

# Pupil Detection and Tracking Using Multiple Light Sources

Carlos Morimoto    Dave Koons    Arnon Amir    Myron Flickner

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

**Abstract.** We present a fast, robust, and low cost pupil detection technique that uses two infra red (IR) time multiplexed light sources, composed of two rings of 8 LED's each, synchronized with the camera frame rate. One light source is placed very close to the camera's optical axis, and the second source is placed off-axis. The pupil appears bright in the camera image during on-axis illumination (similar to the red eye effect from flash photography), and dark when illumination is off-axis. The off-axis light source is calibrated to provide roughly-equivalent whole-scene illumination. Pupil detection follows from thresholding the difference of the dark from the bright pupil images. 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, and up to 5m from the camera.

## 1 Introduction

Eye tracking has the potential to become an important component in future computer interfaces. Jacob [11] describe several ways of using eye movements for human-computer interaction. Robust techniques for eye detection are of particular importance to eye-gaze tracking systems. Information about the eyes can also be used to detect 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 one 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 [20, 10] (see Figure 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 [9, 1, 18] 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 [5], skin color [8, 13, 16, 18], geometric models and templates [1, 15, 19], machine learning and artificial neural networks techniques [2, 14, 17], etc.

Other techniques are specifically targeted to detect eyes. Kothari and Mitchell [12] 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), so by accumulating along these lines, the center of the iris can be estimated by selecting the bin with highest count (similar to a Hough transform technique). Heuristic rules and a large temporal support are used to filter erroneous candidates.

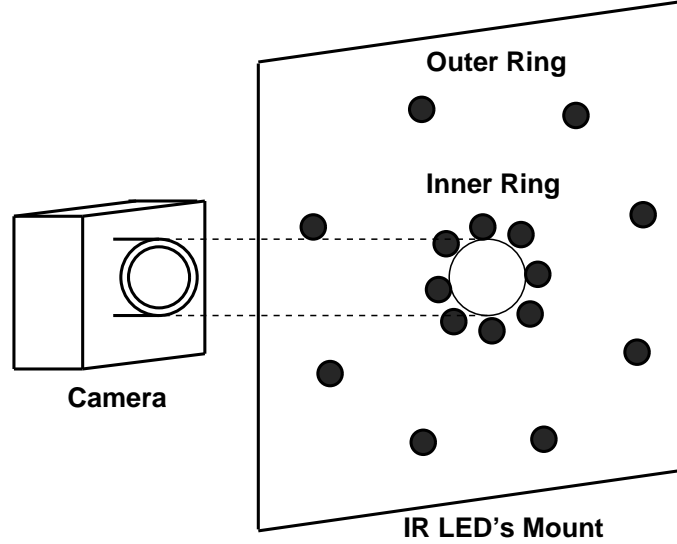
Our detection technique does not require models (color, geometry, templates, examples, etc), and is based on geometrical and physiological properties of the eye. Although special lighting and synchronization schemes are required, the scene background becomes irrelevant and the pupils can be detected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

Ebisawa and Satoh [7] have developed a system very similar to the one presented in this paper. It is based on a differential lighting scheme using two light sources (on and off camera axis) to generate the bright/dark pupil images. The primary application for their system is to track pupils for remote eye-tracking systems, which also requires that the corneal reflection created by the light sources to be also detected, requiring a narrow field of view camera since the reflection is in general very small. In a more recent publication, Ebisawa [6] 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 different geometric disposition of the illuminators, so that both sources generate similar corneal reflections and also increase the operational field of view. 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 they 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 camera and two light sources. For convenience, we use near infra red light sources (875 nm), which are almost invisible to the human eye, and a camera which is sensitive to that wavelength. Figure 1 shows a diagram of the configuration of the camera and the illuminators.



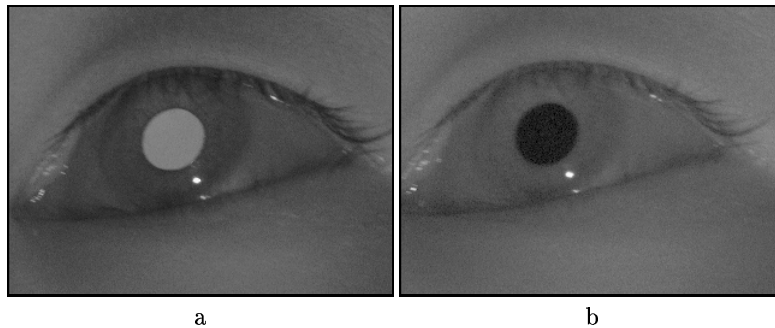
**Fig. 1.** Camera and infra red illuminators

The illuminators consist of 16 IR LED's (HDSL-4220) distributed around two concentric rings (8 LED's in each ring). 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 (15mm), and the diameter of the outer ring is about 90mm. These values were obtained empirically, and are very dependent on the camera and type of lens mount.

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 LED's 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. The outer ring was actually configured to generate slightly brighter images, which facilitates the segmentation process.

For the purpose of eye tracking for gaze estimation, it is also important to detect and track the glint (corneal reflection) [3, 4, 10]. Depending on the illumination conditions and position of the glint, it is often difficult to detect the glint in the bright pupil image. Ebisawa [6] has developed a pupil brightness stabilization 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. Figure 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.



**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

### 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<sup>1</sup>.

<sup>1</sup> Motion artifacts could be considerably reduced by increasing the frame rate.

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 [6]. 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 closeup images of the eye. Hence, for accurate tracking, the pupil must be tracked using the grey 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 grey 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.

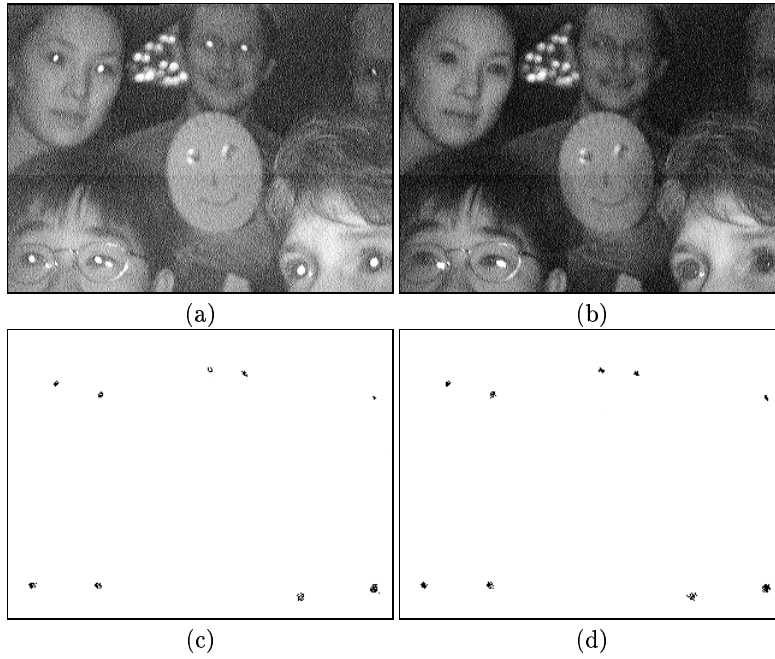
## 4 Experimental Results

A real-time implementation of the pupil detection system is currently running at 15 frames per second 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 synchronized the LED rings with the frame grabber.

The current implementation uses a 1/3" CCD B&W board camera with a micro lens mount (12mm thread), with a 12mm micro lens, and a visible light block filter (780nm). Despite the fact that the micro camera provides images of very low quality, mainly because it does not have controls for the gain, iris, and shutter speed, the system proves to be very robust, being able to detect pupils

from people up to 5 meters from the camera. The system also works well for people with glasses.

Figure 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.

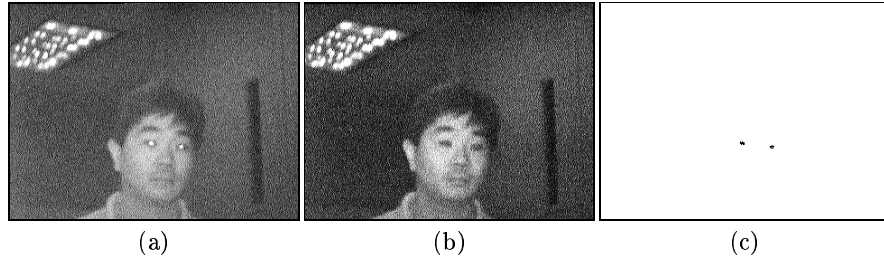


**Fig. 3.** (a) Bright and (b) dark pupil images. (c) and (d) shows the pupils detected from subtraction and thresholding using the corresponding previous frames, after erosion

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

## 5 Conclusion

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.



**Fig. 4.** (a) Bright and (b) dark pupil images; (c) shows the pupils detected using a temporal support of 4 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 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 currently running at 15 frames per second, 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 is 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. S. Birchfield. An elliptical head tracker. In *Proceeding of the 31st Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, CA, November 1997.
2. A.J. Colmenarez and T.S. Huang. Face detection with information-based maximum discrimination. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 782–787, Puerto Rico, PR, June 1997.
3. T.N. Cornsweet and H.D. Crane. Accurate two-dimensional eye tracker using first and fourth purkinje images. *Journal of the Optical Society of America*, 63(8):921–928, August 1973.

4. H.D. Crane and C.M. Steele. Accurate three-dimensional eyetracker. *Journal of the Optical Society of America*, 17(5):691–705, March 1978.
5. T. Darrell, B. Moghaddam, and A. Pentland. Active face tracking and pose estimation in an interactive room. Technical Report 356, M.I.T. Media Laboratory Perceptual Computing Section, Cambridge, MA, 1996.
6. Y. Ebisawa. Unconstrained pupil detection technique using two light sources and the image difference method. *Visualization and Intelligent Design in engineering and architecture*, pages 79–89, 1995.
7. 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.
8. P. Fieguth and D. Terzopoulos. Color based tracking of heads and other mobile objects at video frame rates. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 21–27, Puerto Rico, PR, June 1997.
9. A. Gee and R. Cipolla. Fast visual tracking by temporal consensus. *Image and Vision Computing*, 14(2):105–114, February 1996.
10. T.E. Hutchinson, K.P. White Jr., K.C. Reichert, and L.A. Frey. Human-computer interaction using eye-gaze input. *IEEE Transactions on Systems, Man, and Cybernetics*, 19:1527–1533, Nov/Dec 1989.
11. 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):152–169, April 1991.
12. R. Kothari and J.L. Mitchell. Detection of eye locations in unconstrained visual images. In *Proc. International Conference on Image Processing*, volume I, pages 519–522, Lausanne, Switzerland, September 1996.
13. 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.
14. H. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, San Francisco, CA, June 1996.
15. L.C. De Silva, K. Aizawa, and M. Hatori. Detection and tracking of facial features. In *Proceedings of the SPIE Com. and Image Proc. 95*, volume 2501, pages 1161–1172, Taipei, Taiwan, May 1995.
16. R. Stiefelhagen, J. Yang, and A. Waibel. A model-based gaze tracking system. In *Proceedings of the Joint Symposia on Intelligence and Systems*, Washington, DC, 1996.
17. K. Sung and T. Poggio. Example-based learning for view-based human face detection. Technical Report AI-1521, M.I.T. AI Laboratory, Cambridge, MA, December 1994.
18. 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.
19. A.L. Yiuille, P.W. Hallinan, and D.S. Cohen. Feature extraction from faces using deformable templates. *International Journal of Computer Vision*, 8(2):99–111, April 1992.
20. L. Young and D. Sheena. Methods & designs: Survey of eye movement recording methods. *Behavioral Research Methods & Instrumentation*, 7(5):397–429, 1975.