Frame-Rate Pupil Detector and Gaze Tracker

C.H. Morimoto† D. Koons A. Amir M. Flickner

†Dept. Ciência da Computação
IME/USP - Rua do Matão 1010
São Paulo, SP 05508, Brazil
hitoshi@ime.usp.br

IBM Almaden Research Center
650 Harry Road K57
San Jose, CA 95120, USA
{dkoons,arnon,flick}@almaden.ibm.com

Abstract

We present a robust, frame-rate pupil detector technique, based on an active illumination scheme, used for gaze estimation. The pupil detector uses two light sources synchronized with the even and odd fields of the video signal (interlaced frames), to create bright and dark pupil images. The retro-reflectivity property of the eye is exploited by placing an infra-red (IR) light source close to the camera's optical axis resulting in an image with a bright pupil. A similar off axis IR source generates an image with dark pupils. Pupils are detected from the thresholded difference of the bright and dark pupil images. After a calibration procedure, the vector computed from the pupil center to the center of the corneal glints generated from light sources is used to estimate the gaze position. The frame-rate gaze estimator prototype is currently being demonstrated in a docked 300 MHz IBM Thinkpad with a PCI frame grabber, using interlaced frames of resolution 640×480×8 bits.

1 Introduction

Robust face detection and tracking will be fundamental to future human computer interaction, and any reliable technique for detecting eyes would greatly simplify this task. The requirement for interaction imposes severe constraints on the response time of these image processing tasks, which are also known to have high computational demands. In this paper we describe a novel robust frame-rate pupil detector technique that is suitable for desktop and kiosk applications.

Current research on real-time face detection and tracking are model-based, i.e., they use information about skin color [5, 7] or face geometry [1] for example. The technique described in this paper explores physical properties of eyes (i.e., their retro-reflectivity) to segment them using an active illumination scheme described in Section 2. Eye properties have been used before in commercial eye gaze trackers such as those available from ISCAN Incorporated, Applied Science Laboratories (ASL), and LC Technologies, but they use only bright or dark pupil images for tracking.

Due to the retro-reflectivity of the eye, a bright pupil image is seen by the camera when a light source is placed very close to its optical axis (Figure 1). This effect is well known as the red-eye effect from flash photographs [8]. Under regular illumination (when the light source is not on the camera's optical axis), a dark pupil is seen. The trick for robust pupil detection is to combine dark and bright pupil images, where pupil candidates are detected from the thresholded difference of the dark from the bright pupil image, as seen in Figure 2.

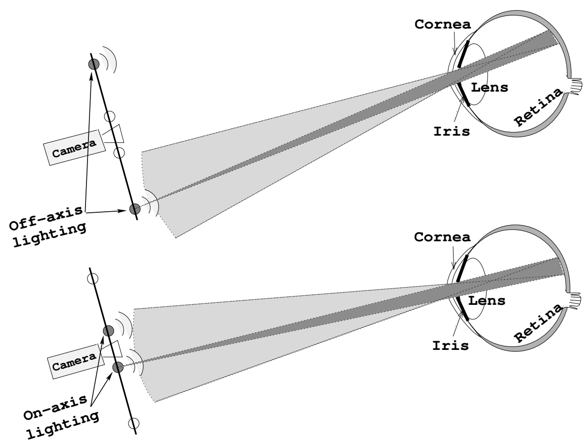
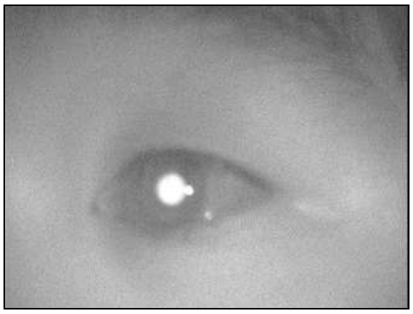
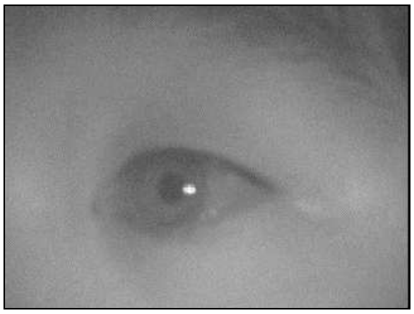


Figure 1: Retro-reflectivity of the eye. Observe that when the light source is placed off-axis (top), the camera does not capture the light returning from the eye.

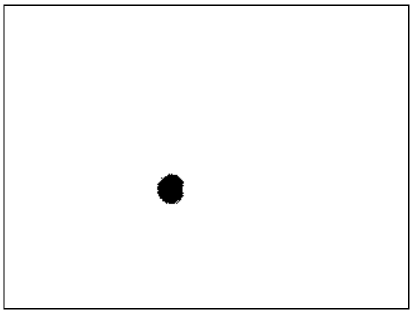
The pupil detection systems presented in [6, 2] are also based on a differential lighting with thresholding scheme. These systems are used to detect and track the pupil and estimate the point of gaze, which also requires the detection of the corneal reflections created by the light sources. The corneal reflection from the light sources can be easily seen as the bright spot close to the pupils in Figures 2a and 2b. Our system differs from these due to its simplicity, and the



(a)



(b)



(c)

Figure 2: (a) Bright and (b) dark pupil images. (c) Difference of the dark from the bright pupil after thresholding

constraint to use "off-the-shelf" hardware. In a previous paper [3] we have described a real-time eye and face detector system based on the differential lighting with thresholding scheme. This paper introduces several enhancements we have made to build the frame-rate (30 frames/second) pupil tracker and gaze estimator.

The next section describes several issues related to the implementation of the pupil detector based on the active illumination scheme, and Section 3 presents the eye gaze tracker built on top of the pupil detector. Experimental results for both pupil detector and eye gaze tracker are given in Section 4. Section 5 concludes the paper and discusses future work.

2 Implementation Issues

Figure 3 shows the pan-tilt camera with the illuminators used in the eye tracking system. The pan-tilt camera is a Sony EVI D30, and the illuminators are constituted by two sets of light sources. For convenience, near infra-red (IR) light with wavelength 875nm is used, which is invisible to the human eye. The IR illumination also makes the system insensitive to changes in indoor ambient illumination, i.e., the room lights can be turned on/off without affecting the operation of the system. We slightly modified the original camera optics to adjust it to operate in near IR (by removing its IR block filter) and introducing a pair of extra lenses to increase its optical magnification.

The light sources LIGHT1 and LIGHT2 in Figure 3 are composed of sets of 7 IR LEDs each. LIGHT2 is composed of two sets of LEDs, symmetrically placed on the left and right sides of the optical axis. Symmetry around the optical axis is desired because it reduces shadow artifacts by producing more uniform illumination, but asymmetrical configurations also perform adequately. LIGHT1 is placed near the camera's optical axis, so it generates the bright pupil image (Figure 2a) when it is on, and LIGHT2 is placed off-axis to generate a dark pupil image (Figure 2b), adjusted for similar brightness in the rest of the scene.

The video signal from the camera is composed of interlaced frames, where one frame can be decomposed into an even field and an odd field. Thus, a field has half the vertical resolution of a frame. Let F_t be an image frame taken at time instant t, with resolution c columns (width) by r rows (height), or $c \times r$. F_t can be de-interlaced into E_t and O_t, where E_t is the even field composed by the even rows of F_t and O_t is the odd field composed by the odd rows of F_t.

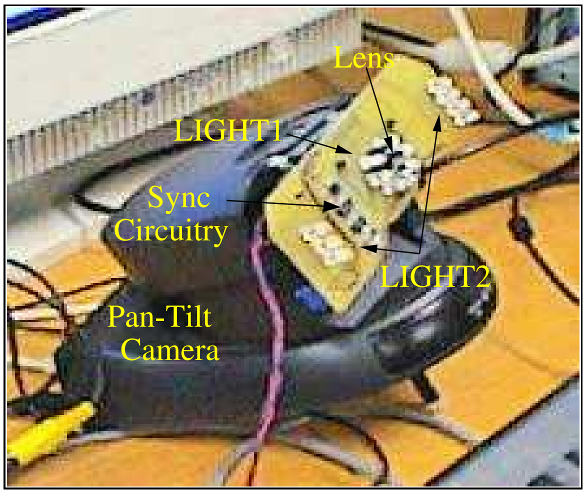


Figure 3: Camera and IR illumination setting.

When LIGHT1 is synchronized with the even fields and LIGHT2 with the odd fields, i.e., each illuminator stays on for just half the frame period, one interlaced frame will contain both bright and dark pupil images. Figure 4 shows a block diagram of the pupil detection process. Once an interlaced frame F_t is captured, it is de-interlaced and the odd field O_t is subtracted from the even field E_t (dark from the bright pupil images). Thresholding of the difference image then creates a binary image, which is the input of a connected component labeling algorithm. Each connected component (blob) is checked for particular geometric properties, such as size and aspect ratio, and those that satisfy these constraints, are output as pupils. Observe that it is also possible to detect pupils using (O_t, E_{t-1}), i.e., between frames, thus increasing the detection rate to 60 fields per second.

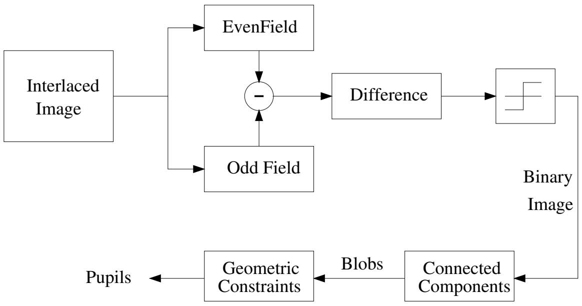


Figure 4: Pupil detection block diagram.

The only piece of hardware build for the system was a very simple device to keep the even and odd fields synchronized with LIGHT1 and LIGHT2 respectively. Figure 5 shows a block diagram of the light synchronization device. The video signal from the camera is received by a video decoder module that separates the even and odd field signals. The video decoder is a National LM1881 chip, that is mounted on the same board that supports the IR LEDs (see Figure 3). The signal is fed to amplifying buffers that provide power for the IR LEDs.

3 Eye Gaze Tracking

The purpose of an eye gaze tracker is to estimate the position on the screen to where the user is fixating her/his gaze. This is accomplished by tracking the user's pupil and the corneal glint, after a brief calibration procedure, that determines the mapping from coordinates of the pupil tracker to user screen coordinates. Assuming a static head, an eye can only rotate in its socket, and the surface of the eye can be approximated by a sphere. Since the light sources are also fixed, the glint on the cornea of the eye can be taken as a reference point, thus the vector from the glint to the center of the pupil will describe the gaze direction.

To estimate the screen coordinates to where the user is looking, a simple second order polynomial transformation is used. After the calibration procedure, a simple possible application is to control the mouse using eye gaze, which provides an estimate about the accuracy of the system. We have obtained an accuracy of about 1 degree of resolution, that corresponds to about 1cm on the screen viewed from 50cm.

3.1 Calibration Procedure

The calibration procedure is very simple and brief. Nine points are arranged in a 3×3 grid on the screen, and the user is asked to fixate his/her gaze on a certain target point, press a key, and move to a next target, until all the points are fixated. On each fixation, the vector from the center of the glint to the center of the pupil is saved, so that 9 corresponding points are obtained. The transformation from a glint-pupil vector $\mathbf{E} = (x_e, y_e)^t$, to a screen coordinate $\mathbf{S} = (x_s, y_s)^t$ is given by:

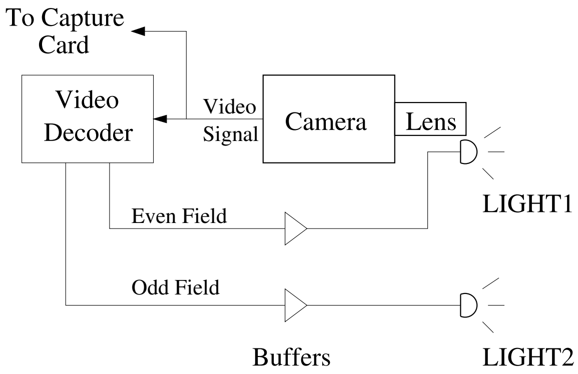


Figure 5: Synchronization device block diagram.

$$\begin{aligned} x_e &= a_0 + a_1 x_s + a_2 y_s + a_3 x_s y_s + a_4 x_s^2 + a_5 y_s^2 \\ y_e &= a_6 + a_7 x_s + a_8 y_s + a_9 x_s y_s + a_{10} x_s^2 + a_{11} y_s^2 \end{aligned} \quad (1)$$

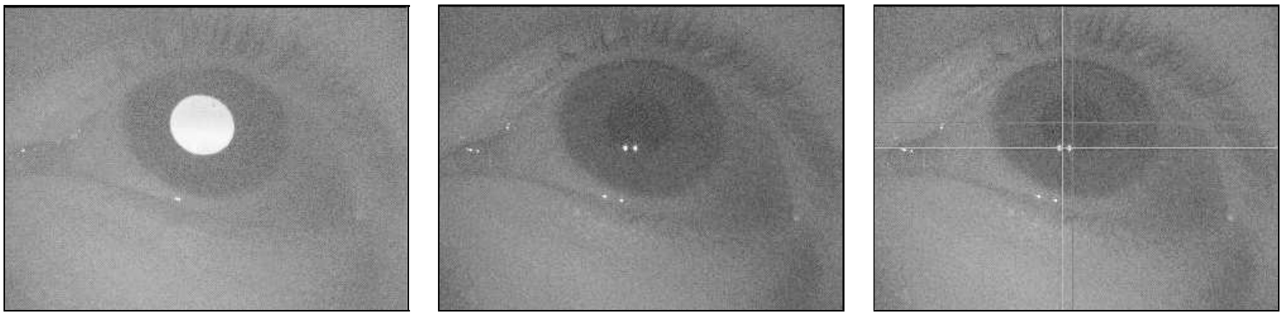
where a_i are the coefficients of this second order polynomial. Each corresponding point gives 2 equations from (1), thus 18 equations are produced and an over determined linear system is obtained. The polynomial coefficients for x_e and y_e can be obtained independently, so that 2 simpler over determined systems are solved, using a least squares method.

3.2 Glint-Pupil Vector

To compute the glint-pupil vector it is necessary to extract the centers of the pupil and the glint. Since the field of view is very narrow, the pupil is the biggest round blob obtained after the labeling algorithm. To estimate the center the pupil a window slightly larger than the enclosing box of the pupil is created. Then gray scale pixels in the difference image are summed horizontally and vertically (Radon transform). The x, y center is computed as the center of mass of the horizontal and vertical projections (sums). A search procedure for very bright pixels around the pupil is used to detect the glint and compute its center of mass. Ideally all coordinates are computed with subpixel repeatability. Figure 6 shows the bright, dark, and the dark pupil image with two crosses superimposed, that correspond to the computed centers of the pupil and glint.

3.3 Pan-Tilt Servo Mechanism

In order to allow some head motion, it is required to keep the pupil centered in the image. The magnitude of the rotation angle of the camera which brings the pupil to the center of the image (assuming the rotation is around



(a) (b) (c)

Figure 6: (a) Bright and (b) dark pupil images. (c) Dark pupil image superimposed by two crosses, marking the center of the pupil and the glint (for monitoring and debugging purposes).

its principal point), will only depend of the image size and the field of view (FOV) of the camera. If the center of the pupil is at the pixel (x, y) (assume the pixel $(0, 0)$ to be the center of the image), and given the FOV $= (\phi_x, \phi_y)$, and image size $W \times H$, the pan and tilt are given by:

$$\text{pan} = \phi_x \frac{x}{W} \tag{2}$$

$$\text{tilt} = \phi_y \frac{y}{H} \tag{3}$$

4 Experimental Results

The current prototype was implemented on a dual Pentium II 400MHz machine running Windows NT4, using a commercial PCI capture card compatible with Video for Windows. The eye tracker runs at frame-rate (30 frames per second), processing interlaced frames of resolution $640\times480\times8$ bits. We have also achieved this frame rate with an IBM 300 MHz Pentium II Thinkpad 770X machine running Windows NT4, using a PCI capture card installed in an IBM dock III.

Eyeglasses and contact lens do not change the retro-reflectivity of the eye and unless the glasses or contacts are tinted with an IR blocking coating, they do not inhibit detection. Figure 7 shows the bright, dark, and difference images for a person with glasses. Observe in Figure 7c that spurious false pupil candidates are generated by the specular reflections from the eyeglasses, and these reflections can also block the dark pupil response under very particular head orientations. In such cases, if the head motion must be restricted, a slight change in the orientation of the glasses is enough to reestablish detection and gaze estimation. Pupil detection using only the dark or only the bright pupil images, as it is done by most commercial eye trackers for gaze estimation, would have a lot more spurious responses, which can be expected from images of the same kind shown in Figure 7.

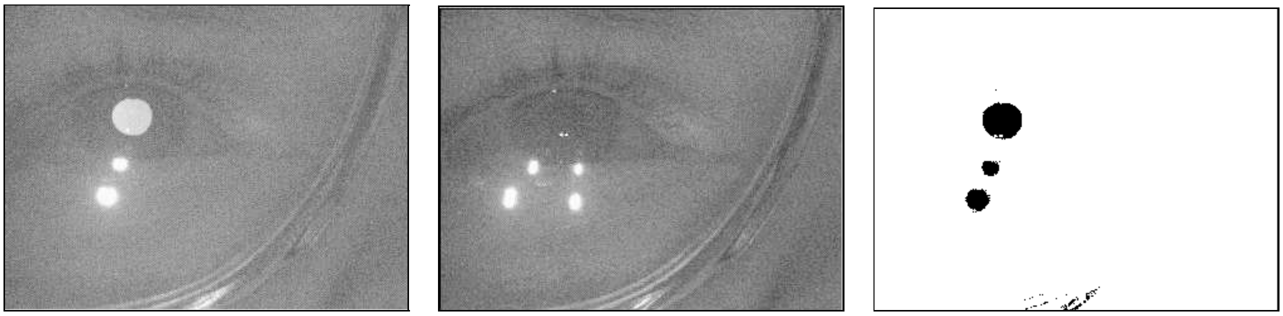
The retro-reflectivity property of eyes is uncommon in man-made and natural objects resulting in pupils generally being the only objects appearing with high contrast between the two pupil images. Pupil detection is greatly facilitated by the enhanced signal-to-noise ratio, and the simple process of thresholding the difference between bright and dark pupil images is generally sufficient, as shown in Figure 2c. Our experience shows that most retro-reflectors we tested, used for example in running shoes, reflect light in a reasonable wide angle, so that they appear bright in both images and do not cause artifacts. Certain lamps, like table lamps and ceiling lamps, have reflectors that when pointed to the camera can cause artifacts.

The one degree accuracy mentioned in Section 3 and the small head motion allowed by the system is comparable with commercial systems. The limitations on head motion is due to the simple motion model adopted, because the calibration changes with different head positions. We are currently working on more complex models to allow free head motion. Other applications such as real-time face tracking [3] and enhanced human-computer interaction [4, 9] using the frame-rate pupil detector are described in other publications.

5 Conclusion

We have presented a robust frame-rate pupil detector and eye gaze tracker, with high potential to be used in human-computer interaction, particularly in desktop and kiosk applications. The even and odd frames of a video camera are synchronized with two IR light sources. A pupil is alternately illuminated with an on-axis IR source when even frames are being captured, and with an off-axis IR source for odd frames. The on camera axis illumination generates a bright pupil, and the off axis illumination keeps the scene at about the same illumination, but the pupil remains dark. Detection follows from thresholding the difference between even and odd frames.

Once the pupil is detected, the corneal glint from the light sources is searched for, near the center of the pupil. The eye gaze tracker uses the center of the pupil and glint



(a) (b) (c)

Figure 7: (a) Bright and (b) dark pupil images with glasses. (c) Difference image after thresholding. The strong glints on the glasses can be avoided with a slight change in its orientation.

to estimate the position on the screen to where the user is fixating her/his gaze, after a brief calibration procedure that determines the mapping from coordinates of the pupil tracker to user screen coordinates.

The eye gaze tracker has been successfully tested for a very large number of people, and it has proven to be very robust. Future extensions include the generalization of the problem to a 3D model in order to allow for large head motion and enhancements on the pupil detector to increase its accuracy, that includes changes in the calibration procedure and mapping functions.

References

[1] S. Birchfield. An elliptical head tracker. In *Proceeding of the 31st Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, CA, November 1997.

[2] Y. Ebisawa and S. Satoh. Effectiveness of pupil area detection technique using two light sources and image difference method. In A.Y.J. Szeto and R.M. Rangayan, editors, *Proceedings of the 15th Annual Int. Conf. of the IEEE Eng. in Medicine and Biology Society*, pages 1268–1269, San Diego, CA, 1993.

[3] C. Morimoto, D. Koons, A. Amir, and M. Flickner. ”real-time detection of eyes and faces”. In *Proceedings of 1998 Workshop on Perceptual User Interfaces*, pages 117–120, San Francisco, CA, November 1998.

[4] C.H. Morimoto, D. Koons, A. Amir, M. Flickner, and S. Zhai. Keeping an eye for hci. In *Proc. of the XII Brazilian Symposium on Computer Graphics and Image Processing - Sibgrapi 99*, Campinas, SP, October 1999.

[5] N. Oliver, A. Pentland, and F. Berard. Lafter: Lips and face real time tracker. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 123–129, Puerto Rico,PR, June 1997.

[6] A. Tomono, M. Iida, and Y. Kobayashi. A tv camera system which extracts feature points for non-contact eye movement detection. In *Proceedings of the SPIE Optics, Illumination, and Image Sensing for Machine Vision IV*, volume 1194, pages 2–12, 1989.

[7] J. Yang and A. Waibel. A real-time face tracker. In *Proceedings of the Third IEEE Workshop on Applications of Computer Vision*, pages 142–147, Sarasota, FL, 1996.

[8] L. Young and D. Sheena. Methods & designs: Survey of eye movement recording methods. *Behavioral Research Methods & Instrumentation*, 7(5):397–429, 1975.

[9] S. Zhai, C.H. Morimoto, and S. Ihde. Manual and gaze input cascaded (magic) pointing. In *Proc. ACM SIGCHI - Human Factors in Computing Systems Conference*, pages 246–253, Pittsburgh, PA, May 1999.