

A Non-invasive Computer Vision System For Reliable Eye Tracking

Antonio Haro, Irfan Essa
GVU Center / College of Computing
Georgia Institute of Technology
Atlanta, GA 30332-0280
{haro,irfan}@cc.gatech.edu

Myron Flickner
Computer Vision Enhanced User Interfaces
IBM Almaden Research Center
San Jose, CA 95120
flick@almaden.ibm.com

ABSTRACT

Knowing what the user is attending to and what they are looking at is essential for creating attentive user interfaces. Towards this end, we are building a reliable, real-time, non-invasive eye tracker using computer vision. Our system can robustly locate and track eyes without any calibration, and estimate the user's focus of attention. We have built several higher-level processes on top of this tracking system and have done some user studies to test the viability of our approach.

Keywords

Visual tracking, gaze tracking, face tracking, computer vision

INTRODUCTION

Tracking and locating the eyes of a user is an integral part of any attentive user interface. For this purpose, we are developing a real-time, non-invasive eye tracking system [2]. Our system works on multiple people simultaneously, and is able to recover high-level features such as eye blinks and facial regions. Here, we present a system that uses some of these features to monitor a single user's focus of attention by observing their eye movements.

Our eye tracker is an inexpensive (under \$50) black and white camera with structured infrared lighting that works for all people, reliably, and in real-time. It is connected to a consumer-end personal computer for video processing. The system exploits the physiological properties of eyes, their appearance, and their dynamics to provide very reliable detection and tracking.

We have used this system under a variety of conditions to locate eyes, to track eye movements, and to locate multiple users. To validate our system, we undertook simple testing where subjects were asked to sit in front of a computer equipped with our eye-tracker, and

to perform different tasks. In addition to tracking their eyes, we also measured the velocities and accelerations of their eye movements. We found the eye accelerations to be a strong cue for determining a shift in focus as the user changed tasks.

Related Work: Tracking eyes has been a subject of research and development in many fields. There are many hardware devices for eye tracking [4, 3]. Some are mounted on the head directly and have a high degree of accuracy while others utilize structured infrared lighting and precalibration to track where the user is looking on the screen. Our eye tracker is similar in principle to the ones that use structured infrared lighting. However, as we are doing our processing in software, we can add higher-level processing such as face finding and multiple person tracking without increases in hardware. Another difference is that at present, our system does not need calibration to deal with variations in the environment and in appearance of the user's eyes. Our work has a lot in common with the work of Rasmussen, *et al.* [5]. In both works, trackers are given multiple sources of information to update their estimates, and achieve robustness as a result. The main difference is that our system runs in real-time as our models are simpler, without any degradation in performance.

EYE TRACKER

The eye tracker we use consists of a black and white camera with structured infrared lighting, attached to the actual tracking software, which runs in real-time on a consumer-end processor. The tracker is very accurate for smooth head movements, however, it suffers some lack of accuracy for fast head movements. We combine the physiological properties, appearance and dynamics of eyes to track them reliably.

Structured infrared lighting allows us to use the *physiological properties* of the eye, exploiting the *red-eye effect*. In this manner, only regions that reflect light back as the cornea does when light is shining directly on it will be tracked.

This is not enough, however, as there are many objects in complex scenes that exhibit this light reflection property. For example, some shiny surfaces at some orientations can be indistinguishable from eyes if only the



Figure 1: *Our experimental setup*

reflectiveness is used. We use an *appearance model of an eye* to ensure that regions must look like eyes to be tracked.

Another problem that is encountered is that the eyes move dynamically. To model the *dynamics of the eye* and to robustly track it over time, we use Kalman trackers. These trackers allow us to obtain a measure that an observed region is moving like an eye.

These three properties, along with temporal information are combined to yield a reliable eye tracker. Only regions that look, move, and reflect light back as eyes do are tracked¹. The regions that are found provide good starting points for doing higher level processing.

EXPERIMENTS

To measure the robustness of our tracking system and to evaluate the viability of our methods for tracking a shift in attention between tasks, we undertook simple subject studies. Ten users were asked to watch some video segments with breaks. They were told to entertain themselves during the breaks but were not told the purpose of the experiment. Our goal was to determine when the breaks in the video occurred by watching their eyes' movements. We located these breaks from the movement data by looking at points in time where there are clusters of large movements.

The experimental setup used is shown in Figure 1. Our eye tracker system's camera was mounted on the top of the test computer's monitor. The testing video was shown in a window on the test computer's desktop. It consisted of five three minute long segments of video. Three of the five segments were from the evening news, and two were three minute long shots of static black screens. The black screen segments were the second and fourth segments played.

RESULTS

Figure 2 shows the average results of tracking the eyes of the 10 test subjects. The y -axis denotes the acceleration of the eyes, and the x -axis shows the current frame number. Data was captured at 20 frames per second since writing this data to file slowed the tracker down. Note how the segments of time from frame 0 – 3800, 7600 – 11400, and 15200 – 19000 are similar to each other. These correspond to the three news segments that the subjects watched. They are distinctly different from segments 3800 – 7600 and 11400 – 15200 which are

¹See <http://www.gvu.gatech.edu/cpl/projects/pupil/pupil.mov> for tracking results.

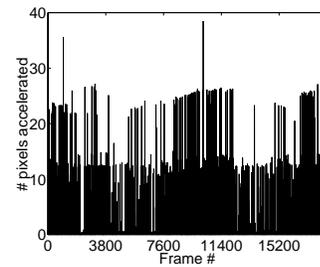


Figure 2: *Average acceleration plot*

the "breaks" where users were allowed to do whatever they wanted. These breaks are not completely devoid of activity, however, since users were allowed to get up, to look away from the machine, surf the web, or to play games (which increased ocular activity).

The portions where the news played always consist of three acceleration bands: one around 0, one around 10, and one around 25. These bands occur because users shift their eyes while watching the news to read the names of the reporters, to look at the background of a crime scene, or to look at the network logo (which some said was distracting). The dip from frame 2232 – 2501 is due to helicopter footage of a suspect's car in the woods; users focused on the car. Further statistical analysis of this data is required to see if there are more relationships in it between adjacent video breaks.

DISCUSSION

Our primary goal is to determine what the users are doing so that the interaction can be effectively supported. In this paper we have shown what can be accomplished by looking at just eye movements. Our eye detection system can extract even higher level behaviors and processes. For example, we can use our system to group eyes into faces which would allow us to do face recognition on multiple people in real-time. Our system can also track eye-blinking. This would allow us to determine if someone is falling asleep by looking at how the rate of their blinking changes [1]. An exciting application is to install such a system in a car to determine driver attention or sleepiness.

REFERENCES

- [1] M. Eriksson and N. Papanikotopoulos. Eye tracking for detection of driver fatigue. In *IEEE Conference on Intelligent Transportation Systems*, pages 314–319, 1997.
- [2] A. Haro, M. Flickner, and I.A. Essa. Detecting and tracking eyes by using their physiological properties, dynamics, and appearance. Submitted to CVPR2000. A copy is available from the Gvu Technical Reports WWW page at <http://www.gvu.gatech.edu/gvu/reports/1999/>.
- [3] LC Technologies Inc. <http://www.lctinc.com/>.
- [4] Applied System Labs. Models 501 and 504. <http://www.a-s-l.com/>.
- [5] C. Rasmussen and G. Hager. Joint probabilistic techniques for tracking multi-part objects. In *CVPR98*, pages 16–21, 1998.