

Improving the Accuracy and Reliability of Remote System-Calibration-Free Eye-Gaze Tracking

Craig A. Hennessey* and Peter D. Lawrence, *Senior Member, IEEE*

Abstract—Remote eye-gaze tracking provides a means for non-intrusive tracking of the point-of-gaze (POG) of a user. For application as a user interface for the disabled, a remote system that is noncontact, reliable, and permits head motion is very desirable. The system-calibration-free pupil-corneal reflection (P-CR) vector technique for POG estimation is a popular method due to its simplicity, however, accuracy has been shown to be degraded with head displacement. Model-based POG-estimation methods were developed, which improve system accuracy during head displacement, however, these methods require complex system calibration in addition to user calibration. In this paper, the use of multiple corneal reflections and point-pattern matching allows for a scaling correction of the P-CR vector for head displacements as well as an improvement in system robustness to corneal reflection distortion, leading to improved POG-estimation accuracy. To demonstrate the improvement in performance, the enhanced multiple corneal reflection P-CR method is compared to the monocular and binocular accuracy of the traditional single corneal reflection P-CR method, and a model-based method of POG estimation for various head displacements.

Index Terms—Binocular, eye-gaze, multiple corneal reflection, pupil-corneal reflection (P-CR), remote, single camera, system-calibration free, tracking.

I. INTRODUCTION

YE-GAZE tracking can be used as a human-machine interface technique for individuals with high-level spinal-cord injuries or motor-neuron disorders, who are unable to operate standard interface tools such as the keyboard and mouse [1]. While video-based eye-gaze tracking has great potential for improving the quality of life of these individuals, a number of key technical issues need to be improved upon. While the requirements for remote eye-gaze tracking are application dependent, in general, improvements are needed in accuracy, precision, response time, reliability, cost, ability to tolerate head motion, and simplification of system and user calibration requirements [2]. Reducing the need for system calibration greatly simplifies the system design, potentially lowering costs, while also simplifying the initial user setup of the system. The objective of this paper is to describe the methodology and experimental results showing that tracking a pattern of multiple corneal reflections can be used to improve the point-of-gaze (POG) accuracy for

the system-calibration-free pupil-corneal reflection (P-CR) vector technique for POG estimation.

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

1) *Head Mounted*: Head-mounted eye-gaze trackers typically use the P-CR vector method for POG estimation [3]. The accuracy of the P-CR method degrades as the head is displaced from the calibration position especially in depth [2]. Head-mounted systems offer the benefit of fixed head-to-camera displacement, however the system must be firmly mounted to the head to prevent slippage. Head-mounted systems, if used over an extended period of time, can result in fatigue and comfort can be a concern [4], [5].

The single corneal reflection used in the traditional P-CR method can be problematic as larger eye rotations can result in distortion of the reflection when it nears the boundary between the cornea and sclera resulting in increased error. At even larger eye rotations the corneal reflection may be lost entirely, resulting in a complete failure of the system to estimate the POG. To develop a more reliable P-CR image feature, Hua *et al.* [6] recently proposed a technique for head-mounted POG estimation that used a symmetric arrangement of four LEDs to generate a cross-shaped pattern of corneal reflections. A virtual point located at the intersection of the horizontal and vertical lines connecting the matching pairs of reflections was then used in forming the P-CR vector. To compensate for the loss of reflections, the two pairs of LEDs must be placed orthogonally with respect to each other (i.e., vertically or horizontally) and parallel to the camera image plane.

2) *Remote*: Remote eye-gaze tracking offers greater comfort and ease of use as the user is not required to wear head-mounted equipment. Early remote eye-gaze tracking systems using the P-CR technique however required a relatively motionless head using bite bars or chin rests to prevent head to camera displacements. A recent attempt by Cerrolaza *et al.* to overcome this limitation showed promise by tracking the relative displacement of two corneal reflections and normalizing the P-CR vector accordingly [7]. While a chin rest was still used to stabilize the head, it was observed that the normalized P-CR vector performed better than the traditional P-CR vector when the head is displaced with depth. As will be shown later in this paper, unconstrained head and eye movements can lead to the loss and distortion of the corneal reflections and therefore result in inaccurate POG estimation. Using multiple corneal reflections, along with a means to differentiate the reflections, will be shown to compensate for corneal reflection distortion and loss.

To allow for natural head motion in remote eye-gaze tracking, more complex techniques for POG estimation have been

Manuscript received April 21, 2008; revised January 10, 2009. First published March 4, 2009; current version published June 12, 2009. Asterisk indicates corresponding author.

*C. A. Hennessey is with the Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC V6T 1Z4, Canada (e-mail: craigah@ece.ubc.ca).

P. D. Lawrence is with the Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC V6T 1Z4, Canada.

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/TBME.2009.2015955

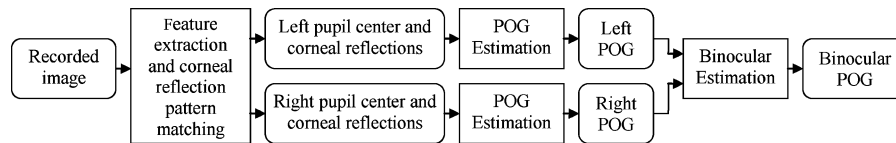


Fig. 1. High-level binocular eye-gaze tracking system block diagram is shown. The final POG may be estimated from either the left or right eye, increasing reliability to the loss of an eye due to head motion. Alternatively, the final POG can be estimated as the average of the POG estimates from the left and right eyes providing a more accurate estimate of the true POG.

developed based on models of the eye, camera, and physical system. An early remote system developed by Shih and Liu [8] tracked both eyes using two remote cameras imaging at 30 Hz and mounted close to the subject's eyes. System calibration included stereo camera lens calibration [9], physical system modeling of the computer screen, LEDs, and camera positions, and per-user calibration to approximate parameters of the eyes. While only two corneal reflections were required for estimation of the POG, three reflections were used to provide redundancy should one be lost due to eye rotation. An average accuracy over six subjects of slightly better than 1° of visual angle was reported. For this system however only a small degree of head motion (4×4 cm with little depth motion) was possible due to the proximity of the cameras to the eyes and the limited depth of field of the lens.

In the system by Ohno *et al.*, two cameras and a pan/tilt/zoom mechanism were used to increase the allowable range of head motion while achieving similar accuracy results to Shih *et al.* A wide-angle lens camera was used to direct the narrow angle lens pan/tilt/zoom camera to track the eye, however, the mechanical tracking mechanism was too slow to keep up with fast head motions [10]. High-speed galvanometer mechanisms were investigated by Beymer *et al.* for providing fast mechanical tracking, however, the tracking mechanism was significantly more complicated. The system required a complex system calibration, and the use of two pairs of stereo cameras led to a low system update rate of 10 Hz. [11]

For remote systems, typically only a single eye is tracked due to the limited field of view of the camera. Tracking a single eye is in general sufficient as both eyes tend to point to the same position [12]. If both the left- and right-eye POG estimates are known however, averaging of the two can potentially reduce the overall error, as was observed by Cui *et al.* for head-mounted eye-gaze tracking [13]. Tracking both eyes also allows the system to continue operating when a single eye is lost due to head motion.

3) *Present Work*: In the work presented here, a novel technique for tracking a pattern of corneal reflections is presented to enhance the performance of the system-calibration-free P-CR POG-estimation method. Tracking the corneal reflection pattern improves the reliability of POG estimation by detecting the loss and distortion of corneal reflections when both head and eye rotations cause the reflections to move off the surface of the cornea. The tracking technique presented here has fewer restrictions on the placement of the light sources than the method by Hua *et al.*, as well as providing a mechanism for detecting distortion of the reflections and not just the complete loss. Tracking the scale and translation parameters of the corneal reflection

pattern will be shown to enhance the performance of the P-CR method for operation in remote, system-calibration-free, eye-gaze tracking to match that of the more complex model-based method which requires considerable system calibration. As the system provides binocular estimates, a comparison of the binocular and monocular performance will also be provided.

II. METHODS

A high-level overview of the proposed system is outlined in Fig. 1. In this system a single camera is used to record images of the face in which both left- and right-eye tracking is attempted. The identified image features are labeled as coming from either the left or right eye and are then used in the POG-estimation algorithm. The image processing and POG-estimation stages are described in greater detail in the following subsections.

A. Image Feature Processing

The P-CR and model-based POG-estimation algorithms require accurately identified pupil and corneal reflection image features. The purpose of the image processing stage of the eye-gaze tracking system is to extract these image features accurately and rapidly. A novel corneal reflection identification algorithm is then applied to the extracted corneal reflection image features to check for possible loss or distortion of the reflections.

1) *Feature Extraction*: The eye image features required from the recorded images are the centers of the pupils and the locations of the corneal reflections. IR light is used for system illumination to enhance the performance of the feature extraction, using the bright-pupil and image-difference techniques [14], [15]. Using IR light generates the necessary reflections off the cornea, as well as reducing the sensitivity of the system to ambient lighting conditions. A ring of on-axis lights generate the high contrast pupil in the bright-pupil image, while a collection of off-axis point light sources are used to generate an image with a dark pupil as well as multiple corneal reflections. An ellipse is fit to each image feature contour with the contour center location then identified as the center of the ellipse [16]. The image feature extraction procedure is described in greater detail in Hennessey *et al.* [17]. 中心替代椭圆

If the image feature extraction algorithm identifies a pupil and corresponding pattern of corneal reflections in the image, the first identified eye image is blanked out and a second image feature search is performed for the second eye. When both eyes are visible, the extracted image features can easily be associated with either the left or right eye based on their relative displacements in the image. When only one eye is visible to the camera some form of eye and face tracking is required

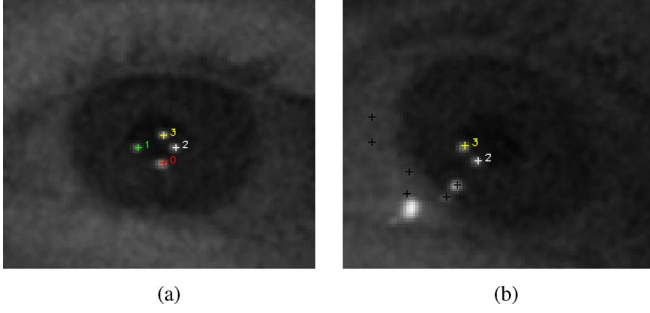


Fig. 2. (a) Set of four valid corneal reflections labeled corresponding to the generating IR light sources. (b) Same set of four light sources are used to generate the pattern in which the eye has been rotated resulting in corruption of the reflection pattern. The corrupted pattern illustrates the loss of one of the valid corneal reflections [reflection 1 from Fig. 2(a)], the distortion of another [reflection 0 from Fig. 2(a)] as well as the addition of spurious reflection off the surface of the sclera. (a) Valid corneal reflections. (b) Invalid corneal reflections.

to identify which eye remains visible. A wide variety of eye- and face-tracking techniques have been developed including Kalman filtering [18], support vector machine [19], and eigen-features [11]. Since system lighting is already controlled, background segmentation [20] is used here.

2) **Corneal Reflection Pattern Matching:** In a new approach to corneal reflection tracking, the off-axis light sources are used to generate a pattern of corneal reflections in the dark-pupil image. The corneal reflection pattern can then be used to enhance the performance of the POG-estimation techniques. For the P-CR POG-estimation method, a single corneal reflection is required. For the model-based method two corneal reflections are required for triangulation of the 3-D center of the cornea. Using three or more light sources to generate multiple corneal reflections can be used to provide redundancy should any reflection be corrupted or lost. Distortion or loss of corneal reflections occurs when the reflection occurs near the boundary between the cornea and the sclera or on the sclera itself. The distortion of the reflections are due to the different radius of curvature between the sclera and the cornea, while the rougher surface of the sclera can cause valid image reflections to disappear or spurious reflections to appear. An image reflection is defined as valid if it is correctly matched with a reference reflection and the corresponding source of the IR light. A valid pattern of four corneal reflections are shown in Fig. 2(a) while the same pattern is shown corrupted due to eye rotation in Fig. 2(b).

Using multiple corneal reflections requires a means for distinguishing the corneal reflection image points from one another, as the POG-estimation methods require the correspondence between the light source and the generated reflection. Many general techniques for point-pattern matching have been developed. For a literature survey see Cox and Jager [21]. The corneal reflection point-pattern matching technique described here is based on inter-point distances and is customized for corneal reflection detection. The algorithm compensates for translation, distortion, addition, and deletion of corneal reflections. For proper operation, the IR point light sources must be placed such that at least two valid reflections off of the surface of the cornea will always be visible to the camera, as a single reflection is insufficient

Algorithm 1 Corneal reflection pattern matching

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

```

1:  $dist_{min} = inf$ 
2: // Find index for closest image point  $Q_\alpha$  to center of pupil
3:  $\alpha = \arg \min_i \|Q_i - P_{center}\|$ 
4: // Identify corresponding image and reference points
5: for  $j = 1..N$  do
6:   // Translation from image to reference
7:    $T_j = R_j - Q_\alpha$ 
8:   for  $i = 1..M$ ,  $i \neq \alpha$  do
9:     for  $k = 1..N$ ,  $k \neq j$  do
10:      // Dist. from each image pt. to reference pt.
11:       $d_k = \|(T_j + Q_i) - R_k\|$ 
12:      // Label  $Q_\alpha$  at minimum overall dist.
13:      if  $d_k < dist_{min}$  then
14:         $dist_{min} = d_k$ 
15:        Label  $Q_\alpha$  as  $R_j$ 
16:      end if
17:    end for
18:     $\beta = \arg \min_k \{d_k\}$ 
19:    // Label  $Q_i$  if minimum dist. is under threshold
20:    if  $d_\beta \leq thresh$  then
21:      Label  $Q_i$  as  $R_\beta$ 
22:    end if
23:  end for
24: end for

```

for the matching procedure. In addition, unique displacements between all pairs of reflections are required in order to provide a means for matching the valid reflections with the corresponding IR point light sources.

To perform the matching operation a reference pattern is required in which the valid corneal reflections are identified and associated with their corresponding IR point light sources as shown in Fig. 2(a). The reference pattern is created by initially recording a valid pattern of image reflections formed on each of the eyes and manually identifying the corresponding corneal reflections and IR light sources. Subsequent system operation extracts the coordinates of the corneal reflection image points (Q_i) and searches for matching pairwise displacements within the reference pattern points (R_j). A match is identified if a displacement is found under a certain tunable threshold value. This tunable threshold may be set to allow for slight distortions in the corneal reflection pattern, while corneal reflections resulting in larger distortions of the pattern are rejected. Once an appropriate threshold level has been determined, either experimentally or based on the maximum distortion requirements of the POG-estimation algorithms, the parameter may be fixed and need no longer be adjusted. For the system presented here a distortion threshold of five pixels was used successfully for all subjects.

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

采用最接近pupil center的点
作为初始匹配点 畸变可能最小

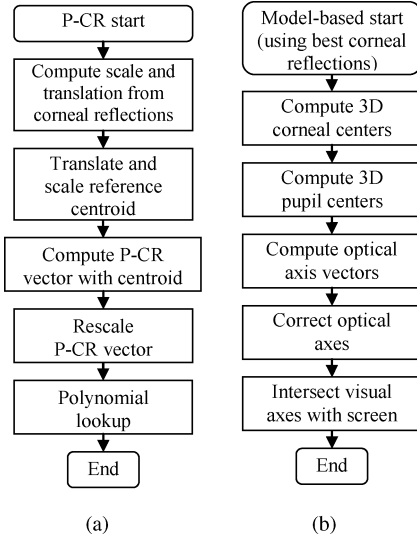


Fig. 3. POG estimation. (a) POG-estimation processes are shown for the enhanced P-CR method. (b) Processes for the model-based method. The enhanced P-CR method integrates the corneal reflection tracking for centroid estimation with the rescaling of the P-CR vector to compensate for head motion. For the model-based method the “best corneal reflections” are the two reflections located nearest to the center of the pupil, as these are least likely to have been distorted. (a) Enhanced P-CR Method. (b) Model-Based Method.

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

B. POG Estimation

The two main techniques used for POG estimation in remote eye-gaze tracking are the P-CR and model-based methods. The traditional P-CR and model-based methods have been enhanced here to take advantage of the multiple redundant corneal reflections to enhance both reliability and the ability to handle head motion. Each POG-estimation algorithm is outlined in Fig. 3 and will be described in the following subsections.

1) *Enhanced P-CR Vector*: Traditionally the P-CR vector ($V = (v_x, v_y)$) is formed in the recorded bright-pupil image of the eye from the on-axis corneal reflection to the center of the pupil. Through a user-calibration procedure in which the subject observes known points on the screen, the P-CR vector is mapped to the POG ($U = (u_x, u_y)$) on the screen in pixels. The mapping typically uses a second-order polynomial (1) where the parameters a_i and b_i are determined during calibration [2].

$$\begin{aligned} u_x &= a_0 + a_1 v_x + a_2 v_y + a_3 v_x v_y + a_4 v_x^2 + a_5 v_y^2 \\ u_y &= b_0 + b_1 v_x + b_2 v_y + b_3 v_x v_y + b_4 v_x^2 + b_5 v_y^2. \end{aligned} \quad (1)$$

Using a single corneal reflection to form the P-CR vector can be problematic, as the reflection may be distorted or lost during large eye rotations. To avoid the problem of lost or distorted

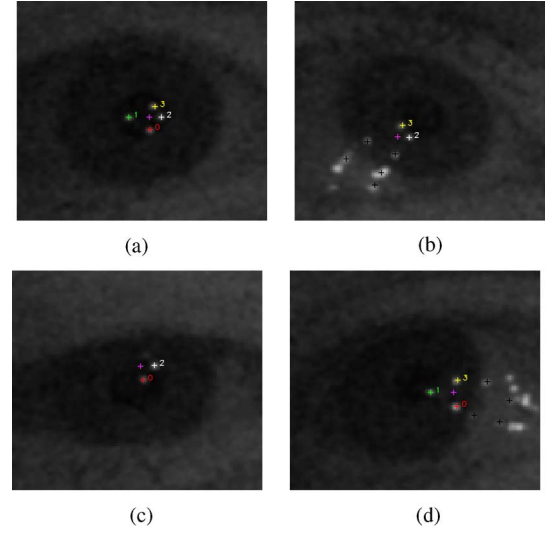


Fig. 4. In the figures shown, the centroid maintains its position relative to the corneal reflection image points regardless of the loss or distortion of the corneal reflections making up the pattern. In (b) through (d) the centroid was correctly determined while up to two corneal reflection points of the pattern were lost. (a) Reference pattern. (b) Missing reflections. (c) Missing reflections. (d) Missing reflections.

reflections, the P-CR vector is instead formed from the centroid of the corneal reflection pattern, based on the valid corneal reflections identified by Algorithm 1.

The centroid R_c of the 2-D corneal reflection reference pattern is determined once based on the N reference positions $R_j = (r_{jx}, r_{jy})$ using (2).

$$R_c = \frac{1}{N} \sum_{j=1, \dots, N} R_j. \quad (2)$$

The POG is then determined in each system loop as shown in the flowchart of Fig. 3(a). A set of M corneal reflection points $Q_i = (q_{ix}, q_{iy})$, are first extracted from the recorded images and matched with the reference points R_j using Algorithm 1. For each correctly matched point Q_i to R_j an equation is formed in which only the scale (s) and translation ($T = (t_x, t_y)$) parameters are unknown (3).

$$R_j = s \cdot Q_i + T. \quad (3)$$

For each iteration of the POG-estimation algorithm, provided two or more matching pairs Q_i and R_j are identified, an overdetermined set of equations for s , t_x and t_y are formed which are easily solved using a least-squares approach.

To compute a robust estimate of the corneal reflection pattern centroid Q_c at run-time, the original reference pattern centroid R_c is scaled and translated according to the determined scale and translation factors (4). Examples of the resulting centroid at different eye rotations are shown in Fig. 4.

$$Q_c = \frac{1}{s} \cdot (R_c - T) \quad (4)$$

The P-CR vector may then be formed from the corneal reflection pattern centroid Q_c to the pupil center P_{center} using (5). The resulting vector is robust to loss and distortion of the

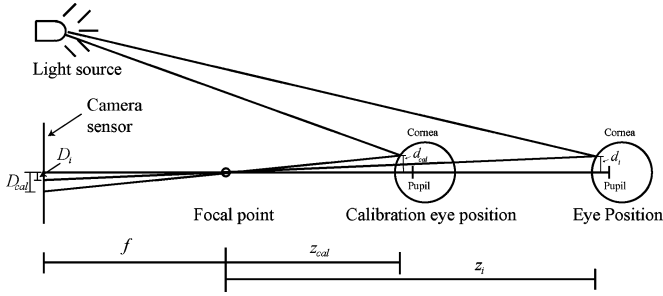


Fig. 5. Simplified schematic (two dimensional and not to scale) of the eye located at two depths with the pupil looking along the x -axis, illustrating the change in the P-CR image vector with head displacements. The P-CR vectors d_i are projected from distances z_i , through the focal point of the camera lens with focal length f , onto the camera surface forming the P-CR image vectors D_i .

corneal reflections, which otherwise would have distorted the P-CR vector V had it been formed with a single Q_i point alone.

$$V = P_{\text{center}} - Q_c. \quad (5)$$

One additional problem with the P-CR vector is that head displacements in z from the calibration z -plane leads to position degradation of POG-estimation accuracy. Translation of the head in depth results in a change in magnitude of the P-CR image vector, even though the subject's POG on the screen may have remained constant. A simplified 2-D system with a single light source is shown in Fig. 5 to illustrate this problem. The light rays follow the pin-hole lens equation (6) with the projection d_i of the P-CR vector in the real world at distance z_i imaged through the pin-hole lens with focal length f onto the camera surface with image size D_i .

$$\frac{D_i}{f} = \frac{d_i}{z_i}. \quad (6)$$

照片大小
和焦距

In the 2-D example shown $d_i \cong d_{\text{cal}}$, due to the tangential reflection of rays off the surface of the cornea and the much larger values of z_i than the radius of the spherical cornea surface which leads to (7).

$$D_{\text{cal}} = \frac{z_i}{z_{\text{cal}}} D_i. \quad (7)$$

Since each of the corneal reflections in a corneal reflection pattern behaves as in (7), in order to compensate for the degradation in POG accuracy of the P-CR technique due to head motion in depth, the P-CR vector is scaled back to the calibration position using the ratio of the size of the corneal reflection pattern in place of the ratio of z_i to z_{cal} in (7). When calibration of the system first takes place, the average scale s_{cal} of the corneal reflection patterns is recorded. The ratio of the current scale s to the calibration scale s_{cal} is then used to rescale all future P-CR vectors. The P-CR vector equation using the corneal reflection pattern centroid in (5) is improved further with the addition of rescaling in (8).

$$V' = \frac{s}{s_{\text{cal}}} (P_{\text{center}} - Q_c) \quad (8)$$

Scal 修正
时的均值
目前值

2) *Model-Based*: The second method of estimating the POG is based on 3-D models of the eye and system as shown in Fig. 3(b). The model-based method for POG estimation was designed to inherently compensate for motion of the head, at the expense of an increasingly complex algorithms and system calibration. The model-based method requires 3-D models of the camera and lens, computer screen, and eye. In addition to the pupil image center, two corneal reflection images are also required to estimate the POG. The details of the model-based POG-estimation procedure can be found in Hennessey *et al.* [22].

The accuracy of the model-based method for POG estimation allowing free head motion will be compared with the accuracy of the enhanced P-CR method in the following section. The POG-estimation methods use image features extracted from the same images in the comparison. The model-based method requires exactly two corneal reflections however, and since multiple corneal reflections are found in the images a selection of two of the reflections must be made. To ensure that the most reliable image information is used in the model-based POG-estimation method, only the corneal reflections least likely to be distorted are used. Since the distortion and loss of reflections occurs as the points approach the boundary of the cornea and sclera, the two reflections located closest to the center of the pupil are chosen for the model-based method.

3) *Binocular Estimation*: The POG-estimation methods can be performed independently for the left and right eyes resulting in two POG estimates, one for each eye. The use of binocular averaging of the left- and right-eye POG estimates has the potential for improving overall accuracy [13]. As healthy eyes generally observe the same point in space, the left- and right-eye POG estimates should be located at the same position. Approximations used in the POG-estimation algorithms however may lead to symmetric errors for the left and right eyes. Averaging of the two estimates may then result in improving overall accuracy for both the P-CR and model-based systems by cancelling some of the error as will be shown in the following experiments. Additionally, in the event that one eye is located outside the field of view of the camera, or the POG for one eye is unable to be computed due to corrupt image features, the results of POG estimation for the remaining eye may be used.

III. EXPERIMENTAL METHODS AND RESULTS

A. Experimental Hardware

The evaluation of the proposed methods was performed on the eye-gaze tracking system shown in Fig. 6. A single dragonfly express camera from point grey research with a fixed focal length lens was located below the computer screen and used to record images of the face and eyes. The single camera used had a sensor resolution of 640×480 pixels which streamed video over the firewire 1394b data bus at 200 frames per second. An IR ring surrounding the camera lens was used to generate the on-axis lighting for the bright pupil. The off-axis light sources were comprised of four clusters of seven, 880 nm, IR LEDs located around the computer screen. A microcontroller was used to synchronize the camera shutter with the on and off-axis LED

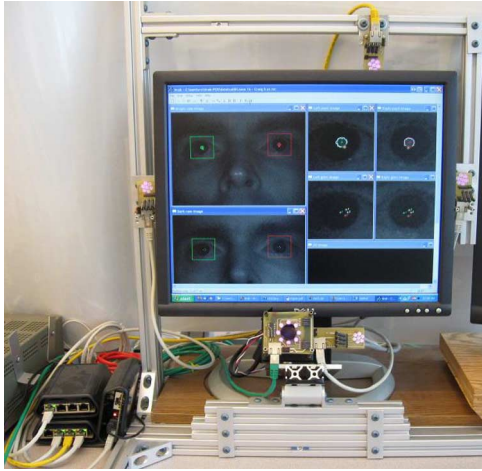


Fig. 6. Eye-gaze tracking system. The camera is located below the computer screen, with the camera lens surrounded by the ring of on-axis lighting. There are four off-axis point light sources located around the computer screen. The microcontroller used for synchronization of the on and off-axis lighting with the camera shutter is located in the lower left portion of the image.

TABLE I
PROCESSING TIMES

Activity	Processing Time (ms)
Time budget per frame	5.0
Image feature extraction	1.65
Corneal reflection pattern matching	0.022
P-CR POG estimation	0.15
Model-based POG estimation	0.30
Data logging & display	0.38
Processing time used	2.5

lighting. Extruded aluminum rails formed a mechanical mounting structure to which the IR point light sources were attached and the displacements between the lights, camera, and screen were fixed. For the model-based POG-estimation method, the camera lens was calibrated with the MATLAB camera calibration toolbox [23], while the physical locations of the camera, screen and LEDs were measured manually. For the P-CR method no system calibration was required.

The computer screen was a 17" liquid crystal display (LCD) with a resolution of 1280×1024 pixels and dimensions of 38 cm in width and 30 cm in height. The computer had a 2.66 GHz Intel Core 2 processor and 2 GB RAM. All algorithms were implemented in C++ with the open source OpenCV [24] computer vision libraries used for image processing. Given the processing power available, the software developed was capable of processing the single camera video stream at full frame rates. The average execution time required for each processing stage was recorded as listed in Table I. The recorded times were averaged over 1 s of operation and measured when both eyes were visible to the camera.

With the high-speed sampling rate of 200 Hz, filtering was used to smooth out noise from the system and the inherently jittery eye motions [1], [17]. A rectangular FIR low-pass filter (moving window average) with a filter order of seven samples, or 35 ms, was used to smooth the POG estimates.

TABLE II
CALIBRATION AND EVALUATION POSITIONS

Position	Calibration (pixels)		Evaluation (pixels)	
	X	Y	X	Y
1	67	0	0	0
2	703	67	636	0
3	1270	67	1273	0
4	67	438	0	508
5	568	438	636	508
6	1203	572	1273	508
7	0	943	0	1017
8	568	943	636	1017
9	1203	1010	1273	1017

B. Evaluation of Reliability and Accuracy

1) *Methods*: The performance of the proposed algorithms were evaluated with an experiment comprised of seven different subjects. The subjects included six males and one female, with ages ranging from 24 to 31 years. Two subjects wore contact lenses while the remaining had uncorrected vision. The ethnicity of the subjects was three Caucasian, one Hispanic, two Middle Eastern, and one Indian.

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

The experimental procedure had each test subject begin with a nine-point user calibration at the midpoint of the depth of focus of the camera lens, approximately 62 cm from the screen. After calibration, each subject was asked to move his/her head toward the screen until just before the image features became too blurred to properly track the eyes, due to the limited depth of focus. At this point, the extracted image features and the POG using each POG-estimation method was recorded at each point on a 3×3 grid across the screen. The nine-point data collection procedure was repeated with the head located back at the middle of the depth of focus (roughly the original calibration distance) and again with the head as far back as possible before the image features again became out of focus. At each calibration or evaluation position the subject orally indicated that they were observing the known position, at which point a marker was set in the recorded data stream and the test point moved to the subsequent position. None of the screen positions used for the evaluation coincided with any of the screen positions used for the calibration. The screen positions used for calibration and evaluation are listed in Table II.

2) *Results*: The number of times each of the on-axis or off-axis corneal reflections were lost at each of the nine points, at each of three depths, for the seven subjects was determined from the recorded image feature data. The number of times that fewer than two valid off-axis corneal reflections were available, resulting in an inability to estimate the centroid, was also

TABLE III
CORNEAL REFLECTION LOSS FOR EACH EYE AS A PERCENTAGE OF TOTAL
POSSIBLE AT THREE HEAD DEPTHS

Corneal Reflection	Close (%)		Middle (%)		Far (%)	
	L	R	L	R	L	R
Off-axis (0)	6	27	2	5	5	16
Off-axis (1)	8	2	3	0	11	2
Off-axis (2)	10	6	10	2	6	6
Off-axis (3)	14	16	2	5	5	8
On-axis	0	3	0	2	0	0
Centroid loss	0	0	0	0	0	0

TABLE IV
AVERAGE POG-ESTIMATION ERROR AT THREE HEAD DEPTHS

Method	Average Error (cm)			Standard Dev. (cm)
	Left	Right	Binocular	Binocular
Close Position ($d = 58cm$)				
Trad. P-CR	2.92	3.11	2.84	1.89
Enha. P-CR	1.35	1.67	1.02	0.53
Model-Based	1.18	1.25	1.01	0.63
Middle Position ($d = 62cm$)				
Trad. P-CR	1.62	1.64	1.34	0.77
Enha. P-CR	1.26	1.37	0.85	0.44
Model-Based	1.18	1.16	0.87	0.58
Far Position ($d = 64cm$)				
Trad. P-CR	3.07	2.35	2.48	2.30
Enha. P-CR	1.97	1.54	1.34	2.10
Model-Based	1.33	1.23	1.00	0.62

* Convert to $^\circ$ visual angle using ($e = \text{error}$): $\theta = 2 \tan^{-1} \left(\frac{e}{d} \right)$.

determined. The percentage of lost corneal reflections compared with the percentage of lost centroid positions (which required at least two valid reflections) was determined for each of the subject's left and right eyes and summarized in Table III, where the off-axis corneal reflections are identified as labeled in Fig. 4(a). The 3-D positions of the eyes were determined from the model-based method for POG estimation, which also provides estimates for the 3-D position of the center of the cornea of each eye. The average eye depth over the seven subjects for the close position was 58 cm, for the middle position was 62 cm, and for the far position was 66 cm.

At each test position at each depth the POG was estimated for both the left and right eyes using each of the three POG-estimation algorithms, traditional P-CR, enhanced P-CR and model-based. Calculating the POG with each POG-estimation algorithm at the same time on the same recorded images allowed for a direct comparison between the performance of the different methods. For each test position, the accuracy was computed as the distance between the evaluation marker and the output of the FIR smoothing filter. The error averaged over the seven subjects is shown in Table IV. In addition to the average POG error from the left and right eyes, the binocular POG error is also shown. The POG accuracy can be converted from centimeters to degrees of visual angle given the depths of the eyes. For the enhanced P-CR method an accuracy of 0.85 cm was achieved at the middle position for the binocular average of the left- and right-eye POG, which would then correspond to a visual angle accuracy of 0.79° . At each head depth the average scaling factor used to correct the P-CR vector was also recorded. At the close position (58 cm) an average corneal reflection pattern scale of 1.12 was observed, at the middle position (62 cm) an average of

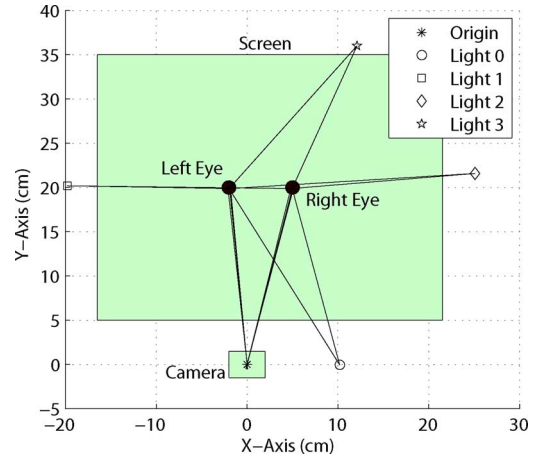


Fig. 7. System model used for simulation is comprised of four light sources, a single camera, and the computer screen. The world coordinate origin is located at the focal point of the pin-hole camera, with the positive x -axis to the right, the positive y -axis up, and the positive z -axis toward the user. Rays are shown cast from the IR light sources, reflected off the spherical cornea of the left and right eyes, through the focal point of the lens and onto the surface of the camera sensor.

0.99 was observed while at the far position (66 cm) an average of 0.90 was observed.

C. System Model Simulation

1) *Methods*: To verify the observed scaling behavior of the P-CR vector and the resulting performance of the enhanced P-CR rescaling technique proposed, a model of the eye-gaze tracking system was developed. The model used the MATLAB-based open-source software framework for simulating eye trackers developed by Böhme *et al.* [25]. Using this framework, a model of the real system shown in Fig. 6 was developed consisting of a single camera and four light sources.

To determine the corneal reflection image locations the cornea of the eye is modeled as a spherical surface, off which light rays cast from the IR point light sources are reflected. The reflections off the cornea are traced through the focal point of the pin-hole camera model and onto the surface of the camera sensor as shown in Fig. 7. Given the pixel size and resolution of the camera sensor, the simulated camera image locations of the corneal reflections can be determined. To generate the simulated pupil center image, the 3-D center of the pupil is ray-traced out of the eye, accounting for refraction at the surface of the cornea, through the pin-hole camera focal point and onto the camera sensor surface. The algorithms outlined in Section II were then applied to form the P-CR vector used in calibration and evaluation of the system.

The nine-point calibration procedure took place at the 62 cm head depth, and used the same calibration screen positions as listed in Table II. The simulated eyes were then located from 40 to 120 cm in depth at 1 cm intervals and both the average POG-estimation accuracy across the nine evaluation points, as well as the scale of the corneal reflection patterns were determined.

2) *Results*: The average accuracy of the simulation and the scale of the corneal reflection pattern, normalized with the

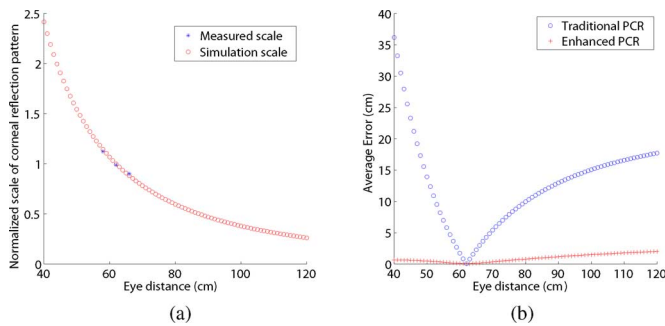


Fig. 8. (a) Simulator was used to determine the change in scale of the corneal reflection pattern as the head was displaced in depth from 40 to 120 cm at 1 cm intervals, normalized at 62 cm by s_{cal} as in (8). Also shown are the average corneal reflection pattern scale values from the multiple subject experiment at the three head depths evaluated. (b) Average error for the traditional versus enhanced P-CR POG-estimation methods for the simulated eye-gaze tracking system at 1 cm intervals from 40 to 120 cm. (a) Scale factor of P-CR vector. (b) Head displacement average error.

TABLE V
AVERAGE POG-ESTIMATION ERROR AT THREE HEAD DEPTHS
FROM SIMULATION

Method	Average Error (cm)		
	Left	Right	Binocular
Close Position (58 cm)			
Trad. P-CR	3.38	3.14	3.26
Enha. P-CR	0.29	0.53	0.41
Middle Position (62 cm)			
Trad. P-CR	0.51	0.58	0.54
Enha. P-CR	0.51	0.58	0.54
Far Position (64 cm)			
Trad. P-CR	3.64	3.59	3.61
Enha. P-CR	0.74	0.66	0.70

calibration position scale s_{cal} , over the head positions evaluated are shown in Fig. 8.

For comparison, the scale of the simulated corneal reflection patterns at each of the three real-world experiment head depths evaluated were determined to be 1.15 at the close position, 1.0 at the middle position (the same position as the calibration position), and 0.88 at the far position. The simulation average error at the three head depths for the left, right, and binocular averaged eyes are shown in Table V.

IV. DISCUSSION

When the eye is rotated to view different points on the screen, the corneal reflections translate across the surface of the cornea. In remote eye-gaze tracking, translation of the head also results in corneal reflection translations, unlike head-mounted systems where the head to camera displacements are fixed. **At certain orientations of the head and eye, the corneal reflections may be blocked by eyelashes, distorted on the boundary between the cornea and sclera, or lost on the rougher surface of the sclera due to diffuse reflection.** An experiment was performed in which the corneal reflections were tracked over 189 different head and eye positions as listed in Table III. As seen from the table, over the seven subjects were tested, if only a single on-axis or off-axis corneal reflection was used to form the P-CR vector the system would have been unable to determine the POG up to 3% or 27% of the time, respectively. Using the centroid of the

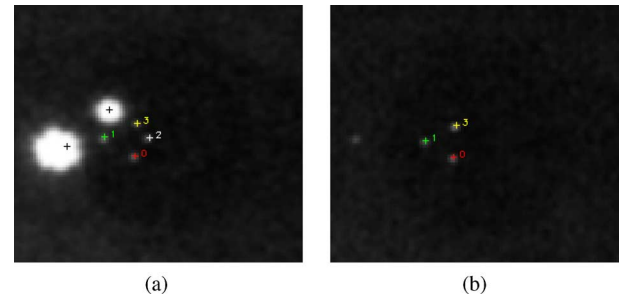


Fig. 9. When eye glasses are worn the IR light sources may reflect off the surfaces of the lenses. (a) Additional reflections can be seen off both the front and back surface of the eye-glass lens, however, the pattern matching algorithm is able to reject these additional reflections and correctly identify the corneal reflection pattern. (b) Eye-glass lens reflections were positioned to obscure the corneal reflection pattern, at which point the offending light source was turned OFF and the remaining corneal reflections identified through pattern matching. (a) Eye-glass Reflections. (b) Reflections Avoided.

corneal reflection pattern to compute the scaled P-CR vector however, resulted in a POG estimate for all head positions and eye rotations as there was no loss of the centroid shown in Table III. **Consequently the use of multiple redundant corneal reflections results in a more reliable system for POG estimation in which head motion is allowed.**

An additional benefit of the pattern matching and redundant corneal reflections is apparent when the system user is wearing eye glasses. Eye glasses commonly cause problems due to reflections off the surface of the lenses. These reflections can be identified and discarded through the use of the corneal reflection pattern matching algorithm, as shown in Fig. 9. Provided redundant light sources are available, in the event that the eye glasses reflection overlaps and obscures the corneal reflection pattern, the offending IR light source may be turned OFF and the remaining reflections were identified with the pattern matching algorithm.

Tracking the corneal reflections provides an estimate of the scale and translation of the corneal reflection pattern. In the multisubject experiment, the traditional P-CR, enhanced P-CR, and model-based methods were each used to estimate the POG at the same time, using the same source image data, to compare the accuracy of the three POG-estimation methods. The average error across the entire screen for the binocular (averaged) eyes using each of the three POG-estimation methods was compared using a one-way repeated measures analysis of variance (ANOVA) at each of the three depths tested. The SPSS software package was used for the statistical analysis.

At the close distance ($F(2,18) = 7.75, p < 0.05$) and far distance ($F(2,18) = 13.00, p < 0.05$) a statistically significant difference was found between the POG-estimation methods. A *post-hoc* analysis showed that the statistically significant difference (at the 0.05 level) was between the traditional P-CR method and both the enhanced P-CR and model-based methods. The ability of the enhanced P-CR method to handle changes in head depth was shown to improve to match that of the model-based method as no statistically significant difference was observed between the two methods. At the middle depth, no statistically significant differences between the accuracies of the

POG-estimation techniques were found ($F(2,18) = 3.89, p \geq 0.05$). No improvement of the enhanced P-CR method over the traditional method at the middle depth was expected however, as the user calibration was originally performed at approximately the same depth. Overall POG-estimation accuracies as good as 0.85 cm or 0.79° of visual angle were observed with the enhanced P-CR. It should be noted that while the enhanced P-CR method was shown to improve the POG-estimation accuracy over the head depths evaluated, a number of limitations restrict the practical capabilities of this technique. These limitations include the loss of image focus due to limited depth of field as well as the reduction in spatial resolution as the head is displaced further from the camera sensor, resulting in features imaged over a smaller number of sensor pixels.

The average error achieved by the enhanced P-CR procedure presented here is comparable to the results observed by Cerrolaza *et al.* [7], under similar experimental conditions. Their best average error was 1.08, 0.42, and 1.33 cm at the close, middle, and far distances, respectively, assuming a similar screen size and resolution, as their results must be converted from pixels. Unlike the more robust multiple corneal reflection pattern matching technique proposed here, their method uses only two corneal reflections that may be susceptible to error should distortion of one corneal reflection occur due to certain eye or head rotations. A chin rest was used in their experiments to fix the subject's heads, ensuring proper head placement and valid image features. In addition, the chin rest made it possible to exactly relocate the head to the middle (calibration) position, resulting in the improved performance with respect to the middle head position of the free-head system presented here.

Binocular eye-gaze tracking allowed for averaging of the left- and right-eye POG estimates to potentially increase the overall system accuracy through averaging. For the remote eye-gaze tracking system presented here, the left, right, and binocular POG accuracy results of the three POG-estimation methods at the middle depth were analyzed with one-way repeated measures ANOVAs. For both the enhanced P-CR ($F(2,18) = 8.86, p < 0.05$) and model-based ($F(2,18) = 7.71, p < 0.05$) methods, the binocular POG-estimation accuracy was statistically better (at the 0.05 level) than the POG accuracy of both the left or the right eyes alone. For the traditional P-CR method, no statistically significant difference was observed ($F(2,18) = 2.85, p \geq 0.05$). Binocular tracking with averaging of the left- and right-eye POG estimates in remote eye-gaze tracking therefore equals or improves on the accuracy of monocular tracking alone.

A simulation was developed to further validate the proposed methods and results of the real-world experiment. The scale of the corneal reflection patterns at each head depth were found to be closely matched between the simulation and real-world experiment at each head depth: 1.15 versus 1.12 at the 58 cm head depth, 1.0 versus 0.99 at the 62 cm head depth, and 0.88 versus 0.90 at the 66 cm head depth, respectively. The similarity in scale between the simulation and average of the multiple subject experiment indicates a reasonable match between the approximations used in the simulation (spherical cornea surface

and population averages for the eye model parameters) and the real world.

As with the real-world experiment, an analysis and simulation showed that rescaling of the P-CR vector can be used to improve the accuracy of POG estimation when the head is displaced in depth. In the simulation however, there was no improvement in accuracy from averaging the left- and right-eye POG estimates, since there was no difference between the simulated left and right eye (other than their respective positions) and no cancellation of symmetric errors were possible.

V. CONCLUSION

Remote eye-gaze tracking requires the ability to handle both head and eye motion since the camera-to-eye displacement is not fixed as it is with head-mounted systems. The use of the system-calibration-free P-CR method for POG estimation in remote eye-gaze tracking systems has been problematic as the P-CR method suffers from degradation in accuracy with head motion. To prevent the degradation in accuracy with head displacements toward or away from the screen, multiple corneal reflections and point-pattern matching techniques were developed to provide a scaling correction for the P-CR vector. **Using the centroid of the corneal reflection pattern to form the P-CR vector also improved the robustness to loss and distortion of the corneal reflections due to head and eye movement.**

With displacements of the head in depth it was shown that the performance of the enhanced P-CR method matched that of the model-based method, while avoiding the need for complex system calibration. Accuracy of the enhanced P-CR method was found to be 2.8 times better than the traditional method (from an average error of 2.8 to 1.0 cm) at the near depth displacement of the head from the user calibration position, while an improvement of 1.8 times was observed at the far displacement (2.5 to 1.3 cm average error). A simulation based on a model of the system verified the scaling behavior of the P-CR vector with head displacements, as well as the improvement in performance when rescaling takes place. It was also shown that over seven different subjects, **binocular averaging of the left- and right-eye POG estimates resulted in an accuracy that was statistically equal to or better than the monocular performance for the traditional P-CR, enhanced P-CR, and model-based POG-estimation methods.** 平均双眼

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

ACKNOWLEDGMENT

The authors would like to express their appreciation for the support of the Natural Sciences and Engineering Research

Council of Canada (NSERC) Chair in Design Engineering, and NSERC Discovery Grant #4924.

REFERENCES

- [1] A. T. Duchowski, *Eye Tracking Methodology: Theory and Practice*. New York: Springer-Verlag, 2003.
- [2] C. H. Morimoto and M. R. M. Mimica, "Eye gaze tracking techniques for interactive applications," *Comput. Vis. Image Underst.*, vol. 98, no. 1, pp. 4–24, 2005.
- [3] T. Hutchinson, J. White, 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.
- [4] S. K. Schnipke and M. W. Todd, "Trials and tribulations of using an eye-tracking system," in *Proc. CHI' 2000 Extended Abstracts Human Factors Comput. Syst.*. New York: ACM, pp. 273–274.
- [5] R. Jacob and K. Karn, "Eye tracking in human-computer interaction and usability research: Ready to deliver the promises (section commentary)," in *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*. Amsterdam, The Netherlands: Elsevier, 2003, pp. 573–605.
- [6] H. Hua, P. Krishnaswamy, and J. P. Rolland, "Video-based eyetracking methods and algorithms in head-mounted displays," *Opt. Exp.*, vol. 14, no. 10, pp. 4328–4350, 2006.
- [7] J. J. Cerrolaza, A. Villanueva, and R. Cabeza, "Taxonomic study of polynomial regressions applied to the calibration of video-oculographic systems," in *Proc. 2008 Symp. Eye Tracking Res. Appl.*. New York: ACM, pp. 259–266.
- [8] 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.
- [9] 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. 3, no. 4, pp. 323–344, Aug. 1987.
- [10] T. Ohno and N. Mukawa, "A free-head, simple calibration, gaze tracking system that enables gaze-based interaction," in *Proc. 2004 Symp. Eye Tracking Res. Appl.*. New York: ACM, pp. 115–122.
- [11] D. Beymer and M. Flickner, "Eye gaze tracking using an active stereo head," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 18–20, 2003, vol. 2, pp. II-451–II-458.
- [12] R. J. K. Jacob, *Eye Movement-Based Human-Computer Interaction Techniques: Toward Non-Command Interfaces*, vol. 4. Norwood, NJ: Ablex, 1993, pp. 151–190.
- [13] Y. Cui and J. M. Hondzinski, "Gaze tracking accuracy in humans: Two eyes are better than one," *Neurosci. Lett.*, vol. 396, no. 3, pp. 257–262, Apr. 2006.
- [14] 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.
- [15] 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, 2000.
- [16] 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.
- [17] C. Hennessey, B. Noureddin, and P. Lawrence, "Fixation precision in high-speed noncontact eye-gaze tracking," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 38, no. 2, pp. 289–298, Apr. 2008.
- [18] Z. Zhu, Q. Ji, and K. Fujimura, "Combining Kalman filtering and mean shift for real time eye tracking under active IR illumination," in *Proc. 16th Int. Conf. Pattern Recognit.*, vol. 4. Washington, DC: IEEE Comput. Soc., 2002, p. 40318.
- [19] Z. Zhu, K. Fujimura, and Q. Ji, "Real-time eye detection tracking under various light conditions," in *Proc. 2002 Symp. Eye Tracking Res. Appl.*. New York: ACM, pp. 139–144.
- [20] Q. Ji and X. Yang, "Real-time eye, gaze, and face pose tracking for monitoring driver vigilance," *Real-Time Imag.*, vol. 8, pp. 357–377, Oct. 2002.
- [21] G. Cox and G. de Jager, "A survey of point pattern matching techniques and a new approach to point pattern recognition," in *Proc. 1992 South African Symp. Commun. Signal Process.*, Sep. 11, 1992, pp. 243–248.
- [22] C. Hennessey, B. Noureddin, and P. Lawrence, "A single camera eye-gaze tracking system with free head motion," in *Proc. 2006 Symp. Eye Tracking Res. Appl.*. New York: ACM, 2006, pp. 87–94.
- [23] J.-Y. Bouguet, *Camera calibration toolbox for MATLAB* [Online]. Available: www.vision.caltech.edu/bouguetj/
- [24] OpenCV: opencvlibrary.sourceforge.net.
- [25] M. Böhme, M. Dorr, M. Graw, T. Martinetz, and E. Barth, "A software framework for simulating eye trackers," in *Proc. Eye Tracking Res. Appl. (ETRA)*, 2008, pp. 251–258. Available: <http://portal.acm.org/citation.cfm?id=1344529>



Craig A. Hennessey received the B.A.Sc degree in systems engineering from Simon Fraser University, Burnaby, BC, Canada, in 2001, and the M.A.Sc. and Ph.D. degrees in electrical engineering from the University of British Columbia, Vancouver, BC, in 2005 and 2008, respectively.

His current research interests include applications of eye-gaze tracking, computer vision, image processing, and human-computer interaction.



Peter D. Lawrence (S'64–M'73–SM'06) received the B.A.Sc. degree in electrical engineering from the University of Toronto, ON, Canada, in 1965, the Masters 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.

He was a Guest Researcher at Chalmers University's Applied Electronics Department, Goteborg, Sweden between 1970 and 1972, and between 1972 and 1974, as a Research Staff Member and Lecturer in the Mechanical Engineering Department, Massachusetts Institute of Technology, Cambridge. Since 1974, he has been with the University of British Columbia and is currently a Professor in the Department of Electrical and Computer Engineering. He is also registered as a Professional Engineer. His current 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.