Image and Vision Computing 18 (2000) 331–335
www.elsevier.com/locate/imavis

Pupil detection and tracking using multiple light sources

C.H. Morimoto[a,*], D. Koons[b], A. Amir[b], M. Flickner[b]

[a]Dept. de Ciência da Computação—IME/USP, Rua do Matão 1010, São Paulo, SP 05508-900, Brazil
[b]IBM Almaden Research Center, 650 Harry Road, B2 K57, San Jose, CA 95120, USA

Received 29 September 1998; received in revised form 14 July 1999; accepted 20 July 1999

Abstract

We present a fast, robust, and low cost pupil detection technique that uses two near-infrared time multiplexed light sources synchronized with the camera frame rate. The two light sources generate bright and dark pupil images, which are used for pupil segmentation. To reduce artifacts caused mostly by head motion, a larger temporal support is used. This method can be applied to detect and track several pupils (or several people). Experimental results from a real-time implementation of the system show that this technique is very robust, and able to detect pupils using wide field of view low cost cameras under different illumination conditions, even for people with glasses, from considerable long distances.

Keywords: Pupil detection; Gaze tracking

1. Introduction

Eye-tracking using computer vision techniques has the potential to become an important component in future computer interfaces. Jacob [1] describes several ways of using eye movements for human–computer interaction. Robust techniques for eye detection in images are of particular importance to eye-gaze tracking systems. Information about the eyes can also be used to detect and track human faces and bodies, for applications in face recognition, monitoring human activity, multi-modal interfaces, etc.

Commercial remote eye-tracking systems used for the estimation of a person's gaze (point of regard), such as those produced by ISCAN Incorporated, LC Technologies (LCT), and Applied Science Laboratories (ASL), rely on a single light source that is positioned off-axis in the case of the ISCAN ETL-400 systems, and on-axis in the case of the LCT and the ASL E504 systems. Illumination from an off-axis source (and normal illumination) generates a dark pupil image. When the light source is placed on-axis with the camera optical axis, the camera is able to detect the light reflected from the interior of the eye, and the image of the pupil appears bright [2,3] (see Fig. 2). This effect is often seen as the red-eye in flash photographs when the flash is close to the camera lens. These systems require the initial localization of the pupil in order to begin tracking. Other eye and face tracking systems such as those described in Refs. [4–6] could also benefit from robust pupil detection techniques.

Eyes can also be detected as a sub-feature following face-region detection and segmentation. Face detection can be done using several different methods, such as from head motion in front of a known background [7], skin color [8–10], geometric models and templates [4,11,12], machine learning and artificial neural networks techniques [13–15], etc.

Other techniques are specifically targeted to detect eyes. Kothari and Mitchell [16] use spatial and temporal information to detect the location of the eyes. Their process starts by selecting a pool of potential candidates using gradient fields. The gradient along the iris/sclera boundary always point outward the center of the iris (dark pupil), thus by accumulating along these lines, the center of the iris can be estimated by selecting the bin with highest count. Heuristic rules and a large temporal support are used to filter erroneous pupil candidates.

Our detection technique does not require models (color, geometry, templates, examples, etc), and is based on geometrical and physiological properties of the eye. Although special lighting and synchronization schemes are required, the scene background becomes irrelevant and the pupils can be detected in a wide range of scales and illumination conditions.

Tomono et al. [17] and Ebisawa and Satoh [18] have developed systems very similar to the one presented in

* Corresponding author. Tel.: +55-11-818-6135; fax: +55-11-814-4135.
E-mail address: hitoshi@ime.usp.br (C.H. Morimoto).

0262-8856/00/$ - see front matter
PII: S0262-8856(99)00053-0

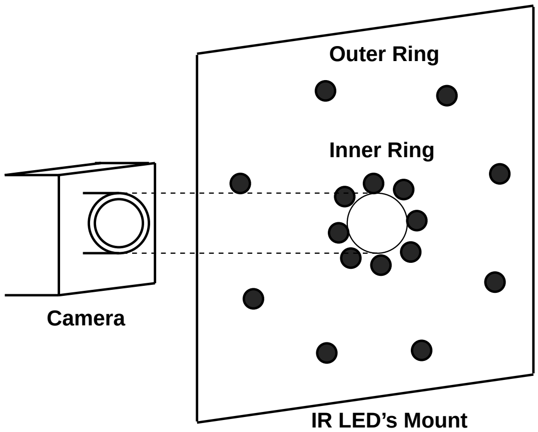

Fig. 1. Camera and infrared illuminators.

this paper. Both systems have the purpose of tracking eyes for estimating the point of gaze, which also requires the detection of the corneal reflections created by the light sources.

Tomono et al. [17] developed a real-time imaging system composed of three CCD cameras and two near-infrared (IR) light sources. The light sources have different wavelengths (λ_1, and λ_2), and the light source with wavelength λ_1 (or λ_1 for simplicity) is polarized. λ_2 is placed near the camera optical axis, and λ_1 is placed slightly off-axis, generating the bright and dark pupil images, respectively. CCD3 is sensitive to λ_2 only, thus it outputs bright pupil images. CCD1 and CCD2 are sensitive to λ_1 (λ_2 is filtered out), and CCD1 also has a polarizing filter in order to receive only the diffuse light components, i.e. the corneal reflection due to λ_1 does not appear in the images from CCD1. Once the three images are available, the pupil is segmented from differencing and thresholding the images from CCD3 and CCD2, and the corneal reflection used for gaze estimation is segmented using the images from CCD2 and CCD1.

The system from Ebisawa and Satoh [18] is also based on a differential lighting scheme using two light sources with same wavelength (on and off camera axis) to generate the bright/dark pupil images. The detection of the corneal reflection created by the light sources requires the use of a narrow field of view camera (long focal length) since the reflection is in general very small. In a more recent publication, Ebisawa [19] presents a real-time implementation of the system using custom hardware and pupil brightness stabilization for optimum detection of the pupil and the corneal reflection.

The system presented in this paper introduces a much simpler and inexpensive solution, and it is also based on the differencing followed by thresholding technique using bright and dark pupil images. The system uses a single wide field of view CCD camera, and two IR sources (with single wavelength) disposed symmetrically around the camera's optical axis in order to generate concentric corneal reflections. The wide field of view allows for the detection of multiple pupils. Geometric constraints could also be used to group the pupils, so that several heads could be detected and tracked. More robust and accurate filtering and tracking algorithms were also developed, and are presented in Section 3. The next section describes the basic configuration of the system, and Section 4 presents experimental results from a real-time implementation of the pupil detection and tracking system. Section 5 concludes the paper.

2. System configuration

The pupil detection and tracking system consists of a single CCD B&W camera and two light sources. For convenience, we use near infra red light sources, which are almost invisible to the human eye, and a camera sensitive to that wavelength. Fig. 1 shows a diagram of the configuration of the camera and the illuminators.

The illuminators used in the current implementation consist of two concentric rings of IR LEDs. The center of the rings coincide with the camera optical axis. Both rings are mounted on the same plane; the inner ring diameter is approximately the same as the camera lens, and the diameter of the outer ring is adjusted in order to generate a dark pupil image.

The even and odd frames of the camera are synchronized with the inner and outer rings respectively, i.e. when the inner ring is on, an even frame is grabbed, and alternately, when the outer ring is on, an odd frame is grabbed. The inner ring is sufficiently close to the camera optical axis and generates the bright pupil image. The diameter of the outer ring is sufficiently large (the LEDs are far from the camera axis) in order to generate a dark pupil image, and sufficiently bright to generate approximately the same illumination as the inner ring.

For the purpose of eye-tracking for gaze estimation, it is also important to detect and track the glint (corneal reflection) [2,20,21]. Depending on the illumination conditions and position of the glint, it is often difficult to detect it in the bright pupil image. Tomono et al. [17] use polarized light and two CCDs to segment the corneal reflection. Ebisawa [19] has developed a pupil brightness stabilization

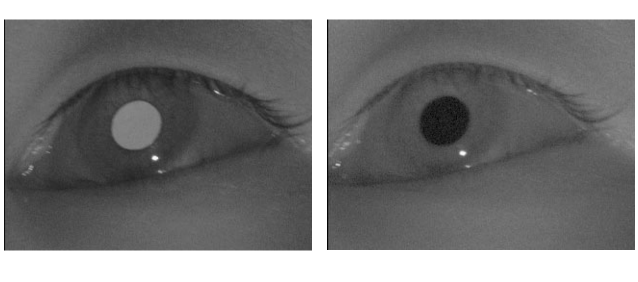

Fig. 2. (a) Bright and (b) dark pupil images. The glints, or corneal reflections, from the on and off-axis light sources can be easily identified as the bright regions in the iris.

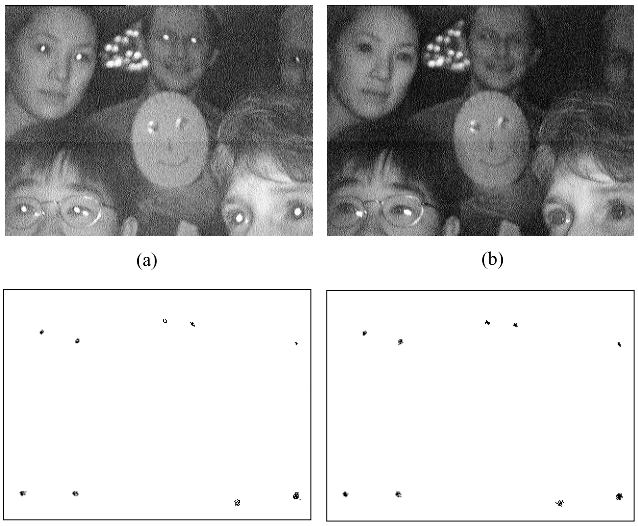

Fig. 3. (a) Bright and (b) dark pupil images. (c) and (d) show the pupils detected from subtraction and thresholding using the corresponding previous frames, after erosion.

technique to keep the illumination of the on-axis light source from saturating the camera, making it possible to detect the glint under different illumination conditions. His system does not use the glint from the dark pupil image, because in their implementation the glint from the off-axis LED moved with the zoom level and was also hard to detect due to low contrast with its surrounding neighborhood.

Since the rings used in our implementation are concentric, the position of the centers of the glints are the same in the dark and bright pupil images. The glint can be easily detected from the dark pupil images, without the need for stabilizing the pupil brightness. Fig. 2 shows examples of bright and dark pupil images. The glints can be easily seen from these pictures, and correspond to the bright regions in the iris, below the pupils.

3. Pupil detection and tracking algorithms

Pupils can be detected from simple thresholding of the difference of the dark from the bright pupil images. In the case of large pupil displacement between frames, due to fast motion of the head and/or eyes, the pupil is lost and detected again as soon as there is some pupil overlap between frames. If the motion is small, so that some overlap exists, pupils can still be detected, but motion artifacts might make this task more difficult.[1]

Most motion artifacts can be filtered by considering a larger temporal support, i.e. more than two images are considered for pupil detection. If the eyes do not move much during the sampling period of F frames, the pupils can be detected as high contrast regions from the differences between all consecutive pairs, while most motion artifacts are detected only between some of the pairs (motion from some particular textured surfaces could also be present after temporal filtering). This method introduces a delay of $F - 1$ frames every time the pupil is lost, since the pupil must remain approximately still for at least F frames until it is detected again.

Once a pupil is reliably detected, it can be tracked using the thresholded difference image as in Ref. [19]. This solution is computationally cheap to be implemented, but the tracking accuracy can be severely affected since the center of the blob detected from the thresholded difference image corresponds to the center of the overlapping region between two consecutive frames. This overlapping region might be significantly different from the true center of the pupil in either frame, especially for close-up images of the eye. Hence, for accurate tracking, the pupil must be tracked using the gray level images, either the dark or bright pupil images, instead of the thresholded difference.

Tracking can be implemented using the bright, the dark, or both pupil images. As mentioned in Section 2, the use of the dark pupil images helps glint detection, and the host computer can control the alternating lighting scheme so that only the outer ring is activated during tracking. For the current implementation though, we use both dark and bright pupils for tracking.

Blobs from the thresholded difference image are segmented using a connected component algorithm. Geometric constraints are used to determine the best two pupil candidates, and a gray level histogram is build for the selected blob regions from the bright or dark pupil images. The histogram is used to grow the regions to the true boundaries of the pupil. Tracking is then performed as a search for a dark/bright region in the next frame close to the previous pupil position. Other tracking techniques such as the Kalman filter are being implemented.

4. Experimental results

A real-time implementation of the pupil detection system is currently running at 15 frames/s on a single processor Pentium 200 machine. The images are of resolution $320 \times 240 \times 8$ bits. A PCI frame-grabber with no on-board processing capabilities is used for image acquisition. One of the computer's serial communication ports is used to synchronize the LED rings with the frame-grabber.

The current implementation uses a 1/3″ CCD B&W board camera with a micro lens mount (12 mm thread), with a 12 mm lens, and a visible light block filter. Despite the fact that the micro camera provides images of very low quality, mainly because it does not have controls for the gain, iris, shutter speed, etc. the system proves to be very robust. The current prototype is able to detect pupils up to

[1] Motion artifacts could be considerably reduced by increasing the frame rate.

(a) (b) (c)

Fig. 4. (a) Bright and (b) dark pupil images; (c) shows the pupils detected using a temporal support of four frames.

5 m from the camera, but this range depends on room illumination conditions and particularly on the subjects. Our tests using over 100 subjects show that 100% of healthy pupils were detected under regular indoor illumination, within desktop range (up to 1.5 m). We had subjects with severe eye damage where only one healthy eye was detected. The tests also show that false candidates are mostly due to high specularity surfaces and retro-reflector objects (see Fig. 5). To improve the overall detection range, we are currently investigating the causes of pupil brightness response variation for different subjects.

Fig. 3 shows the pupil detection system operating with typical noisy data. Simple thresholding after subtraction is used. To eliminate artifacts, high contrast regions only one pixel wide were eroded. Observe that the pupils from all subjects were detected, and the false eyes (glass marbles) in the center of the images were rejected since they do not have the correct optic and geometric properties that produce the bright/dark eye effect. Note also that the system is quite robust even for people wearing glasses.

Fig. 4 shows the result of the pupil detection system operating with a temporal support of four frames. Fig. 4a shows an even bright pupil image, Fig. 4b shows the consecutive odd dark pupil image, and Fig. 4c shows the output of the temporal filter, which eliminates all but the pupils.

Fig. 5 shows close-up pictures of the bright and dark pupil images when the user is wearing glasses. In order to reduce the physical dimensions of the device, we experimented a different configuration of light sources, where we split the off-axis source into two groups of LEDs positioned on the sides of the lens. This is the reason why there are two horizontal pairs of glints on the glasses in the dark pupil image shown in Fig. 5b. Each pair of glints corresponds to the reflection from one of the surfaces (back and front) of the glasses. The on-axis source remains unchanged, and Fig. 5a shows the bright pupil image with two extra glints from the front and back reflections on glasses surfaces. Observe from the pupil candidates shown in Fig. 5c, that the biggest blob still corresponds to the pupil. This is in general true in most of our experiments, and that explains why the system works well for people with glasses. It is possible though to lose the pupil when the glints from the surfaces of the glasses hide the dark pupil response, but these are particular situations that in practice only happen for a very short period.

5. Conclusions

We have described a simple, fast, robust, and inexpensive way to detect and track eyes. The even and odd frames of a video camera are synchronized with the IR light sources, and the face 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. Motion artifacts

(a) (b) (c)

Fig. 5. (a) Bright and (b) dark pupil images of a person wearing glasses. (c) shows the pupils candidates. The false candidates are caused by the reflections of the on-axis light source from the back and front surfaces of the glasses.

caused by the subject's head movements and other moving objects are filtered using a larger temporal support and other geometrical constraints on the size and relative position of the pupils. Once the pupils are detected, they can be tracked using several techniques such as the Kalman filter. The robustness of the technique is demonstrated by a real-time implementation, which is current running at 15 frames/s, using images of resolution $320 \times 240 \times 8$ bits.

The system has been successfully tested for a very large number of people, and it has proven to be very robust indoors, particularly for office environments, although it has not been tested outdoors, where high intensity illumination might introduce difficulties. The system is inexpensive and very compact (the dimensions of the current implementation are about $9 \times 9 \times 3$ cm), and we are confident that this method could greatly increase the robustness and accuracy of current remote face and eye-tracking systems.

Future extensions of this work include a system for gaze estimation, face and facial features detection and tracking, and a faster implementation which will synchronize the LED rings with the even and odd fields, increasing the processing speed to 60 fields per second.

References

[1] R. Jacob, The use of eye movements in human–computer interaction techniques: what you look at is what you get, ACM Transactions on Information Systems 9 (3) (1991) 152–169.
[2] T.E. Hutchinson, K.P. White Jr., K.C. Reichert, L.A. Frey, Human–computer interaction using eye-gaze input, IEEE Transactions on Systems, Man, and Cybernetics 19 (1989) 1527–1533.
[3] L. Young, D. Sheena, Methods and designs: Survey of eye movement recording methods, Behavioral Research Methods and Instrumentation 7 (5) (1975) 397–429.
[4] S. Birchfield, An elliptical head tracker, Proceedings of the 31st Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, November 1997.
[5] A. Gee, R. Cipolla, Fast visual tracking by temporal consensus, Image and Vision Computing 14 (2) (1996) 105–114.
[6] J. Yang, A. Waibel, A real-time face tracker, Proceedings of the Third IEEE Workshop on Applications of Computer Vision, Sarasota, FL, 1996, pp. 142–147.
[7] T. Darrell, B. Moghaddam, A. Pentland, Active face tracking and pose estimation in an interactive room, Technical Report 356, MIT Media Laboratory Perceptual Computing Section, Cambridge, MA, 1996.
[8] P. Fieguth, D. Terzopoulos, Color based tracking of heads and other mobile objects at video frame rates, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Puerto Rico, PR, June 1997, pp. 21–27.
[9] N. Oliver, A. Pentland, F. Berard, Lafter: lips and face real time tracker, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Puerto Rico, PR, June 1997, pp. 123–129.
[10] R. Stiefelhagen, J. Yang, A. Waibel, A model-based gaze tracking system, Proceedings of the Joint Symposia on Intelligence and Systems, Washington, DC, 1996.
[11] L.C. De Silva, K. Aizawa, M. Hatori, Detection and tracking of facial features, Proceedings of the SPIE Com. and Image Proc. 95, Taipei, Taiwan, May 1995, vol. 2501, pp. 1161–1172.
[12] A.L. Yiulle, P.W. Hallinan, D.S. Cohen, Feature extraction from faces using deformable templates, International Journal of Computer Vision 8 (2) (1992) 99–111.
[13] A.J. Colmenarez, T.S. Huang, Face detection with information-based maximum discrimination, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Puerto Rico, PR, June 1997, pp. 782–787.
[14] H. Rowley, S. Baluja, T. Kanade, Neural network-based face detection, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, CA, June 1996.
[15] K. Sung, T. Poggio, Example-based learning for view-based human face detection, Technical Report AI-1521, MIT AI Laboratory, Cambridge, MA, December 1994.
[16] R. Kothari, J.L. Mitchell, Detection of eye locations in unconstrained visual images, Proceedings of the International Conference on Image Processing, Lausanne, Switzerland, September 1996, vol. I, pp. 519–522.
[17] A. Tomono, M. Iida, Y. Kobayashi, A TV camera system which extracts feature points for non-contact eye movement detection, Proceedings of the SPIE Optics, Illumination, and Image Sensing for Machine Vision IV, 1989, vol. 1194, pp. 2–12.
[18] Y. Ebisawa, S. Satoh, Effectiveness of pupil area detection technique using two light sources and image difference method, in A.Y.J. Szeto, R.M. Rangayan (Eds.), Proceedings of the 15th Annual International Conferemce of the IEEE Engineering in Medicine and Biology Society, San Diego, CA, 1993, pp. 1268–1269.
[19] Y. Ebisawa, Unconstrained pupil detection technique using two light sources and the image difference method, Visualization and Intelligent Design in Engineering and Architecture (1995) 79–89.
[20] T.N. Cornsweet, H.D. Crane, Accurate two-dimensional eye tracker using first and fourth Purkinje images, Journal of the Optical Society of America 63 (8) (1973) 921–928.
[21] H.D. Crane, C.M. Steele, Accurate three-dimensional eyetracker, Journal of the Optical Society of America 17 (5) (1978) 691–705.