Official Proceedings of

VIRTUAL REALITY SYSTEMS '93

CONFERENCE AND EXHIBITION
MARCH 15-17, 1993
GRAND HYATT HOTEL, NEW YORK, NY

INCLUDES

VIRTUAL REALITY: NEW DIRECTIONS IN HUMAN COMPUTER INTERACTION

CONFERENCE AND EXHIBITION

FEBRUARY 4-5, 1992
Sheraton Hotel and Towers, New York, NY

Published by SIG-ADVANCED APPLICATIONS, INC.

Producers of
VIRTUAL REALITY SYSTEMS '93
Publishers of
VIRTUAL REALITY SYSTEMS MAGAZINE

What You Look At is What You Get: Using Eye Movements as Computer Input

Robert J.K. Jacob

Human-Computer Interaction Lab
Naval Research Laboratory
Washington, D.C.

Introduction

Current user interface technology has been stronger in the computer-to-user direction than user-to-computer, hence today's user-computer dialogues are rather one-sided, with the bandwidth from the computer to the user far greater than that from user to computer. Using eye movements as an input medium can help redress this imbalance. While the technology for measuring a user's visual line of gaze (where he or she is looking in space) and reporting it in real time has been improving, what is needed are appropriate interaction techniques that incorporate eye movements into the user-computer dialogue in a convenient and natural way.

Eye Movement Input in Virtual Reality

Eye movements are a particularly promising medium for use in virtual reality, for hardware reasons as well as because of the complementary relationship between their interaction styles. The essence of virtual reality and other advanced interface approaches is to exploit the user's pre-existing abilities and expectations. Navigating through a conventional computer system requires a set of learned, unnatural commands, such as keywords to be typed in, or function keys to be pressed. Navigating through a virtual environment exploits the user's existing "navigational commands" such as positioning his or her head and eyes, turning his or her body, or walking toward something of interest. By exploiting skills that the user already possesses, advanced interfaces hold out the promise of reducing the cognitive burden of interacting with a computer by making such interactions more like interacting with the rest of the world. Our approach to eye movement interaction relies upon natural eye movements as a source of user input and thus extends this philosophy. Here, too, the goal is to exploit more of the user's pre-existing abilities to perform interactions with the computer.

Characteristics of the Eye

To see an object clearly, it is necessary to move the eyeball so that the object appears on the fovea, a small area at the center of the retina. Because of this, a person's eye position provides a rather good indication (to within the one-degree width of the fovea) of what specific portion of the scene before him he is examining. The most common way of moving the eyes is a sudden, ballistic, and nearly instantaneous saccade. It is typically followed by a fixation, a 200-600 ms. period of relative stability during which an object can be viewed. During a fixation, however, the eye still makes small, jittery motions, generally covering less than one degree.

Eye Tracking Technology

A variety of technologies can be used for measuring eye movements. The simplest eye tracking technique is electronic recording, using electrodes placed on the skin around the eye to measure changes in the orientation of the potential difference that exists between the cornea and the retina. However, this method is more useful for measuring relative eye movements than absolute position. Perhaps the least user-friendly approach uses a contact lens that fits precisely over the bulge at the front of the eyeball and is held in place with a slight suction. This method is extremely accurate, but suitable only for laboratory studies. More practical methods use remote imaging of a visible feature located on the eyeball, such as the boundary between the sclera and iris, the outline of the pupil, or the corneal reflection of a light shone at the eye. All these require the head to be held absolutely stationary (a bite board is customarily used), to be sure that any measured movement represents movement of the eye, not the head. However, by tracking two features of the eye simultaneously, it is possible to distinguish head movements (the two features move together) from eye movements (the two move with respect to one another), and the head need not be rigidly fixed. This is currently the most practical method for use in a user-computer interface. In our laboratory, we use an Applied Science Laboratories (Waltham, Mass.) Model 4250R eye tracker.

Designing Eye Movement-Based Interaction Techniques

The most naive approach to using eye position as an input might be as a direct substitute for a mouse: changes in the user's line of gaze would cause the mouse cursor to move. This is an unworkable (and annoying) approach, because people are not accustomed to operating devices just by moving their eyes. They expect to be able to look at an item without having the look "mean" something. Normal visual perception requires that the eyes move about, scanning the scene before them. It is not desirable for each such move to initiate a computer command. At first, it is empowering simply to look at what you want and have it happen. Before long, though, it becomes like the Midas Touch. Everywhere you look, another command is activated; you cannot look anywhere without issuing a command. The challenge in building a useful eye tracker interface is to design interaction techniques that avoid this Midas Touch problem.

Example of an Eye Movement-Based Interaction Technique: Object Selection

The task here is to select one object from among several displayed. With conventional devices, this is often done by pointing at the object and then pressing a button. With the eye tracker, there is no natural counterpart of the button press. We reject using a blink for a signal because it detracts from the naturalness possible with an eye movement-based dialogue by requiring the user to think about when he or she blinks. We tested two alternatives. In one, the user looks at the desired object then presses a button on a keypad to indicate that the looked-at object is his choice. The second uses dwell time—if the user continues to look at the object for a sufficiently long time, it is selected without further operations. The two techniques can be implemented simultaneously, where the button press is optional and can be used to avoid waiting for the dwell time to expire, much as an optional menu accelerator key is used to avoid traversing a menu.

At first this seemed like a good combination. In practice, however, the dwell time approach is much more convenient. While a long dwell time might be used to ensure that an inadvertent selection will not be made by simply "looking around" on the display, this mitigates the speed advantage of using eye movements for input and also reduces the responsiveness of the interface. To reduce dwell time, we make a further distinction. If the result of selecting the wrong object can be undone trivially (selection of a wrong object followed by a selection of the right object causes no adverse effect—the second selection instantaneously overrides the first), then a very short dwell time can be used. For example, if selecting an object causes a display of information about that object to appear and the information display can be changed instantaneously, then the effect of selecting wrong objects is immediately undone as long as the user eventually reaches the right one. This approach, using a 150-250 ms. dwell time gives excellent results. The lag between eye movement and system response (required to reach the dwell time) is hardly detectable to the user, yet long enough to accumulate sufficient data for our fixation recognition and processing. The subjective feeling is of a highly responsive system, almost as though the system is executing the user's intentions before he expresses them. For situations where selecting an object is more difficult to undo, button confirmation is used.

Checklist

Eye movement input is more useful when:

- Very rapid response is required
- It is important to reduce the user's workload (by exploiting an additional, natural and convenient communication channel)
- The application can make use of position to within one degree of accuracy (i.e., for selecting relatively large objects or areas of a screen)
- A very large display area (real or virtual) is to be traversed
- The user interface is designed to make effective use of the additional information about the user without annoying him or her (this requires careful design of interaction techniques)
- It is acceptable to make an occasional wrong selection if it is immediately followed by the correct selection (e.g., retrieving data, designating a currently selected object for further processing)
- The user's hands are busy (e.g., fighter pilot)
- The user cannot operate other input devices (e.g., disabled user)

Eye movement input is less useful when:

- High accuracy is required (e.g., selecting a single character of small text)
- Very positive indication is required for executing a command (e.g., firing a missile)
- It is not acceptable to require the user to keep his or her head fairly still nor to wear a head-mounted camera (if the user is already wearing a head-mounted virtual reality display, this is not an issue)
- Cost is the driving factor

Biography

Robert Jacob is a Computer Scientist in the Human-Computer Interaction Lab at the Naval Research Laboratory in Washington, D.C., where he is doing research on interaction techniques and user interface software. He received his Ph.D. from Johns Hopkins University and is also on the faculty of George Washington University. He is Vice-Chair of ACM SIGCHI and a member of the editorial board of the forthcoming ACM Transactions on Computer-Human Interaction.