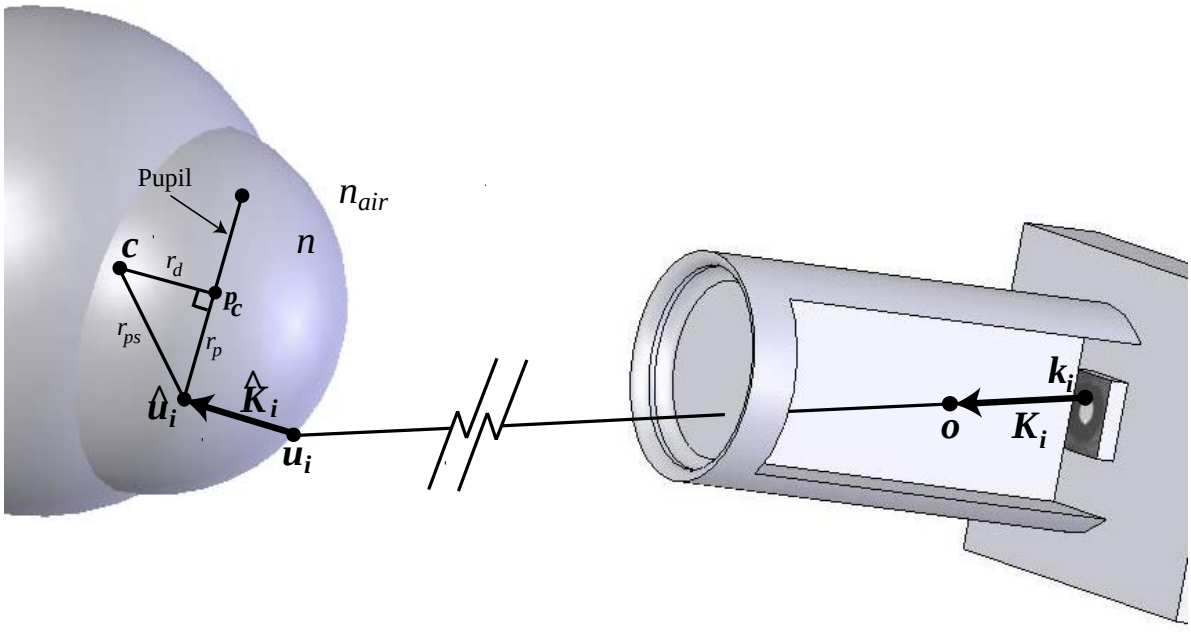


Figure 4 - Estimating the pupil center through ray tracing. The pupil perimeter image point on the surface of the camera sensor is denoted by $\boldsymbol{k_i}$. The ray $\boldsymbol{K_i}$ is traced from the camera sensor to a point $\boldsymbol{u_i}$ on the surface of the cornea. The refracted vector $\hat{\boldsymbol{K}}_i$ points from $\boldsymbol{u_i}$ to the real pupil perimeter point $\hat{\boldsymbol{u}}_i$. The distance from the center of the cornea to the center of the pupil is given by r_d and to the perimeter of the pupil by r_{ps}. The radius of the pupil is given by r_p, the index of refraction of air is given by n_{air} and the index of refraction of the aqueous humor is given by n.

for all of the possible sources of errors. The calibration procedure is automated, in that the system detects when to switch to the next calibration point, and can be performed in under 5 seconds. Figure 5 illustrates an example of the parameters used in performing the calibration and correction of the computed POG.

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

$$\boldsymbol{e}_i = \left(\boldsymbol{m}_i - \boldsymbol{n}_i\right) \tag{12}$$

Future POG estimates are adjusted by applying the four correction factors $\boldsymbol{e}_i$, each weighted inversely proportional to the distance the computed POG is from each of the original calibration POGs as shown in equations 13 through 15.

$$d_i = \left\| \boldsymbol{p}_{computed} - \boldsymbol{n}_i \right\| \tag{13}$$

$$w_i = \frac{1}{d_i \cdot \sum_{k=1..4} \frac{1}{d_k}} \tag{14}$$

$$\boldsymbol{p}_{corrected} = \boldsymbol{p}_{computed} + \left(\sum_{i=1..4} w_i \cdot \boldsymbol{e}_i \right) \tag{15}$$

In the event that any d_i is 0, w_i is set to 1 in equation 14 and the remaining weights set to 0.

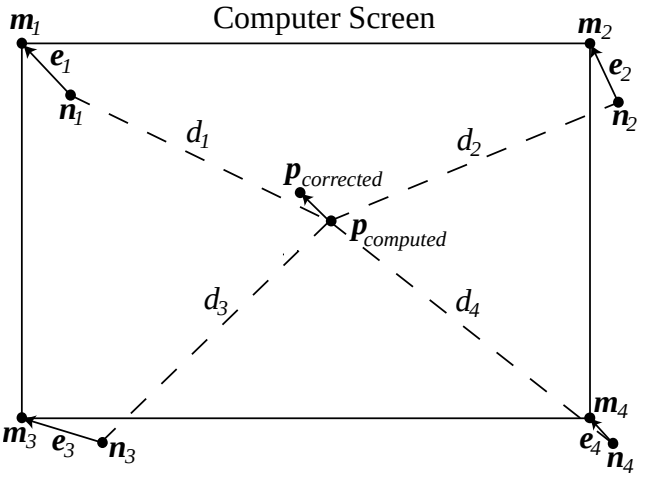
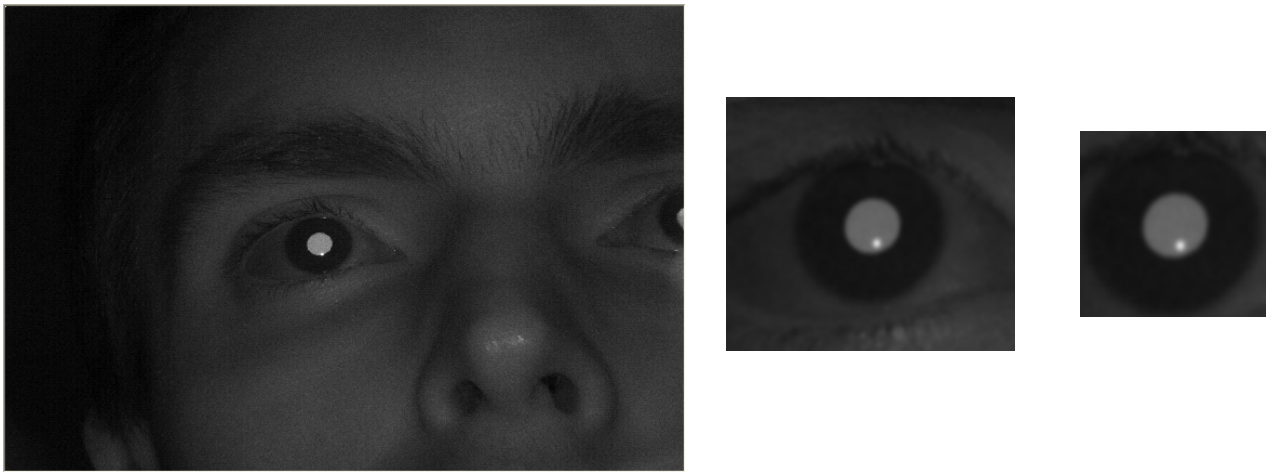


Figure 5 – Example POG calibration and correction. The system is initially calibrated by recording the computed POG locations $\boldsymbol{n}_i$ while the user is looking at known screen locations $\boldsymbol{m}_i$. The error $\boldsymbol{e}_i$ is used to convert future POG estimates $\boldsymbol{p}_{computed}$ to $\boldsymbol{p}_{corrected}$. The distance d_i from the point $\boldsymbol{p}_{computed}$ and each of the calibration locations $\boldsymbol{n}_i$ is used to weight the correction factors.

3.5. Eye and Feature Tracking

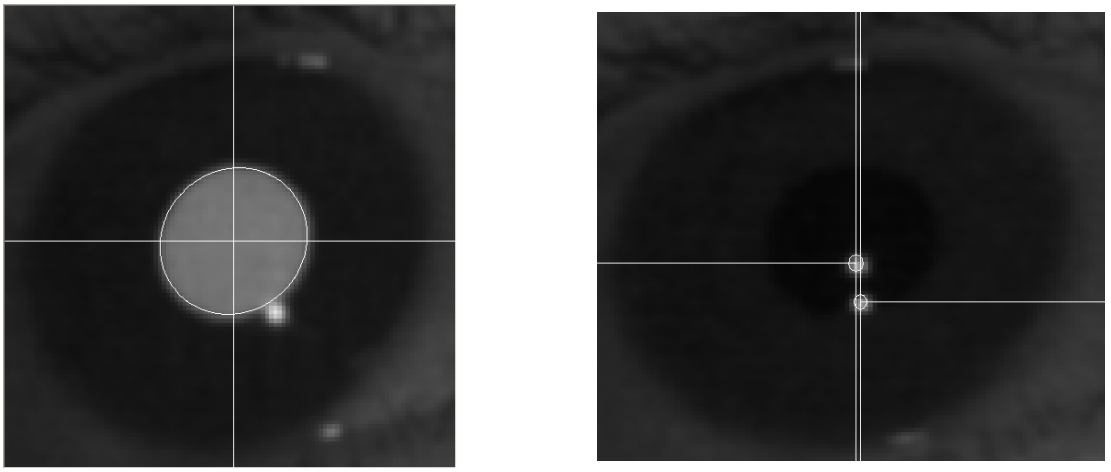
The points $\boldsymbol{i}_i$ used in Section 3.2 and $\boldsymbol{k}_i$ used in Section 3.3 are located on the surface of the camera CCD sensor. These locations are determined from information extracted from the recorded images using video processing techniques. The image processing required to extract these features is the most processor intensive operation of the system. To reduce the required processing time, a series of regions-of-interest (ROI) calculations are employed to reduce the quantity of image information. Initially the full image (Figure 6a) must be processed to detect the location of the eye. The ROI is then applied, sized to contain only the image of the eye (Figure 6b). When the pupil in the eye is detected roughly the size of the ROI reduces further to contain just the cornea and pupil for final processing (Figure 6c). When the image processing has completed, the ROI is increased in size to encompass the eye and re-centered on the estimated center of the pupil image contour for the next processing loop. Re-centering the ROI on the pupil allows the ROI of Figure 6b to effectively track the eye without having to reprocess the entire image. If the eye is lost (due to a blink) or moves outside of the ROI within one frame the entire image is reprocessed and the ROI then reapplied.

To compute the pupil center, points along the perimeter of the pupil contour image are required. The image differencing technique [Ebisawa 1998] is used to aid in identifying the pupil contour. Images are recorded with alternating light sources, one in which the pupil is brightly illuminated from lighting close to the optical axis of the camera and one in which the scene is illuminated by off axis light sources. The off-axis lighting illuminates the face to the equivalent intensity of the bright pupil image but does not cause the pupil to reflect as brightly. A ring of



(a) (b) (c)

Figure 6 - Regions of interest used to decrease processing time; (a) full sized scene image; (b); eye ROI; (c) cornea and pupil ROI.



(a) (b)

Figure 7 - Identified pupil (a) and dual glints (b). An ellipse equation is fitted to the perimeter of each identified contour. The pupil perimeter ellipse is used to estimate the real pupil center location, while the centers of the dual glint ellipses are used to estimate the center of the cornea.

LED's located around the optical axis of the camera are used to generate the bright pupil image. Two lights located beside the computer screen generate the dark pupil image, which will also then contain the required dual glints. Subtracting the dark pupil image from the bright pupil image enhances the pupil contour, making it easier to detect in the scene.

The pupil contour is detected in two stages to improve system accuracy and performance. The pupil is first identified quickly and roughly in the scene using the difference image. A finer pupil detection algorithm is then used to extract the pupil contour from just the bright pupil image. Using just the bright pupil image avoids differencing artifacts due to motion between image frames. The fine pupil detection method also compensates for possible artifacts which may corrupt the pupil perimeter, such as glints or eyelashes. An example of the identified pupil contour is shown in Figure 7a.

The locations of the centers of the dual glints in the recorded images are required for computing the center of the cornea. The glints off the surface of the cornea result in the brightest pixels in the image and are easily detected. Possible artifacts are rejected using the expected displacements between the two glint centers. Examples of the identified dual glints are shown in Figure 7b.

4. Evaluation

4.1. Implementation

The physical implementation of the eye-gaze tracking system is shown in Figure 8. The system was tested on a moderately powerful AMD 1.4 GHz computer and a higher end Pentium IV 2.8 GHz computer. Using different computers provides some insight into how well the system will perform with respect to the available processing power.

The system was also tested with two different cameras, one with a resolution of 1024x768 pixels and a frame rate of 15 Hz, and another camera with a resolution of 640x480 pixels and a frame rate of 30 Hz. Both cameras are versions of the digital Firewire based Dragonfly from Point Grey Research. Using different cameras also provides insight into how the system may perform with respect to available frame rates and image resolutions. The higher resolution camera had an allowable range of motion of approximately 14x12x20 cm (width x height x depth) while for the faster but lower resolution camera the allowable range of motion reduced to approximately 7.5x5.5x19 cm. The width and height are specified at approximately the midpoint of the field of view volume. The focal length of the lens for both cameras was 32 mm.

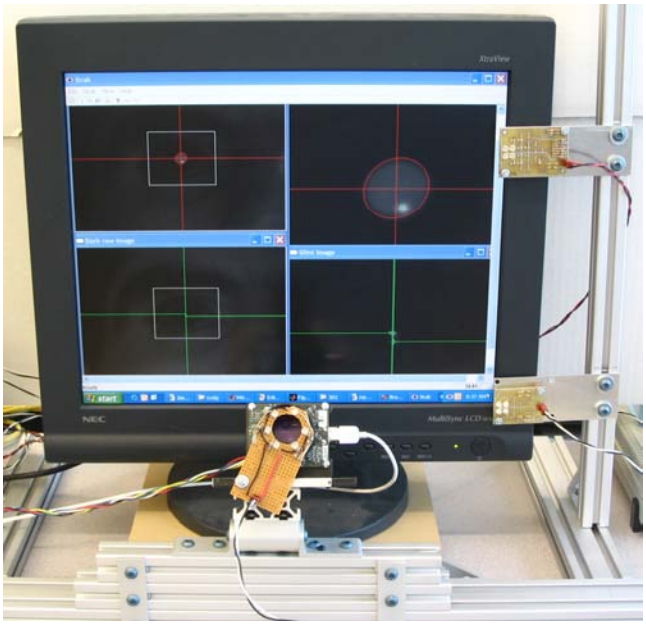


Figure 8 – The physical system implementation. The digital Firewire camera is located below the screen and oriented towards the users face. The on-axis lighting is provided by the ring of LEDs surrounding the camera lens, while the dual glint off-axis light sources are located to the right of the monitor. The entire assembly is mounted on extruded aluminum rails to fix the relative displacements of the LEDs, camera and screen.

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

4.2. Free Head Motion

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

Calibration was performed by the system user at position 1. Accuracy was then measured by recording the POG error when looking at points on a 4x4 grid on the screen. Average accuracy was determined for a total of 7 different head locations within the allowable field of view. An electromagnetic position tracker was worn by the user during this test to verify that the full field of view was spanned. The range of X, Y and Z positions reported by the position tracker across the field of view volume was 14.2 cm, 12.3 cm, and 20.6 cm respectively.

The average accuracies (difference between POG estimate and reference point) at each head location are listed in Table 1 in units of screen pixels. The screen had dimensions of 35 cm and 28 cm in width and height respectively, and a resolution of 1280x1024 pixels. Accuracy in pixels rather than degrees of visual angle is reported here because the distance from the eye to the screen required for computing degrees of visual angle was not readily available. For ease of comparison, accuracy in the subsequent test is reported in pixels as well. An average value for the distance from eye to screen is used to convert from pixels to degrees of visual angle in the Discussion in section 5.

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

Table 1 – Average POG accuracy measured across a 4x4 grid for each different head position.

4.3. Different Hardware Configurations with Different Subjects

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

The test procedure was to perform a calibration and then record a dataset on the 4x4 grid (Trial 1). The user was asked to move away from the system, then to return and sit down in front of the computer again, resulting in a different head position. A second 4x4 grid dataset was recorded away from the calibrated position (Trial 2). The average accuracies for these tests are shown in Table 2.

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

Table 2 – Average POG accuracy measured across a 4x4 grid for multiple trials, subjects and system configurations.

The time required to process one video image for each system configuration was recorded both when the ROI was locked on the eye and when the eye was lost. When the ROI is locked on the eye only a small portion of the image is processed; when the ROI is lost the full image must be processed to reacquire the eye. These processing times are shown in Table 3.

	Processing Time (ms)	
	ROI Lock	Eye Lost
AMD – 30 Hz 640x480	27	35
AMD – 15 Hz 1024x768	28	110
P4 – 30 Hz 640x480	10	32
P4 – 15 Hz 1024x768	10	40

Table 3 - Processing times per system update when the ROI is locked on the eye and when the eye is lost. Each of the four combinations of system configurations were tested.

5. Discussion

Across the span of possible head positions (see Table 1), the best average pixel errors for the uncalibrated positions in X and Y are [20.9, 18.7] pixels and at the worst are [46.5, 21.0] pixels. Across various hardware configurations and different subjects (see Table 2), when the eye was not at the calibration location (Trial 2), the best average errors in X and Y are [14.8, 21.7] pixels and the worst are [29.1, 34.2] pixels. At an average distance of 75 cm from the eye to the screen for Trial 2 the best average accuracy in degrees of visual angle is 0.46° and the worst is 0.90°.

The system was able to estimate the POG over the full range of allowable head positions and with variations in processing power, camera resolution and camera frame rate. When the ROI was locked on to the eye, there was little difference in processing time required between the higher and lower resolution cameras, due to the equivalent ROI size for both cameras. These times indicate that the AMD system could achieve a maximum update rate of 35Hz while the P4 system could achieve a maximum update rate of 100Hz. The maximum update rates however were limited due to the lower frame rates of the cameras to 15 Hz for the higher resolution camera and 30 Hz for the lower resolution camera. The system update rate matches the camera frame rates even though alternating bright and dark pupil images are recorded (required for the image differencing technique). The equivalent rates are achieved by estimating the POG from the latest captured image and the previously captured image (either bright then dark or dark

then bright pupil images). When the ROI lock was lost, the processing time increased in all cases, thus reducing the effective system update rate. The processing time increased more for the higher resolution camera than for the lower resolution camera when the ROI lock was lost, as expected.

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

6. Conclusions

This paper describes the design, implementation and evaluation of an eye-gaze tracking system that meets key requirements as described in the introduction. As quantified below, the single camera, multiple glint system achieves the accuracy claimed in the presence of free head motion within the field of view of the camera. Over various combinations of hardware configurations and subjects the best accuracy achieved with the eye away from the calibration position (Trial 2), averaged over the 4x4 screen grid, was 0.46° and the worst was 0.90° of visual angle, which is comparable to that of other reported systems. System accuracy is highest at the calibrated position and degrades slightly as the head is moved away.

The system developed has an allowable range of head positions of approximately 14x12x20 cm for a 1024x768 pixel resolution camera. As expected, although higher camera resolution increases the allowable range of head positions, for equivalent spatial resolution it does not necessarily improve eye gaze accuracy. There are no moving parts, resulting in fast re-acquisition times. For the P4 system, re-acquisition of the eye after loss of lock can be achieved in 67 ms for a 15 Hz camera and 33 ms for a 30 Hz camera. Employing a single camera with no moving parts also allows the use of a one-time per user calibration procedure that takes less than 5 seconds.

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

Acknowledgements

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

References

BEYMER, D., AND FLICKNER, M. 2003. Eye gaze tracking using an active stereo head. In *Proceedings of Computer Vision and Pattern Recognition*, vol. 2, 451-458.

EBISAWA, Y. 1998. Improved video-based eye-gaze detection method. *IEEE Transactions on Instrumentation and Measurement*, 47:4, 948-955.

GOSS, D. A., AND WEST, R. W. 2002. Introduction to the optics of the eye. *Butterworth-Heinemann*.

HANSON, E. K. S. 2004. Focus of attention and pilot error. In *Proceedings of Eye Tracking Research and Applications (ETRA2004)*, 60.

LOHSE, G. L. 1997. Consumer eye movement patterns on yellow pages advertising. *Journal of Advertising*, 26 (1), 61-73.

MAJARANTA, P. AND RAIHA, K. J. 2002. Twenty years of eye typing: systems and design issues. In *Proceedings of Eye Tracking Research and Applications (ETRA2002)*, 15-22.

MORIMOTO, C., AMIR, A., FLICKNER, M. 2002. Free head motion eye gaze tracking without calibration. In *CHI'02 extended abstracts on Human factors in computer systems*, 586-587.

MORIMOTO, C. H., AND MIMICA, M. R. M. 2005. Eye gaze tracking techniques for interactive applications. *Computer Vision and Image Understanding*, vol 98, 4-24.

NOUREDDIN, B., LAWRENCE, P. D., AND MAN, C. F. 2005. A non-contact device for tracking gaze in a human computer interface. *Computer Vision and Image Understanding*, vol 98, 52-82.

OHNO, T., AND MUKAWA, N. 2004. A free-head, simple calibration, gaze tracking system that enables gaze-based interaction. In *Proceedings of Eye Tracking Research and Applications (ETRA2004)*, 115-122.

OHNO, T., MUKAWA N., AND YOSHIKAWA, A. 2002. FreeGaze: A gaze tracking system for everyday gaze interaction. In *Proceedings of Eye Tracking Research and Applications (ETRA2002)*, 125-132.

PETERSSON, L., FLECTCHER, L., BARNES, N., AND ZELINSKY, A. 2004. An interactive driver assistance system monitoring the scene in and out of the vehicle. In *Proceedings of IEEE International Conference on Robotics and Automation (ICRA2004)*, vol 4, 3475 – 3481.

RAYNER, K. 1998. Eye movements in reading and information processing: 20 Years of Research. *Psychological Bulletin 124* (3), 372-422.

SHIH, S. AND LIU, J. 2004. A novel approach to 3-D gaze tracking using stereo cameras. *IEEE Transactions on Systems, Man, and Cybernetics – Part B*, vol. 34, No 1, 234-245.

YOO, D. H., AND CHUNG, M. J. 2005. A novel non-intrusive eye gaze estimation using cross-ratio under large head motion. *Computer Vision and Image Understanding*, vol. 98, 25-51.

YOUNG, L. AND SHEENA, D. 1975. Survey of eye movement recording methods. *Behavior Research Methods and Instrumentation*, 7, 397-429.

ZHAI, S., MORIMOTO, C., AND IHDE, S. 1999. Manual And Gaze Input Cascaded (MAGIC) Pointing. In *Proceedings of SIGCHI Conference on Computer Human Interaction (CHI1999)*, 246-253.