THE CAMERA MOUSE: PRELIMINARY INVESTIGATION OF AUTOMATED VISUAL TRACKING FOR COMPUTER ACCESS

James Gips, Margrit Betke, and Peter Fleming
Computer Science Department
Boston College
Chestnut Hill, MA 02467

ABSTRACT

A system has been developed that uses a camera to visually track the tip of the nose or the tip of a finger or some other selected feature of the body and moves the mouse pointer on the screen accordingly. People without disabilities quickly learn to use the system to spell out messages or play games. People with severe cerebral palsy have tried the system with some initial success. Our goal is to provide computer access to people who are quadriplegic and cannot speak by developing computer vision systems.

BACKGROUND

People who are quadriplegic and nonverbal, for example from cerebral palsy or traumatic brain injury or stroke, have limited motions they can make voluntarily. Some people can move their heads. Some can blink or wink voluntarily. Some can move the eyes or tongue. Family, friends, and other care providers usually detect these motions visually.

Many computer access methods have been developed to help people who are quadriplegic and nonverbal: external switches, devices to detect small muscle movements or eye blinks, head pointers, infrared or near infrared camera based systems to detect eye movements, electrode based systems to measure the angle of the eye in the head, even systems to detect features in EEG. These have helped many people access the computer and have made tremendous improvements in their lives. Still, there are many people with no reliable means to access the computer. We are interested in developing computer vision systems (1) that work under normal lighting to provide computer access to people who are quadriplegic and nonverbal.

STATEMENT OF THE PROBLEM

Develop a system that uses a camera to visually track a feature on a person's face, for example the tip of the nose, and use the movement of the tracked feature to directly control the mouse pointer on a computer.

THE SYSTEM

The system involves two computers: the vision computer, which does the visual tracking, and the user's computer, which runs a special driver and any application software the user wishes to use.

The Vision Computer

The vision computer is a 550 MHz Windows NT machine with a Matrox Meteor-II video capture board. The vision computer receives 30 frames per second from a Sony EVI-D30 camera mounted above or below the monitor of the user's computer. The image used is of size 320 by 240 pixels. The image sequence from the camera is displayed in a window on the vision computer by the visual tracking program.

Initially the operator uses the camera remote control to adjust the pan-tilt-zoom of the camera so that the person's face is centered in the image. The operator uses the mouse to click on a feature in the image to be tracked, perhaps the tip of the user's nose. The vision computer draws a green 15 by 15 pixel square centered on the point clicked and outputs the coordinates of the center of the square. These will be used for the mouse coordinates by the user's computer.

Thirty times per second the vision computer receives a new image from the camera and decides which 15 by 15 square subimage is closest to the previous selected square. The vision computer program examines 400 15 by 15 trial square subimages around the location of the previously selected square. The program calculates the normalized correlation coefficient $r(s,t)$ for the selected subimage s from the previous frame with each trial subimage t in the current frame

$$r(s,t) = \frac{A\sum s(x,y)t(x,y) - \sum s(x,y)\sum t(x,y)}{s_s s_t}$$

where A is the number of pixels in the subimage, namely 225, and

$$s_s = \sqrt{A\sum s(x,y)^2 - (\sum s(x,y))^2} \quad \text{and} \quad s_t = \sqrt{A\sum t(x,y)^2 - (\sum t(x,y))^2}$$

The trial subimage with the highest normalized correlation coefficient in the current frame is selected. The coordinates of the center of this subimage are sent to the user computer. The process is repeated for each frame.

If the program completely loses the desired feature the operator can intervene and click on the feature in the image and that will become the center of the new selected subimage.

The User's Computer

The user's computer is a Windows 98 machine running a special driver program in the background. The driver program takes the coordinates sent from the vision computer, fits them to the current screen resolution, and then substitutes them for the mouse coordinates in the system. The driver program is based on software developed for the EagleEyes system (2), an electrodes based system that allows for control of the mouse by changing the angle of the eyes in the head.

Any commercial or custom software can be run on the user's computer. The visual tracker acts as the mouse. The NumLock key is used to switch from the regular mouse to the visual tracker and back. The user moves the mouse pointer by moving his head (nose) or finger in space.

The driver program contains adjustments for horizontal and vertical "gain." High gain causes small movements of the head to move the mouse pointer greater distances, though with less accuracy. Adjusting the gain is similar to adjusting the zoom on the camera, but not identical.

Many programs require mouse clicks to select items on the screen. The driver program can be set to generate mouse clicks based on "dwell time." With this feature, if the user keeps the mouse pointer within, typically, a 30 pixel radius for, typically, 0.5 second a mouse click is generated by the driver and received by the application program.

RESULTS

The tracking program works extremely well. The program tracks a person's nose for many minutes without adjustment or intervention. No lighting changes were made in the lab, which has standard overhead fluorescent bulbs. Occasionally the selected subimage creeps along the user's face, for example up and down the nose as the user moves his head. This is hardly noticeable by the user as the movement of the mouse pointer still corresponds closely to the movement of the head.

A person without disabilities has good control very quickly. A person can sit down and spell out a message on an onscreen keyboard after just a minute of practice. Using 0.5 seconds dwell time spelling proceeds at approximately 2 seconds per character, 1.5 seconds to move the pointer to the square with the character and 0.5 seconds to dwell there to select it. People spell out entire messages without intervention by the operator.

We have tried the system with three teenagers with severe disabilities. Two of the teenagers used to have no head control but have had a baclofen pump implanted in the past year to reduce muscle spasticity. They now have some head control and are able to move the cursor around but not yet reliably. One teenager is able to move the cursor at will by moving her head.

We have been working with Rick Hoyt, who was born with severe cerebral palsy. Rick has some voluntary head movement, especially to the left. He and his brother developed an easy to use and increasingly popular spelling system based on just a "yes" movement. We have implemented the spelling system in a computer program (3). When combined with this tracker, messages can be spelled out just by small head movements to the left or right using the Hoyt spelling method.

DISCUSSION

Our current system does not use the tracking history. The subimages in the new frame are compared only to the selected subimage in the previous frame and not, for example, to the original subimage. We plan to investigate methods that would compare the current subimages with past selected subimages, for example using recursive least squares filters or Kalman filters (4).

We are just beginning clinical work with the tracking system. We will invite more people with severe disabilities to try the system. People for whom the system seems appropriate will continue working with it so that we can help them better access the computer and also so we can try to optimize the performance of the system. We will work with Rick Hoyt so he can use the tracking system to spell out messages on the computer using his own spelling method.

Our larger plan is to develop systems to visually recognize the facial movements – head movements, blinks and winks, tongue movements, eye movements (5) – that people with quadriplegia can make so we can provide computer access to as many people as possible. We hope the visual tracker is interesting and useful in itself and a first step in this larger project.

REFERENCES

1. Essa IA, "Computers Seeing People." AI Magazine, Summer 1999, 69-82.
2. DiMattia P, Curran FX, and Gips J, EagleEyes: Emancipating the Intellect of Learners with Severe Disabilities. Edwin Mellen Press, in press. See also http://www.bc.edu/eagleeyes
3. Gips J and Gips J, "A Computer Program Based on Rick Hoyt's Spelling Method for People with Profound Special Needs." Submitted.
4. Haykin S, Adaptive Filter Theory, 3rd edition. Prentice Hall, 1995.
5. Betke M and Kawai J, "Gaze Detection via Self-Organizing Gray-Scale Units." Proceedings of The International Workshop on Recognition, Analysis, and Tracking of Faces and Gestures. IEEE Press, 1999, 70-76. See http://oak.bc.edu/~betke/research-group/default.html

ACKNOWLEDGMENTS

This research has been supported in part by NSF grant 9871219.

James Gips, Computer Science Dept., Fulton Hall 460, Boston College, Chestnut Hill, MA 02467
gips@bc.edu http://www.cs.bc.edu/~gips