Fixation Precision in High-Speed Noncontact Eye-Gaze Tracking

Craig Hennessey, Borna Noureddin, and Peter Lawrence, *Senior Member, IEEE*

***Abstract*—The precision of point-of-gaze (POG) estimation during a fixation is an important factor in determining the usability of a noncontact eye-gaze tracking system for real-time applications. The objective of this paper is to define and measure POG fixation precision, propose methods for increasing the fixation precision, and examine the improvements when the methods are applied to two POG estimation approaches. To achieve these objectives, techniques for high-speed image processing that allow POG sampling rates of over 400 Hz are presented. With these high-speed POG sampling rates, the fixation precision can be improved by filtering while maintaining an acceptable real-time latency. The high-speed sampling and digital filtering techniques developed were applied to two POG estimation techniques, i.e., the high-speed pupil–corneal reflection (HS P-CR) vector method and a 3-D model-based method allowing free head motion. Evaluation on the subjects has shown that when operating at 407 frames per second (fps) with filtering, the fixation precision for the HS P-CR POG estimation method was improved by a factor of 5.8 to 0.035° (1.6 screen pixels) compared to the unfiltered operation at 30 fps. For the 3-D POG estimation method, the fixation precision was improved by a factor of 11 to 0.050° (2.3 screen pixels) compared to the unfiltered operation at 30 fps.**

***Index Terms*—Eye-gaze tracking, fixation precision, head free, high speed, human–computer interface, noncontact, remote.**

I. Introduction

EYE-GAZE tracking systems offer great promise as an interface between humans and machines. The eye gaze can provide insight into the intention of a user, as a user typically looks at objects of interest before acting upon them [1]. Real-time eye-gaze tracking systems allow dynamic interaction between the user and the system using the human visual system for both feedback and control [2]. Tracking the fixations of a user provides a means for using the eye-gaze information as a pointing device [3]. However, the use of eye gaze as an input modality has not had widespread appeal with the general population due in part to the shortcomings of the current eye-gaze tracking technology. Some of the key issues that must be improved upon are accuracy, precision, latency, ease of use, comfort, and cost [4], [5].

Manuscript received January 24, 2007; revised May 28, 2007. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) Chair in Design Engineering and by the NSERC under Discovery Grant 4924-05. This paper was recommended by Associate Editor Q. Zhu.

The authors are with The University of British Columbia, Vancouver, BC V6T 1Z4, Canada (e-mail: peterl@ece.ubc.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSMCB.2007.911378

Recent advances in the development of noncontact video-based eye-gaze tracking systems have removed the need for contact with the user and have greatly improved the user's comfort [6]. Noncontact systems coupled with advanced point-of-gaze (POG) estimation algorithms that compute the location of the eye in 3-D space can now operate without significantly restricting the user's head motion. The increased freedom of motion greatly improves the ease of use of the system.

Eye-gaze tracking systems in general and noncontact video-based systems in particular suffer from low precision or fluctuating fixation estimates. The low precision is caused not just by sensor and system noise but is also due in part to the natural motions of the unconstrained head and eye. Considerable research has focused on developing real-time applications that compensate for the low precision including the use of large pointing targets [7], [8], fisheye lenses [9], and enhanced pointing algorithms such as MAGIC pointing [3] and the Grab and Hold Algorithm [10].

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

II. Background

A. Eye Movements

Although the surrounding world appears stable, the head and eyes are continuously in motion, and the images formed on the retinas are constantly changing. The stable view of the external world is only an artificially stabilized perception. Natural human vision is typically made up of short relatively stable fixations connected by rapid reorientations of the eye (saccades). It is during fixations that the sensory system of the eye collects information for cognitive processing. During saccades, the sensitivity of the visual input is reduced [11].

Fixations typically remain within 1° of visual angle and last from 200–600 ms [1]. While fixating, the eye slowly drifts, with a typical amplitude of less than 0.1° of visual angle and a frequency of oscillation of 2–5 Hz. This drift is corrected by small fast shifts in eye orientation called

microsaccades, which have a similar amplitude to the drift. Superimposed on this motion is a tremor with a typical amplitude of less than 0.008° of visual angle with frequency components from 30–100 Hz and at times up to 150 Hz [12]. These small eye motions during a fixation are thought to be required to continuously refresh the sensors in the eye, as an artificially stabilized image will fade from view [13].

Saccades most frequently travel from 1° to 40° of visual angle and last for 30–120 ms. Between saccades, there is typically a 100- to 200-ms delay [1]. A number of other task-specific eye motions exist, such as smooth pursuit, nystagmus, and vergence, which are not often found in the normal interaction between a user and a desktop monitor. The focus of this paper is on the POG estimates during fixations, which are located between saccadic reorientations of the eye.

B. Fixation Detection and Filtering

A clear identification of the beginning and end of a fixation within the raw eye data stream is important as filtering should only be performed on POG data located within a single fixation. Poor identification of the beginning or end of a fixation may result in a degradation in fixation precision by incorporating POG data from saccades or neighboring fixations. There have been a number of methods developed for identifying the start and end of fixations in raw eye data streams using position, velocity, and acceleration thresholding based on *a priori* knowledge of the behavior of eye-gaze movements [14], [15]. The fixation identification method used in this paper is based on the position variance of the eye data, as described in [1].

Due to the natural motions of the eye, the fixation precision in eye-gaze tracking systems may be low, which limits the range of potential applications. However, as noted in [1], this low precision can be improved by low-pass filtering the estimated POG data to reduce noise at the expense of increased latency. The desired degree of filtering within a fixation will depend on the particular application under consideration. For high precision, a higher-order filter may be used at the expense of a longer latency or lag between the start of a fixation and the desired filter response. Alternatively, a lower-order filter may be used to allow the POG fixation to drift slightly over time to follow the natural drift of the eye. Using digital finite-impulse-response (FIR) filtering techniques allows the filter order to be easily modified. Moreover, clearing the filter history (memory) provides a simple means for resetting the filter when a fixation termination is detected.

C. Eye-Gaze Tracking Systems

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

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

In the late 1980s, Hutchinson *et al.* [16] developed a noncontact video-based system that used the P-CR vector method for computing the POG. The P-CR method greatly enhanced the usability of remote eye-gaze tracking systems by providing tolerance to minor head displacements. The system they developed was targeted to work with the severely disabled who had no other easily available means of communication. Images were recorded with a resolution of 512 × 480 pixels with a POG sampling rate of 30 Hz. After calibration, the average accuracies for this method are typically 0.5° to 1° of visual angle.

Over the past two decades, the P-CR vector method has been the favored means for noncontact video-based POG estimation. However, the P-CR method still required a relatively stable head position. The accuracy of the method considerably degrades as the head is displaced from the calibration position [6].

To allow for free head motion, Shih and Liu [17] developed a novel 3-D model-based method for estimating the eye gaze. Using models of the system, camera, and eye, their algorithm was designed to accurately estimate the POG regardless of head location. Their system used two RS-170-based cameras and frame grabbers to record images with a resolution of 640 × 280 pixels at 30 Hz. The average accuracy was shown to be better than 1° of visual angle. Unfortunately, their system design required the cameras to be quite close to the subjects' eyes to acquire high spatial resolution images, which restricts the freedom of head motion due to the limited camera field of view.

To overcome the limitation of a narrow field of view, Ohno and Mukawa [18] developed a 3-D model-based system with a camera mounted on a pan/tilt mechanism with a narrow-angle (NA) lens and two fixed cameras with wide-angle (WA) lenses. The fixed cameras used stereo imaging to determine the location of the head within the scene and directed the pan/tilt mechanism to orient the NA camera toward the eye. The WA cameras recorded images with a resolution of 320 × 120 pixels, while the NA camera recorded images with a resolution of 640 × 480 pixels, all at frame rates of 30 Hz. The system accuracy was reported as better than 1.0° of visual angle. The pan/tilt mechanism allowed the NA camera to track the motion of the eye with a larger effective field of view; however, the speed at which the mechanism could move was not sufficient to keep up with the faster motion of the head and eye, which results in loss of tracking and slow reacquisition.

Beymer and Flickner [19] used high-speed galvonometers for their 3-D model-based system in an attempt to overcome the limitations of the slow pan/tilt systems. A pair of fixed WA cameras used stereo imaging to direct the orientation of two NA cameras by controlling the pan and tilt of the rotating

lightweight mirrors mounted on galvonometers. The focus of each camera was controlled with a lens mounted on a bellows and driven by another motor. The NA cameras recorded images (with a typical resolution of 640 × 480 pixels) at a frame rate of 30 Hz. Due to the significant processing involved in the system, a POG sampling rate of only 10 Hz was achieved. The accuracy reported for this system was 0.6° of visual angle. While their system was capable of tracking the eye in the presence of natural high-speed head motion, considerable calibration was required, and the overall complexity resulted in a low POG sampling rate.

The 3-D model-based system by Hennessey *et al.* [20] was developed to minimize the physical system complexity while still allowing for fast head motion. The system was based on a single fixed camera with a high-resolution sensor and no moving parts. The higher resolution sensor allowed a larger range of head motion with the eye remaining in the field of view of the camera while still providing images with sufficient spatial resolution for the eye-tracking system to correctly operate. The system algorithms were designed to track the motion of the eye within the image and only operate on the portion of the image containing the eye. Processing only the portion of the image containing the eye allowed the POG to be rapidly computed regardless of the overall image resolution. At the time of system development, a camera with a resolution of 1024 × 768 pixels was available with a maximum frame rate of 15 Hz. Using this system, accuracies better than 1° of visual angle were achieved.

Fixation precision has not often been reported in the evaluations of 3-D model-based eye-gaze tracking systems as the focus tended to be on the basic system functionality and the accuracy of the novel POG algorithms. However, Yoo and Chung [21] did provide some insight into the fixation precision of their free head motion eye-gaze tracking system. Using a similar system design as Ohno and Mukawa [18], they reported an accuracy of 0.98° in horizontal error and 0.82° in vertical error when operating at 15 Hz. The precision in standard deviations was reported in millimeters, which converted to 0.84° of visual angle. We believe that fixation precision is an important parameter in the evaluation of the performance of eye-gaze tracking systems, and the goal of this paper is to present methods for enhancing the fixation precision.

III. Methods

A. POG Estimation

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

1) P-CR Method: The simplicity of the P-CR vector method and its ability to handle minor head motions led to its widespread adoption. As the eye rotates to observe different points, the image of the reflection off the spherical corneal surface remains relatively fixed. The corneal reflection, which is generated by external lighting, provides a reference point for determining the relative motion of the pupil. A simple mapping is used to relate the 2-D POG screen vector to the 2-D image vector formed from the center of the corneal reflection to the center of the pupil, as shown in Fig. 1.

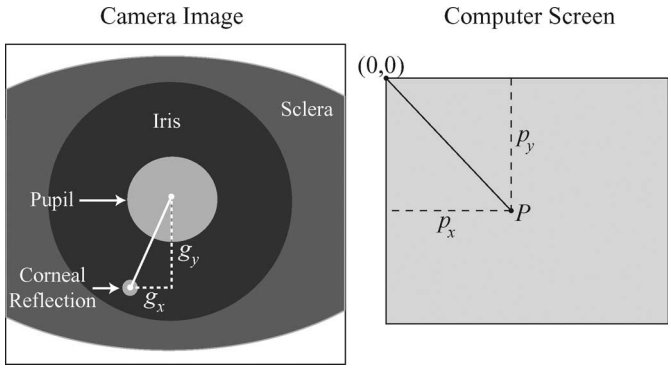


Fig. 1. Example of a portion of a recorded bright pupil image shown to illustrate the P-CR vector. In the P-CR method, the vector (g_x, g_y) is determined from the center of the corneal reflection to the center of the pupil. A mapping is then defined to relate the P-CR vector to the POG screen coordinates (p_x, p_y).

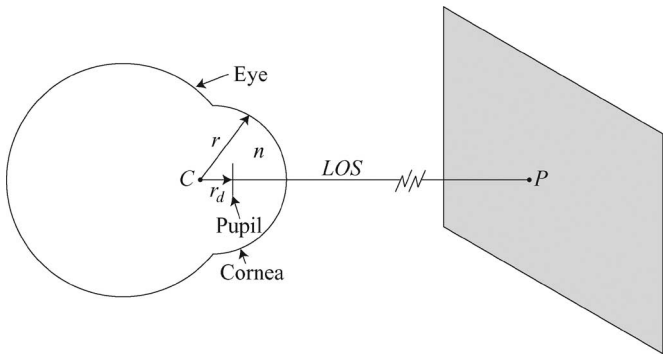


Fig. 2. Three-dimensional model-based method for computing the POG based on determining the location of the center of the cornea and the LOS vector. Using (2), the POG can be found by tracing the LOS vector from C to the surface of the screen P. The model of the eye is based on the schematic description by Gullstrand, which in this case includes three parameters, i.e., the radius of the model of the corneal sphere r, the distance from the center of corneal sphere to the center of pupil r_d, and the index of refraction of the aqueous humor fluid n.

Independent polynomial equations are determined to relate the 2-D P-CR vector (g_x, g_y) to each of the 2-D POG screen coordinates (p_x, p_y). The polynomial order varies between different system designs but is most often of first order as

$$\begin{aligned} p_x &= a_0 + a_1 g_x + a_2 g_y + a_4 g_x g_y \\ p_y &= b_0 + b_1 g_x + b_2 g_y + b_4 g_x g_y. \end{aligned} \tag{1}$$

It has been shown that small increases in accuracy may be achieved by increasing the order of the polynomial at the expense of a decrease in robustness to head motion and the need for an increasing number of calibration points [22].

The parameters a_i and b_i are determined from a calibration procedure in which the user sequentially fixates on a number of known screen locations while the P-CR vector is recorded. In the case of a first-order polynomial fit, a minimum of four calibration points are required to solve for the four unknowns in each of the two equations in (1) typically using a least squares method.

2) 3-D Model-Based Method: Algorithms based on 3-D models have been developed to overcome the degradation in accuracy that the P-CR method suffers with larger head

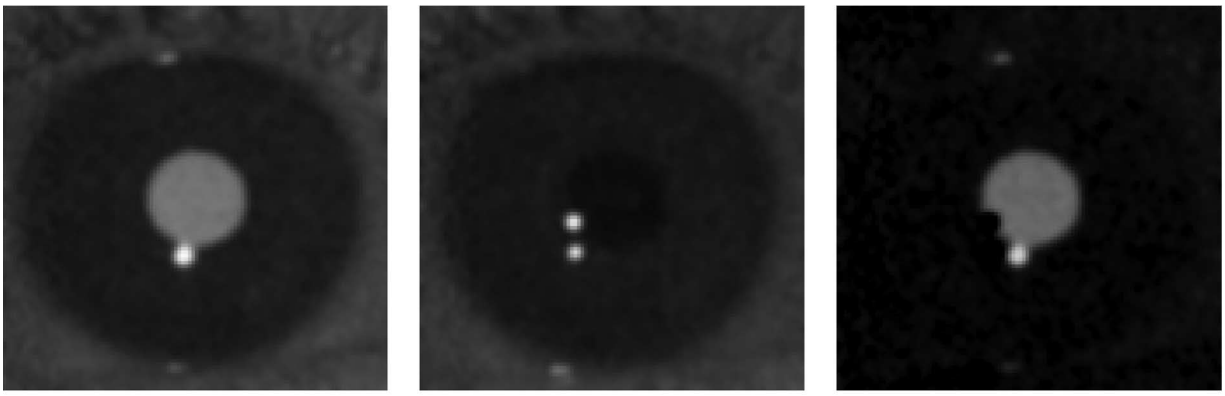


Fig. 3. Illustration of the bright pupil and image differencing techniques. The bright pupil in (a) is illuminated with on-axis lighting, while the dark pupil in (b) is illuminated with off-axis lighting. The background intensity of the two images is similar, which after differencing [(a) and (b)] results in a bright pupil on an almost blank background as shown in (c). (a) Bright pupil image. (b) Dark pupil image. (c) Difference image.

movements. The 3-D model-based methods compute the position of the eye in 3-D space, which is then used in computing the POG regardless of the head and eye position. There are a large variety of 3-D model-based algorithms, although each technique is typically based on a model of the physical system, camera, and eye. The physical system is geometrically modeled through physical measurement or using optical methods as in [17]. The camera lens is modeled as a pinhole with the parameters identified through camera calibration [23], [24]. The models of the eye are most often based, in varying levels of sophistication, on the schematic eye developed by Gullstrand [25]. An example of a typical eye model with three parameters is shown in Fig. 2. A per-user calibration is required to fit the eye model parameters to individual users.

Feature information is extracted from recorded images and fitted to the system models to solve for the location of the eye in 3-D space, the line of sight (LOS), and ultimately the POG, as shown in Fig. 2. The location of the eye in 3-D space is found by determining the center of the cornea C, when modeled as a spherical surface, using triangulation with images of multiple corneal reflections. With the 3-D location of the cornea center and the image location of the center of the pupil, the 3-D LOS vector can be computed. The LOS can be traced from C to intersect with any surface point P in the system by determining the parameter t in (2). The object of intersection is typically the surface of the computer screen that is parameterized as a plane in the system model

$$P = C + t \cdot \text{LOS}. \tag{2}$$

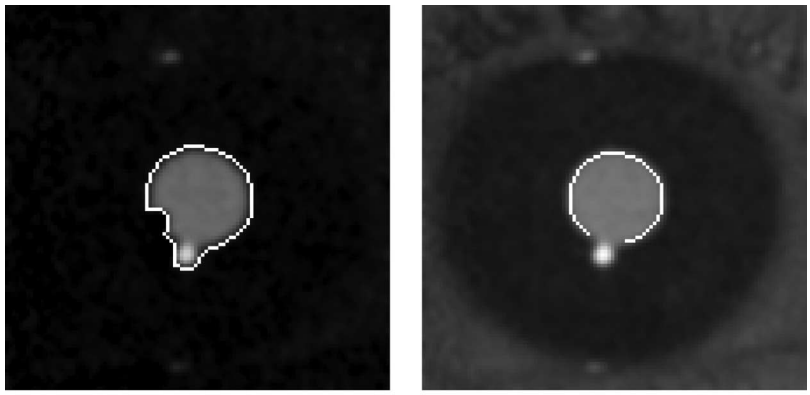
B. Image Processing

Both the P-CR vector and the 3-D model-based methods for estimating the POG require features extracted from the recorded images. The P-CR method requires the location of the pupil and the location of a single corneal reflection, while the 3-D method requires the pupil and at least two corneal reflections for triangulation. The locations of the pupil and corneal reflections are found by identifying the perimeter of their respective image contours. The pupil contour perimeter can be considerably difficult to segment due to the low contrast between the pupil and the surrounding iris. The corneal reflections can be difficult to segment due to their small size, which is often less than 3×3 pixels. Varying levels of ambient light can compound the feature extraction difficulty.

To improve the performance of the feature extraction task, the bright pupil and the image differencing techniques are used to create a high-contrast image of the pupil [26], [27]. Computing a difference image from alternating bright and dark pupil images removes most of the background features, which ideally leaves only the high-contrast pupil on a black background. An example of the bright pupil and the image differencing techniques are shown in Fig. 3. Using a single on-axis light source generates a single corneal reflection that is used in the P-CR POG estimation method. By using two off-axis light sources for the dark pupil image, the two corneal reflections required for the 3-D method are generated.

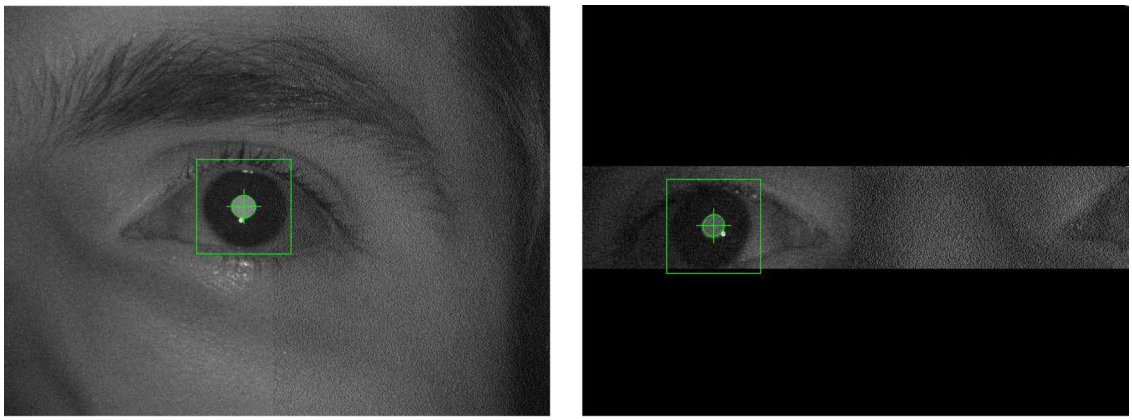
While the image differencing technique aids in the identification of the pupil contour within the image, it is also susceptible to significant artifacts that may corrupt the identified contour. When the difference image is computed, the corneal reflections formed by the off-axis lighting in the dark pupil image can result in removing a portion of the pupil as seen in the lower left side of Fig. 3(c). Moreover, the addition to the pupil contour of the corneal reflection from the on-axis lighting is also seen. The difference image is also susceptible to significant artifacts due to interframe motion. The interframe motion may distort the extracted pupil contour by misaligning the bright and dark pupil images, which will significantly impact the accuracy of the POG estimation algorithms.

To avoid the inaccuracies resulting from interframe motion and image differencing, a two-stage approach to pupil detection was used. The first stage of pupil extraction determines the image difference pupil as described above. The corneal reflections are then identified in both the bright and dark pupil images based on their proximity to the roughly identified difference pupil [see Fig. 4(a)]. In the second stage of pupil identification,



(a) (b)

Fig. 4. Example of the results of the two-stage pupil detection algorithm. In (a), the detected perimeter of the identified image difference pupil contour is shown overlying the difference image. Using the difference pupil contour as a guide, the pupil perimeter is detected in the bright pupil image as shown in (b). The gap in the pupil perimeter is a result of masking off the on-axis corneal reflection, which is subsequently compensated for by fitting an ellipse to the bright pupil contour perimeter. (a) Difference pupil contour. (b) Bright pupil contour.



(a) (b)

Fig. 5. ROIs are used to reduce the quantity of image information to process as well as to increase the camera frame rate. In (a), only the software ROI is applied to the original full-sized bright pupil image (640 × 480 pixels). Only the portion of the image within the rectangular box (110 × 110 pixels) surrounding the eye will be processed. In (b), the hardware ROI (640 × 120 pixels) has been applied in addition of the software ROI. (a) Full image and software ROI. (b) Hardware and software ROIs.

the pupil contour is segmented in only the bright pupil image using the previously detected difference pupil as a guide. Using only the bright pupil image avoids errors due to interframe motion and the accidental removal of the pupil area by the subtraction of the dark pupil–corneal reflections. The final step of the second stage is to mask off the portion of the pupil contour that may be due to the addition of the on-axis corneal reflection [see Fig. 4(b)]. The resulting pupil perimeter retains its elliptical shape when compared with the initial roughly identified pupil perimeter. For a more detailed description of the methods used for pupil and corneal reflection segmentation, see [28].

Before transferring the identified pupil and glint locations to the POG estimation algorithms, the identified contour perimeters are further refined using an ellipse fitting algorithm, which is both fast (computationally efficient) and robust to noise [29]. The subpixel accuracy in the identification of the contour centers may be achieved by using the center of the equation of an ellipse fitted to the contour perimeters [30]. In addition, using an ellipse fitted to the available pupil perimeter points compensates for the loss of data when a gap appears as a result of the masking operation to remove the corneal reflection from the on-axis lighting.

C. POG Sampling Rate

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

1) Software ROI: Image processing algorithms can be considerably time consuming due to the large quantity of information to process. To greatly reduce the processing load for our system, a software-based region of interest (ROI) was employed to constrain the processing to only the image area of interest. In the design of our system, rather than using mechanical tracking, the camera field of view encompasses a large area that allows the eye to move around within the scene. Accordingly, only a small portion of the overall scene contains the information of interest, as shown in Fig. 5(a).

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

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

The frame rate for the camera used by our system (described in Section III-D) only increased by skipping image rows, and no frame rate improvement was achieved for skipping image columns. Using the software ROI in conjunction with the hardware ROI allowed the flexibility to maximize the frame rate while minimizing the required processing. Similar to the software ROI, the location of the hardware ROI was recentered on the pupil image frame to track the motion of the eye. Unfortunately, changing the location of the hardware ROI in real time aborted the exposure of the current image, which results in an underexposed image for one frame. To minimize the number of hardware ROI location changes, the size of the hardware ROI was chosen to be the full width of the original image and slightly larger than the height of the cornea, while the size of the software ROI was set to the width of the cornea and slightly smaller than the height of the hardware ROI, as shown in Fig. 5(b). The software ROI then tracks all horizontal motion and most small vertical motions without requiring a change in the hardware ROI location. The hardware ROI is then only repositioned for larger vertical displacements in the position of the eye.

3) Image Sequencing: Recording alternating bright and dark pupil images for the image differencing technique aids in the detection of the pupil within the overall scene; however, it also reduces the effective POG sampling rate. When a $1:1$ ratio of alternating bright and dark pupil images is recorded, the P-CR method can only generate a unique POG (P_i) at half the camera frame rate, as shown in Table I, since all the information required to compute the POG is contained within the bright pupil image. Recall that for the P-CR POG estimation method, the image features required are the pupil and a single corneal reflection, which are both found in the bright pupil image. The 3-D method uses the image information from both the bright and dark pupil images, and as such can compute a unique POG (P_i) at the camera frame rate by using features from each current image (f_{i+1}) along with the image previously recorded (f_i).

TABLE I
POG SAMPLING SEQUENCES FOR HS P-CR AND 3-D POG ESTIMATION METHODS WITH $1:1$ AND $3:1$ BRIGHT TO DARK PUPIL RATIOS

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

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

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

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

Unfortunately, a similar technique cannot be used for the 3-D method to avoid the interframe motion while increasing the POG update rate. The 3-D method would require two additional corneal reflections in the bright pupil image to compute the POG with the information contained solely in a single image. The extra reflections would have to be masked off of the pupil contour, as described in Section III-B, which potentially removes large portions of the pupil contour and consequently decreases the accuracy of the pupil feature identification. The corneal reflection from the on-axis lighting in the bright pupil image cannot be used with the 3-D method as the on-axis light source is located coaxially with the focal point of the camera, which results in a singularity in the 3-D model algorithm, see [20, eq. (4)].

D. Hardware

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

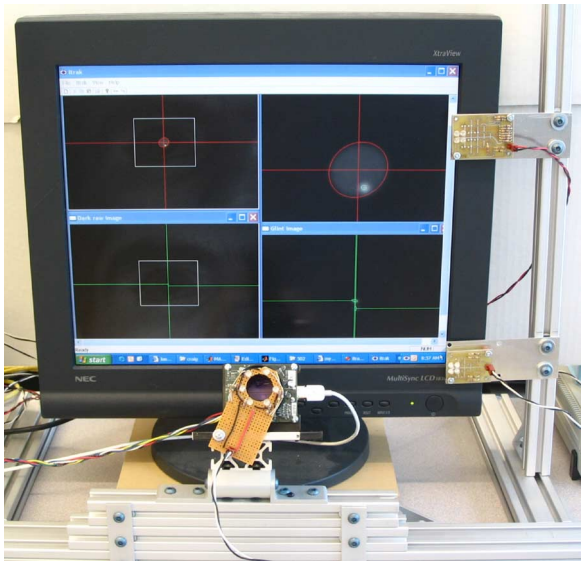


Fig. 6. Physical system showing the camera located beneath the monitor, the on-axis lighting (ring of light-emitting diodes surrounding the lens), the two off-axis point light sources located to the right of the monitor, and the monitor upon which the POG is estimated.

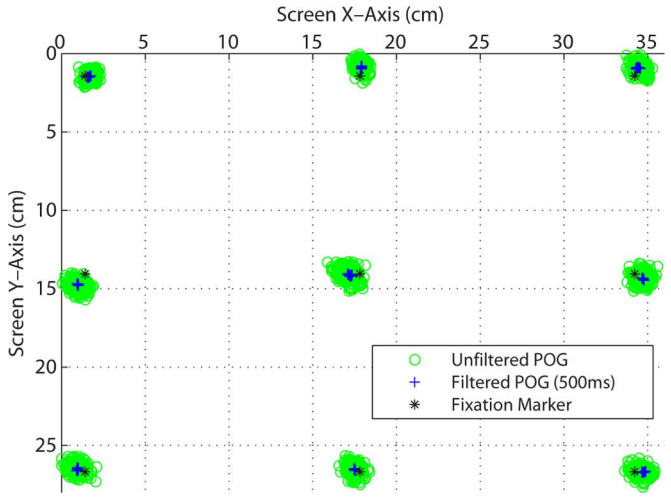


Fig. 7. Example of the fixation task in which the user observed each of the nine points on a 3 × 3 grid. In this example, the POG samples were recorded with the HS P-CR vector method and a camera frame rate of 407 Hz. The original POG data are shown along with the results of filtering with a 500-ms moving window average. The POG screen coordinates have been converted from units of pixels to centimeters in this figure.

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

The system evaluation was performed on a Pentium IV 3-GHz processor with 2 GB of random access memory. A flat-screen liquid-crystal display monitor with a width of 35.8 cm and a height of 29.0 cm was set to a resolution of 1280 × 1024 pixels and located at a distance of approximately 75 cm from the users' eye. The physical system is shown in Fig. 6.

IV. Experimental Design and Results

The techniques to perform the high-speed noncontact eye-gaze tracking described above were evaluated with the HS P-CR and 3-D model-based methods for estimating the POG. Both POG methods were tested at three different camera frame rates to determine the effect of sampling rate on fixation precision. Varying levels of digital filtering were applied to the recorded data for each POG method at each frame rate to show the resulting improvements in precision.

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

The experimental procedure was comprised of a calibration phase, a familiarization phase, and the performance of a simple task during which the POG screen coordinates were recorded. The calibration consisted of having the subject observe the four corners of the screen for approximately 1 s each while the per-user parameters were estimated. After calibration, a short familiarization period was allowed in which the calibration was evaluated with the subject verifying that the computed POG across the screen was in fact the same (or at least very close) to their real POG. The subject was then asked to fixate on nine sequential points on a 3 × 3 grid, which were displayed across the screen. Throughout the fixation task, the screen coordinates of the POG were continuously recorded along with a flag indicating the fixation status at each grid point. The fixation status flag was set to indicate the beginning of a fixation when the relative stability of a fixation was detected, and the flag was cleared when the larger motion of a saccade was detected, as per the position variance algorithm described in [1]. At least 2 s of fixation data were acquired before moving to the next point. An example of the fixation data collected on the 3 × 3 grid for a single subject is shown in Fig. 7, while a subset of ten POG estimates from a single fixation point is shown in Fig. 8.

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

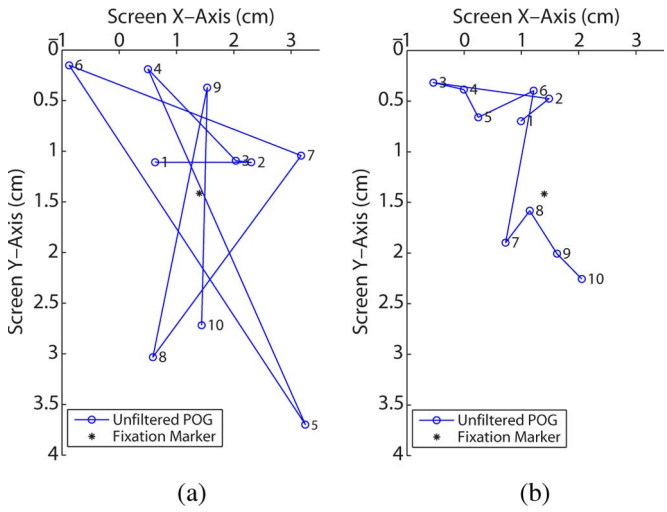


Fig. 8. Labeled sequence of ten unfiltered POG estimates for the 3-D POG estimation method shown from a single fixation marker. The sampling sequences at two camera frame rates are illustrated: (a) 30 Hz, in which the ten-point sequence corresponds to a time interval of 333 ms and (b) 407 Hz, which corresponds to a time interval of 25 ms. (a) The 3-D POG estimation method at 30 Hz. (b) The 3-D POG estimation method at 407 Hz.

TABLE II
IMAGE SEQUENCE PARAMETERS FOR THE HS P-CR POG METHOD

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

TABLE III
FILTER ORDER FOR EACH SAMPLING RATE AND FILTER LENGTH FOR THE HS P-CR AND 3-D POG ESTIMATION METHODS

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

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

After filtering the recorded X and Y POG coordinates with each of the FIR filters, the fixation precision was determined at each of the nine fixation points. The standard deviation was computed on the last 500 ms of the 2 s of data recorded at each fixation point to avoid combining data points from adjacent fixations when high filter orders are used.

TABLE IV
FIXATION PRECISION FOR EACH SYSTEM CONFIGURATION

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

Note: All units in degrees of visual angle

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

$$\theta = 2 \cdot \tan^{-1}\left(\frac{\sqrt{(p_x - r_x)^2 + (p_y - r_y)^2}}{2} \cdot \frac{1}{75}\right) \quad (3)$$

with the assumption that the average distance from eye to screen was 75 cm.

V. DISCUSSION

Using the techniques described above, the operation of the remote eye-gaze tracking system at high sampling rates was achieved. The higher sampling rates more accurately record the faster dynamics of the eye and reduce the signal aliasing. Using the Nyquist criterion, the sampling rate should be at least twice the highest frequency of the microsaccades and tremors (up to 150 Hz [12]) observed during fixations. To illustrate the effect of aliasing, a labeled sequence of POG estimates is shown with a low sampling rate (30 Hz) in Fig. 8(a) and at a much higher sampling rate (407 Hz) in Fig. 8(b). For the lower sampling rate, the details of the trajectory of the POG are missing as illustrated by the erratic and large displacements between the subsequent POG estimates. At the higher sampling rate, the trajectory of the POG estimates can more clearly be seen as the displacement between estimates is smaller.

Processing the incoming images at 200 fps was achieved with only the use of the software ROI. With the addition of the hardware ROI, the camera frame rate increased to 407 fps. Using the 3-D model-based POG estimation algorithm, the POG estimation rate was equal to the camera frame rate: 30 Hz at 30 fps, 200 Hz at 200 fps, and 407 Hz at 407 fps. When using the HS P-CR method for estimating the POG, an update rate

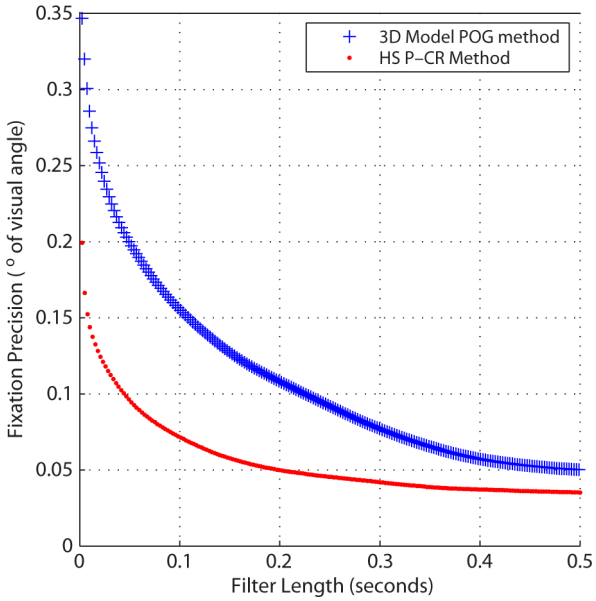


Fig. 9. Fixation precision versus filter length is shown averaged across all four subjects indicating an exponential relationship. The POG screen coordinates were recorded with the system operating at 407 fps for both the HS P-CR and the 3-D POG methods.

of only 15 Hz was achieved when operating at 30 fps due to the requirements of the image differencing technique. With the reduced interframe motion at higher frame rates, it was possible to enhance the P-CR method by increasing the ratio of bright to dark pupil images without losing lock on the eye. Increasing the bright to dark pupil ratio to 9 : 1 for the 200 fps frame rate increased the POG sampling rate to 180 Hz, and increasing the ratio to 19 : 1 at 407 fps increased the sampling rate to 386 Hz. The POG update rates achieved for the HS P-CR and 3-D methods are a significant increase over the rates achieved by similar eye-gaze tracking systems discussed in the background review of this paper.

The fixation precision reported for the 3-D model-based POG method at the lowest sampling rate (30 Hz) was 0.55°. This result is of similar magnitude to the precision reported by Yoo and Chung [21] at 0.84° for their noncontact free-head eye-gaze tracking system, which operated at a rate of 15 Hz. The benefit of our system is the ability to increase the POG sampling rate, which then allows digital filtering to further improve the fixation precision while still maintaining an acceptable latency. Using digital low-pass filtering resulted in an improvement in fixation precision in all the system configurations, as shown in Table IV. In the experiments performed, the best fixation precision was achieved with the longest filter (500 ms), which for the HS P-CR method resulted in a standard deviation of 0.035° or 1.6 screen pixels and 0.050° or 2.3 screen pixels for the 3-D model-based method. The relationship between filter length and fixation precision appears to be exponential as shown in Fig. 9. As the filter length increases, a diminishing return in the tradeoff between the achieved precision and the POG latency is achieved.

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

A comparison of the accuracy between the two methods was not performed as the focus of this paper is on fixation precision. A more detailed investigation of system accuracy is presented in [20]. While not the focus of this paper, the system accuracy was confirmed to be comparable to many contemporary remote eye-gaze tracking systems [6]. Averaged over all subjects and all operating conditions, the HS P-CR method resulted in an accuracy of 0.72°, while the 3-D method accuracy was 1.0° of visual angle. The accuracy of the HS P-CR method appears slightly better in these experiments; however, the measurements were only recorded with the head located near the calibration position and did not explicitly exercise the free head capabilities of the 3-D model-based method.

VI. Conclusions

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

Techniques that enable video-based noncontact eye-gaze tracking systems to operate at high POG sampling rates were presented, which more adequately record the dynamics of high-speed eye movements. A high-speed method for P-CR POG estimation was also presented in which the sampling rate was increased by modifying the ratio of bright pupil to dark pupil images. Increasing the frequency of bright pupil images increased the frequency of the images containing the features required to compute the POG.

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

A comparison between the HS P-CR POG estimation method and the 3-D model-based method showed that the fixation

precision for the HS P-CR method was significantly better than the 3-D method at each of three camera frame rates tested. One possible explanation for this result is that the HS P-CR POG estimation method avoided the misalignment of the image feature data resulting from interframe motion by using information from only a single image to compute the POG. Although the 3-D method is shown to be less precise, it does allow a wider range of head motion [20] than the HS P-CR method [6]. In this paper, however, subjects were asked to maintain a comfortable relatively stationary head position.

Future work will focus on the evaluation of the techniques presented in this paper on a larger sample of subjects. The integration of these methods with an eye-gaze tracking system for use outside the laboratory environment is also desirable to increase the realism of the eye-gaze tracking experiments.

REFERENCES

[1] R. Jacob, "Eye tracking in advanced interface design," in *Virtual Environments and Advanced Interface Design*. New York: Oxford Univ. Press, 1995, pp. 258–288.

[2] R. Jacob and K. Karn, "Eye tracking in human-computer interaction and usability research: Ready to deliver the promises (Section Commentary)," *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*. Amsterdam, Elsevier: The Netherlands, 2003, pp. 573–605.

[3] S. Zhai, C. Morimoto, and S. Ihde, "Manual and gaze input cascaded (magic) pointing," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 1999, pp. 246–253.

[4] K. S. Karn, S. Ellis, and C. Juliano, "The hunt for usability: Tracking eye movements," in *Proc. CHI Extended Abstracts Human Factors Comput. Syst.*, 1999, p. 173.

[5] H. Collewijn, "Eye movement recording," *Vision Research: A Practical Guide to Laboratory Methods*, New York: Oxford Univ. Press, 1999, pp. 245–285.

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

[7] J. P. Hansen, D. W. Hansen, and A. S. Johansen, "Bringing gaze-based interaction back to basics," *Universal Access in HCI*, Lawrence Erlbaum, Mahwah, NJ, 2001, pp. 325–328.

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

[9] M. Ashmore, A. T. Duchowski, and G. Shoemaker, "Efficient eye pointing with a fisheye lens," in *Proc. Conf. Graph. Interface*, 2005, pp. 203–210.

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

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

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

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

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

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

[16] T. Hutchinson, K. White, Jr., W. Martin, K. Reichert, and L. Frey, "Human–computer interaction using eye-gaze input," *IEEE Trans. Syst., Man, Cybern.*, vol. 19, no. 6, pp. 1527–1534, Nov./Dec. 1989.

[17] S.-W. Shih and J. Liu, "A novel approach to 3-D gaze tracking using stereo cameras," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 34, no. 1, pp. 234–245, Feb. 2004.

[18] T. Ohno and N. Mukawa, "A free-head, simple calibration, gaze tracking system that enables gaze-based interaction," in *Proc. Symp. ETRA*, 2004, pp. 115–122.

[19] D. Beymer and M. Flickner, "Eye gaze tracking using an active stereo head," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog.*, Jun. 18–20, 2003, vol. 2, pp. II-451–II-458.

[20] C. Hennessey, B. Noureddin, and P. Lawrence, "A single camera eye-gaze tracking system with free head motion," in *Proc. Symp. Eye Track. Res. Appl.*, 2006, pp. 87–94.

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

[22] Z. Cherif, A. Nait-Ali, J. Motsch, and M. Krebs, "An adaptive calibration of an infrared light device used for gaze tracking," in *Proc. 19th IEEE Instrum. Meas. Technol. Conf.*, May 21–23, 2002, vol. 2, pp. 1029–1033.

[23] R. Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses," *IEEE J. Robot. Autom.*, vol. RA-3, no. 4, pp. 323–344, Aug. 1987.

[24] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000.

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

[26] Y. Ebisawa and S. Satoh, "Effectiveness of pupil area detection technique using two light sources and image difference method," in *Proc. 15th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Oct. 28–31, 1993, pp. 1268–1269.

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

[28] C. Hennessey, "Eye-gaze tracking with free head motion," M.A.Sc. thesis, Univ. British Columbia, Vancouver, BC, Canada, Aug. 2005.

[29] A. Fitzgibbon, M. Pilu, and R. Fisher, "Direct least square fitting of ellipses," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 5, pp. 476–480, May 1999.

[30] J. Zhu and J. Yang, "Subpixel eye gaze tracking," in *Proc. 5th IEEE Int. Conf. Autom. Face Gesture Recog.*, 2002, pp. 124–129.

Craig Hennessey received the B.A.Sc. degree in systems engineering from Simon Fraser University, Burnaby, BC, Canada, in 2001, and the M.A.Sc. degree in electrical engineering from the University of British Columbia, Vancouver, BC, in 2005, where he is currently working toward the Ph.D. degree in electrical engineering.

His research interests include computer vision, image processing, and human–computer interaction.

Borna Noureddin received the B.Eng. degree in computer engineering from the University of Victoria, Victoria, BC, Canada, and the M.A.Sc. degree in electrical engineering from the University of British Columbia, Vancouver, BC, where he is currently working toward the Ph.D. degree.

His research interests include computer vision, biomedical signal and image processing, biophysical modeling, and human–computer interaction.

Peter Lawrence (S'64–M'73–SM'06) received the B.A.Sc. degree in electrical engineering from the University of Toronto, Toronto, ON, Canada, in 1965, the M.S. degree in biomedical engineering from the University of Saskatchewan, Saskatoon, SK, Canada, in 1967, and the Ph.D. degree in computing and information science from Case Western Reserve University, Cleveland, OH, in 1970.

Between 1970 and 1972, he was a Guest Researcher with the Applied Electronics Department, Chalmers University, Goteborg, Sweden. Between 1972 and 1974, he was a Research Staff Member and a Lecturer with the Mechanical Engineering Department, Massachusetts Institute of Technology, Cambridge. Since 1974, he has been with the University of British Columbia, Vancouver, BC, Canada, where he is currently a Professor in the Department of Electrical and Computer Engineering. His main research interests include the application of real-time computing in the control interface between humans and machines, image processing, and mobile hydraulic machine modeling and control.