# The Camera Mouse: Visual Tracking of Body Features to Provide Computer Access for People With Severe Disabilities

Margrit Betke, *Member, IEEE,* James Gips, *Member, IEEE,* and Peter Fleming

***Abstract*—The "Camera Mouse" system has been developed to provide computer access for people with severe disabilities. The system tracks the computer user's movements with a video camera and translates them into the movements of the mouse pointer on the screen. Body features such as the tip of the user's nose or finger can be tracked. The visual tracking algorithm is based on cropping an online template of the tracked feature from the current image frame and testing where this template correlates in the subsequent frame. The location of the highest correlation is interpreted as the new location of the feature in the subsequent frame. Various body features are examined for tracking robustness and user convenience. A group of 20 people without disabilities tested the Camera Mouse and quickly learned how to use it to spell out messages or play games. Twelve people with severe cerebral palsy or traumatic brain injury have tried the system, nine of whom have shown success. They interacted with their environment by spelling out messages and exploring the Internet.**

***Index Terms*—Assistive technology, communication device, real-time tracking, vision-based human–computer interface.**

## I. Introduction

PEOPLE who are quadriplegic and nonverbal—for example, from cerebral palsy, traumatic brain injury, or stroke—often have great difficulties communicating their desires, thoughts, and needs. They use their limited voluntary motions to communicate with family, friends, and other care providers. Some people can move their heads. Some can blink or wink voluntarily. Some can move their eyes or tongue. Assistive technology devices have been developed to help them use their voluntary movements to control computers. People with disabilities can then communicate through spelling or expression-building programs. This allows users to exhibit their thoughts, emotions, and intellectual potential. Along with the increased ability to communicate, users with severe disabilities can benefit from computer access in many other ways. They can acquire knowledge more actively, partake in increased recreational activities, use the Internet, and access computer-controlled technologies such as automated wheelchairs.

Assistive technology systems that use switches to control a computer have been used for a considerable period and are still popular [1]. For entering text and other data into a computer, hitting the switch initiates a scan through a matrix of letters, symbols, words, or phrases. Each matrix entry can be selected with a sequence of switch operations. Current research in this area focuses on dynamically adapting matrix row and column scan delays in order to increase the individual user's text entry rate without complicating the visual display [2]. This is an important issue, since the inability to communicate at an effective rate is a serious barrier for people with disabilities [3].

People with severe disabilities who retained the ability to rotate their heads have other assistive technology options. For example, there are various commercial mouse alternatives. Some systems use infrared emitters that are attached to the user's glasses, head band, or cap.[1] Other systems place the transmitter over the monitor and use an infrared reflector that is attached to the user's forehead or glasses, e.g., [4].[2] The user's head movements control the mouse cursor on the screen. Mouse clicks are generated with a physical switch or a software interface. Evans *et al.* recently described a head-mounted infrared-emitting control system that is a "relative" pointing device and acts like a joystick rather than a mouse [5]. Chen *et al.* developed a system that contains an infrared transmitter, mounted onto the user's eyeglasses, a set of infrared receiving modules that substitute the keys of a computer keyboard, and a tongue-touch panel to activate the infrared beam [6]. Helmets, electrodes, goggles, and mouthsticks are uncomfortable to wear or use. Commercial head-mounted devices can often not be adjusted to fit a child's head. Most important, some users, in particular young children, dislike to be touched on their face and vehemently object to any devices attached to their heads. A noncontact, infrared-based system that tracks the reflected laser speckle pattern of skin is proposed by Reilly and O'Malley [7].

Other commercial systems that allow people with severe disabilities access to a computer are based on measuring corneal reflections.[3] Such systems determine gaze direction by com-

Manuscript received August 22, 2000, revised June 19, 2001. This work was supported by the National Science Foundation under Grants 9871219 and IIS-0093367.

M. Betke is with the Computer Science Department, Boston University, Boston, MA 02215 USA.

J. Gips is with the Computer Science Department, Boston College, Chestnut Hill, MA 02167 USA.

P. Fleming is with Accenture, Wellesley, MA 02481 USA.

Publisher Item Identifier S 1534-4320(02)05848-5.

[1]For example, Penny & Giles HeadWay, Infrared Head-Mounted Mouse Alternative, Don Johnston, Inc., http://www.donjohnston.com.

[2]See also Madentec, http://www.madentec.com.

[3]Eye-Trace System, Permobil Meditech AB, Timra, Sweden, http://www.algonet.se/~eyetrace; Applied Science Laboratories, Bedford, MA, http://www.a-s-l.com; Eyegaze System, LC Technologies, http://www.lctinc.com.

paring the pupil position in an image of a user's eye with the light pattern that occurs when incident light is reflected from the convex surface of the cornea [8]–[10]. Corneal reflection systems have the disadvantages that they need careful calibration, require the user to keep his or her head almost completely still, and are not inexpensive. For example, the Permobil Eye Tracker, which uses goggles containing infrared light emitters and diodes for eye-movement detection, costs between $9900 and $22 460. Recent research advances promise less expensive gaze-tracking solutions [11]. Users with severe disabilities are often not able to keep their heads still enough to use commercial gaze trackers reliably. Chin rests are then used, but they are uncomfortable. In addition, any interruption requires recalibration. The calibration process is too difficult to understand for very young children.

Other control devices measure the electrooculographic potential (EOG) to detect eye movements [8], [12], [13] or analyze features in electroencephalograms [14], [15]. Gips *et al.* [13], [16] have designed "EagleEyes," an EOG-based system that enables people who can move their eyes to control the mouse. Electrodes are attached on the user's face to measure changes in EOG that occur when the position of the eye relative to the head changes. Amplified voltages are translated into the position of the cursor on the screen. The testimonies in [13] show that EagleEyes has made tremendous improvements in children's lives. Still, there are some children who do not want the electrodes to be attached to their faces. Another disadvantage of EOG-based systems is that electrodes can fall off when the user perspires.

Given people's experiences with currently available assistive technology devices, our goal has been to develop a nonintrusive, comfortable, reliable, and inexpensive communication device that is easily adaptable to serve the special needs of quadriplegic people and is especially suitable for children. To attain this goal, we have developed a visual tracking system that interprets visible-light video to provide computer access for people, in particular, children with severe disabilities. Preliminary approaches to video-based computer interfaces for people with disabilities are described in [17]–[20], including our initial work [21], [22]. We call our system the "Camera Mouse" because it tracks a body feature—for example, the tip of the nose—with a video camera and uses the detected motion of the feature to directly control the mouse pointer on a computer. Various choices of features, especially facial features and different computer applications, have been tested for subjects with and without disabilities. Our experiences are very encouraging. Adults and children with severe cerebral palsy or traumatic brain injury have successfully used our system to spell out words or play with educational software.

When compared with other control devices that help people with severe disabilities access a computer, our system's strengths are comfort (no body attachments), ease-of-use (no calibration), and generality (it tracks various body features). Its general tracking ability leads to increased functionality for people who cannot make voluntary head movements but can control, for example, foot movements. Since the Camera Mouse focuses on the motions that a person can make most comfortably, it has the potential to support a high communication rate and result in less perceived exertion.

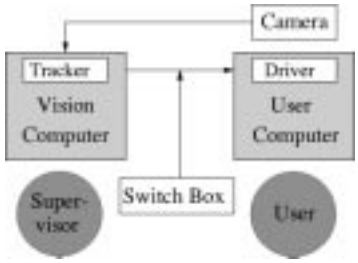


Fig. 1. The Camera Mouse system.

In the near future, standard desktop computers will be equipped with cameras. This will give rise to a new generation of assistive technologies that do not involve customized, expensive electromechanical devices to accomodate special access needs but instead are *software based*. This will dramatically reduce costs and improve availability of assistive technologies [23]. Assistive software, such as presented here for the Camera Mouse system, can make every camera-equipped off-the-shelf computer accessible and therefore give persons with severe disabilities more flexibility. Another advantage of software-based assistive technology is that it is preferred by computer users with disabilities because it draws less attention to their disability than user-borne accessories, such as helmets. Voice recognition software falls into this category and could be applied in combination with our Camera Mouse system by computer users who retained the ability to utter sounds or even speak. In this paper, we report experiences of 12 *nonverbal* people with severe disabilities who have tried to access the computer using the Camera Mouse system. Nine users have shown success.

## II. System Overview

The Camera Mouse system currently involves two computers that are linked together—a "vision computer" and a "user computer." A schematic plan of the system is shown in Fig. 1. The vision computer executes the visual tracking algorithm and sends the position of the tracked feature to the user computer. The user computer interprets the received signals and runs any application software the user wishes to use. The functionalities of the two computers could be integrated into one computer, but the current setup assures sufficient processing power for the visual tracking and allows a supervisor to monitor the tracking performance without interrupting the user's actions.

### A. Vision Computer

The vision computer receives and displays a live video of the user sitting in front of the user computer. The video is taken by a camera that is mounted above or below the monitor of the user computer. Watching this video, the user or an attending care provider clicks with the vision computer's mouse on the feature in the image to be tracked, perhaps the tip of the user's nose. The camera's remote control can be used to initially adjust the pan and tilt angles of the camera and its zoom so that the desired body feature is centered in the image. The vision system determines the coordinates of the selected feature in the initial image and then computes them automatically in subsequent images.

(a)

(b)

Fig. 2. (a) Thirty-month-old Camera Mouse user with cerebral palsy. (b) Visual tracking interface of the vision computer.

Fig. 3. The Camera Mouse user playing with educational software. The video camera is placed underneath the user computer's monitor.

The coordinates of the tracked feature in each image frame are sent to the user computer.

Fig. 2 shows a 30-month-old user of the Camera Mouse system and her tracked face on the monitor of the vision computer. Here the vision algorithm is tracking her lower lip. Fig. 3 shows her in front of the user computer with the camera placed underneath the user computer's monitor.

### *B. User Computer*

The user computer executes any commercial or custom software application the user chooses. It runs a special driver program in the background that takes the signals received from the vision computer, scales them to $x$, $y$ coordinates in the current screen resolution, and then substitutes them for the coordinates of the cursor. The driver program is based on software developed for the EagleEyes system [13] and runs independently from the user's chosen application. The Camera Mouse acts as

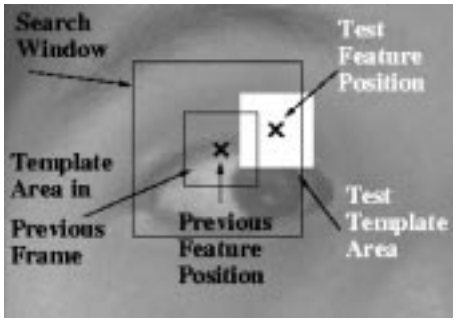


Fig. 4. The search window is shown as a large square. It is centered around the estimated location of the feature in the previous frame. In the current frame, the template is shifted through the window. At each spatial lag, its correlation with the underlying subimage is computed. The location of a test subimage is shown in white. The best correlated subimage is used as the new template in the subsequent frame. Its center is the new estimate of the feature's position.

the mouse, and a manual switch box, shown in Fig. 1, is used to toggle to the standard mouse and back.

The user moves the mouse pointer by moving his or her nose or any other selected feature in space. The driver program contains adjustments for horizontal and vertical "gain" factor. A high gain factor causes small movements of the head to move the mouse pointer greater distances, though with less accuracy. There is a similarity between adjusting the gain and adjusting the camera's zoom. When the gain is adjusted, the coordinates of the feature being tracked are scaled by the specified amount. Changing the zoom of the camera, however, causes the vision algorithm to track the desired feature with either less or more detail. If the camera is zoomed in on a feature, the feature will encompass a greater proportion of the sceen and thus small movements by the user will display larger movements of the cursor. Conversly, if the camera is zoomed out, the feature will encompass a smaller proportion of the screen and thus larger movements will be required to move the cursor.

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." When this option is selected, a mouse click is generated by the driver and received by the application program if the user keeps the mouse pointer within a 30-pixel radius for 0.5 s. These radius and timing parameters are typical choices and can be adjusted according to application needs. The driver program also has an optional exponential smoothing filter. This can be used to smooth out tremors in the feature being tracked.

## III. Tracking Algorithm

When the user initially clicks on the feature to be tracked, a square is drawn around the feature and the subimage within this square is cropped out of the image frame. The cropped subimage is used as a "template" to determine the position of the feature in the next image frame. To find this position, the tracking algorithm uses the template to search for the feature in a "search window" that is centered at the position of the feature in the previous frame. The template is shifted through this search window and correlated with the underlying subimages. The window is defined to contain the centers of all subimages tested. Fig. 4 illustrates template and search window positions.

Since a new frame is received within a thirtieth of a second, the template is usually very similar to the brightness pattern of the feature in the new frame, which can be found by searching for the best correlated subimage. The assumption that corresponding brightness patterns in subsequent frames are constant, the "constant brightness assumption," is often made when designing algorithms for motion analysis in images [24], [25].

Each subimage $s(x, y)$ in the search window is matched with the template subimage $t(x, y)$ that is cropped from the previous frame. As a measure of match, the normalized correlation coefficient

$$r(s,t) = \frac{A\sum s(x,y)t(x,y) - \sum s(x,y)\sum t(x,y)}{\sigma_s \sigma_t} \quad (1)$$

is used, where $\sigma_s = \sqrt{A\sum s(x,y)^2 - \left(\sum s(x,y)\right)^2}$ and $\sigma_t = \sqrt{A\sum t(x,y)^2 - \left(\sum t(x,y)\right)^2}$, and $A$ is the number of pixels in template $t(x, y)$. The normalized correlation coefficient is invariant to uniform variations in shading, i.e., $r(s,t) = r(as + b, t)$ for some constants $a$ and $b$. It can therefore handle some violations of the constant brightness assumption.

The subimage with the highest correlation coefficient among all subimages in the search window is determined. It serves two purposes. First, its center coordinates are transfered to the user computer to be interpreted as cursor coordinates. Second, it is cropped from the current frame and becomes the template that is used to search for the best matching subimage in the next frame. The process repeats and the template is updated for each image frame. This updating of the template ensures that a strong match can be produced in each frame and that the motion of the initially selected feature is tracked. If at any time a low correlation is obtained and the template no longer resembles the desired feature, the supervising operator is free to reclick on the feature on the screen, thereby updating the template manually.

## IV. Choice of Parameters

The tracking performance of the Camera Mouse is a function of template and search window sizes and the velocity of the feature's motion. It also depends on the choice of the feature, as will be shown in Section VI, and the speed of the vision computer's processor. A large *search window* is useful for finding a feature that moves quickly. A large *template* size is beneficial because it provides a large sample size for determining the sample means and variances in the computation of the normalized correlation coefficient. Small templates are more likely to match with arbitrary background areas because they often do not have enough brightness variations, e.g., texture or lines, to be recognized as distinct features. Reference [26] discusses this phenomenon in detail and explains that the size of the template is not the only issue, but more importantly, tracking performance depends on the "complexity" of the template.

Large template or search window sizes require computational resources that may reduce the frame rate substantially. If many incoming frames are skipped, which means that the rate of the frames that are used for tracking drops well below 30 Hz, the constant brightness assumption may not hold for the tracked feature, even if it is still located within the search window. For the worse, when frames are skipped, it is likely that the feature moves outside the search window, far away from its previous position.

To quantify tracking performance, a match between a template and the best-matching subimage within the search window is called *sufficient* if the normalized correlation coefficient is at least 0.8. Correlation coefficients below 0.8 describe *insufficient matches*. Insufficient matches occur when the feature cannot be found in the search window because the user moved quickly or moved out of the camera's field of view. This results in an undesired match with a feature that is different from the initially selected feature. For example, if the right eye is being tracked and the user turns his or her head quickly to the right, so that only the profile is seen, the right eye becomes occluded. A nearby feature, for example, the top of the nose, may then be cropped and tracked instead of the eye.

The threshold of 0.8 was chosen after extensive experiments that resulted in an average correlation of 0.986 over 800 frames for a match between template and best correlated subimage, while the correlation for poor matches varied between 0.7 and 0.8. If the correlation coefficient is above 0.8, but considerably less than one, the initially selected feature may not be in the center of the template anymore and attention has "drifted" to another nearby feature. In this case, however, tracking performance is usually sufficient for the applications tested.

Fig. 5 illustrates how an increase in the width of the search window affects the frame rate and the correlation between template and best-matching subimages. In these experiments, the tip of the computer user's nose was being tracked while the user tried to move the screen pointer at a consistent speed. On average, the tracked feature changed its position by 7 pixels in 1/30 s. Nine parameter values for the search window width were tested. For each value, 800 image frames were captured and processed. The averages and standard deviations reported in Fig. 5 are computed over these 800 frames. As can be seen in the bottom graph, the number of insufficient matches among 800 processed frames is zero for window widths below 44 pixels. If the window width is set to 44 or more pixels, the frame rate drops to about 20 Hz (see top graph). The correlation coefficient of the best match then drops, the standard deviation of the best match increases (see middle graph), and several insufficient matches are found (see bottom graph).

To find good parameter values for window and template sizes that balance the tradeoff between the number of frames examined per second and the sizes of the areas searched and matched, the time it takes to search for the best correlation coefficient is measured as a function of window and template widths (see Fig. 6). A template size of $15 \times 15$ pixels is large enough to capture a meaningful body feature and small enough so that a $40 \times 40$ pixel window can be searched at a frame rate of 30 Hz.

## V. Hardware Specifications

### A. Vision Computer

The vision computer is a 550-MHz Pentium II PC with the Windows NT operating system, a Matrox Meteor-II video capture board, and a National Instruments Data Acquisition Board. The video capture board digitizes the analog NTSC

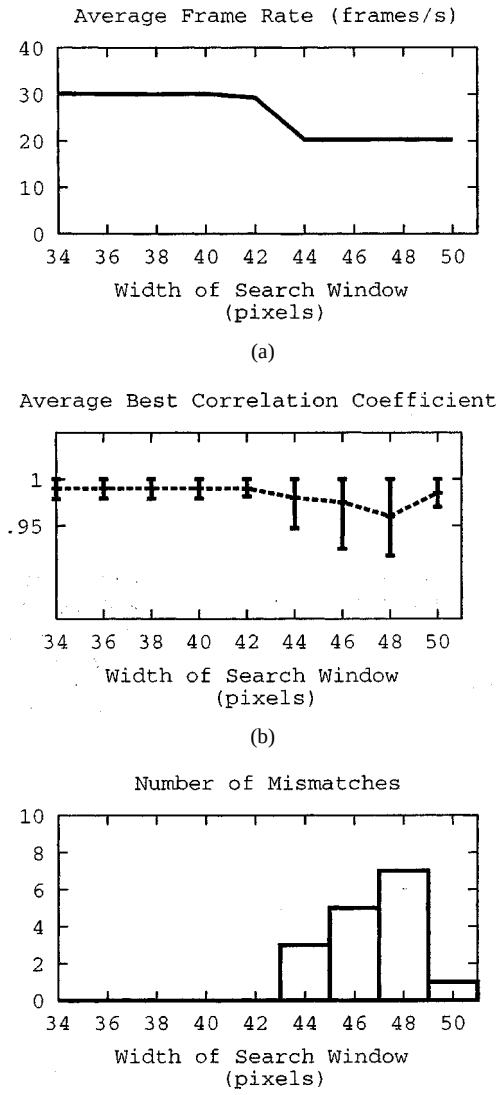


Fig. 5. Tracking experiments to determine the search window size. The tip of the computer user's nose was being tracked while the user moved the screen pointer 7 pixels in 33 ms on average. Nine parameter values for the search window width were tested. For each value, 800 image frames were captured and processed. The averages and standard deviations are taken over the 800 processed frames. An increase in the width of the search window affects the average frame rate (a) and mean and standard deviations of the correlation between template and best matching subimage (b). While the frame rate is close to 30 Hz, the best matching templates correlate highly and the tracking algorithm is able to reliably capture the motion of the feature. However, once the search size becomes large enough to affect the frame rate, the tracking performance suffers and insufficient matches are made (c).

signal received from the video camera, a Sony EVI-D30 CCD camera with zoom, tilt, and pan mechanisms. The video capture board supplies 30 color images of size $640 \times 360$ per second. To guarantee real-time performance, the vision algorithm processes only $320 \times 240$ pixels, i.e., half of the available resolution, at an average frame rate of 30 Hz. It uses a template size of $15 \times 15$ pixels and a search window of $40 \times 40$ pixels. The vision algorithm computes the $x$, $y$ coordinates of the tracked feature and passes them to the National Instruments Data Acquisition Board. This board transforms the coordinates into voltages that are then sent to the user computer.

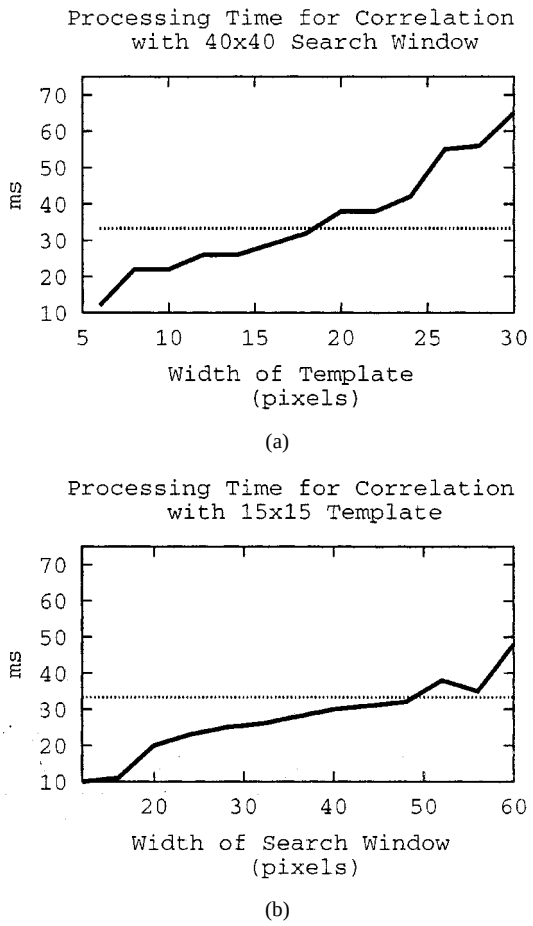


Fig. 6. (a) Processing time of the normalized correlation coefficient as a function of the width of the template, given a search window width of 40 pixels. (b) Processing time of the normalized correlation coefficient as a function of the width of the search window, given a fixed $15 \times 15$ size template. A frame rate of 30 Hz $=$ 33 ms can be ensured if the respective sizes of the cropped template and the search window are $15 \times 15$ and $40 \times 40$ pixels.

### *B. User Computer*

The user computer is a 550-MHz Pentium II PC with the Windows 98 operating system and a National Instruments Data Acquisition Board. A special driver program uses the National Instruments card to obtain the voltages and then converts them into screen coordinates for the mouse pointer. The driver program runs in the background and acts as a regular mouse for any customized or commercial application software.

## VI. Discussion of Feature Choices

A variety of features were tested in an attempt to find body points that can be easily moved by the user and reliably tracked by the system. Since the appearance of body features differs among people, tracking performance varies between individuals. Several features, however, were reliably tracked across the test group. Fig. 7 illustrates the tracking performance of several features, while Fig. 8 illustrates the brightness variations of the template images.

### *A. Nose Tracking*

The nose is a desirable tracking feature for several reasons. First, it is easy for a computer user to point his or her nose in a


Fig. 7. The eye, nose, mouth, and thumb were tracked to determine features that can be tracked reliably (A–E). Sequence F shows that the nose can be tracked even if the laboratory is dark and only the screen illuminates the user's face. Sequence G shows tracking of a set of features. Each test sequence contains 1000 images. A representative subset of each sequence is shown. The tracked subimages surrounding the features have been filled white for better viewing. They are 15 pixels wide in sequences A–F and 10 pixels wide in sequence G.

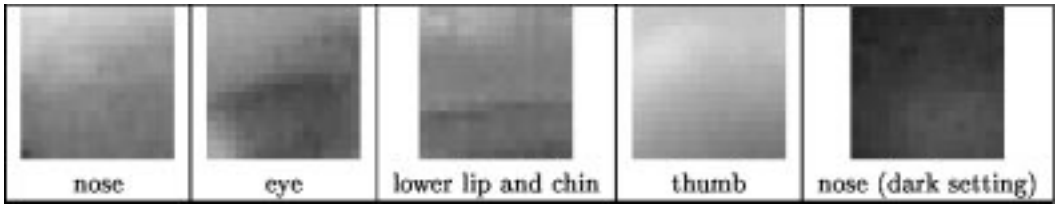


Fig. 8. Examples of the templates that are used to identify the feature being tracked. Each template is $15 \times 15$ pixels in size. Here they have been enlarged in order to view the brightness levels that make up each template. They are templates of the nose, the eye, the lower lip and crevice under the lip, the thumb, and the nose in a room without any light except for that reflected by the monitor.

particular direction while watching the screen. The nose is essentially in the center of the face and does not become occluded when the user's head moves significantly. Second, the nose template tends to contain a good amount of brightness contrast to its surrounding features.

The majority of testing with the Camera Mouse was done under normal overhead lighting. In such an environment, the nose tends to be brighter than the rest of the face, as it is slanted and therefore oriented toward the overhead light.

In image sequence A in Fig. 7, the tip of the nose was tested as a feature. It was tracked throughout the 1000 recorded frames and not lost once. This excellent tracking performance can be reproduced for arbitrarily long time periods as long as the user understands the constraints of the system and cooperates accordingly. If he or she moves so quickly that the nose tip leaves the search window, the system cannot track it. Given the selection of parameters in Section IV, this can only occur with a very drastic motion, for example, a violent shaking of the head or an extremely jerky movement. The system also loses the nose feature if the user covers the nose or moves out of the camera's field of view. The drifting phenomenon, as discussed in Section IV, may occur for some users. For example, the template may slowly drift up the bridge of the nose. For such users, the bottom part of the nose, i.e., the area between the nostrils, is a more useful feature. It works better because the neighboring nostrils provide good contrast points and the shadow that is generally present under the nose distinguishes the bottom part of the nose from the cheeks or lips.

### *B. Eye Tracking*

There has been some success in tracking the eye, but not to the extent of determining gaze direction. Image sequence B in Fig. 7 shows the eye being tracked at various positions. Note that it is not the pupil but the whole eye that is being tracked in sequence B. The brightness contrast between white eye sclera

and dark iris and pupil, along with the texture of the eyelid, provides a distinctive template. Although the eye can be tracked well, it has not been used effectively with a Camera Mouse application because it is a relatively difficult feature to move while viewing the screen. Also, rotating the head may cause the eye to be blocked by the nose and not be visible at all.

### C. Lip Tracking

As shown in image sequences C and D in Fig. 7, tracking the area of the lower lip and cleft has also been tested extensively. This feature can be tracked successfully on many individuals. It is a good tracking location because of the brightness difference between the lip and the cleft. People with large lips tend to have a shadow cast upon the cleft that enhances the brightness difference between the upper and lower image portions of the template and makes vertically drifting templates very unlikely. Furthermore, the outline of the lip forms a curved line that helps control lateral drifting and keeps the template centered on the lips.

When testing this feature in an attempt to track head movement, we noticed that it also works very well in tracking mouth movement alone. This is an important result because it stresses the flexibility of the system. The range of muscle control varies widely between people with severe disabilities, and head movement is not always an option. Opening and closing the mouth is a possibility for many people, though, and can allow cursor motion in either a vertical or a horizontal direction. For example, the *Rick Hoyt Spelling Program* [27] requires only horizontal cursor movements that can be produced by converting the tracked up–down movements of the lips into horizontally changing cursor coordinates.

### D. Thumb Tracking

To test other body features, not just facial features, the thumb was selected. Although it was tracked successfully, as shown in sequence E in Fig. 7, it has two main flaws as a tracking point. First, the camera has difficulties in focusing on it. As can be seen in sequence E, the thumb takes up such a small portion of the screen that the camera's autofocus mechanism focuses on the objects in the background and not the thumb. This distorts the outline of the thumb and makes it difficult to track. Thus, if the thumb is used as a tracking point, it should be held close to the body or some other object in the background, so that the camera is able to focus on the thumb correctly. Another problem in using the thumb is its small surface area, which can move out of the search window easily. If this occurs, the tracking program will lose the thumb entirely and begin tracking a new point in the background. This means that if the thumb is used, slow movements are necessary.

### E. Dark Lighting

A set of test images was taken in a room with no natural or artificial light except for what was produced by the computer monitor. As can be seen in sequence F in Fig. 7, the Camera Mouse was able to track the person's nose in these lighting conditions.

Fig. 9. Aliens game: The user controls the position of the circled cross with either the standard mouse or the Camera Mouse. An alien has appeared at random on top of the screen. In order to "catch the alien," the user must move the circled cross up until it touches the alien.

TABLE I
TIMING COMPARISON BETWEEN REGULAR MOUSE AND CAMERA MOUSE USAGE SAMPLE MEAN AND STANDARD DEVIATION $\sigma$ OF THE LENGTHS OF THREE TESTS WITH THE "ALIENS GAME" PERFORMED BY 20 HEALTHY SUBJECTS

| | Aliens Game | | | |
|---|---|---|---|---|
| | Regular Mouse | | Camera Mouse | |
| | Mean | $\sigma$ | Mean | $\sigma$ |
| Test 1 | 0.50 s | 0.07 s | 0.75 s | 0.12 s |
| Test 2 | 0.43 s | 0.07 s | 0.75 s | 0.18 s |
| Test 3 | 0.40 s | 0.07 s | 0.66 s | 0.10 s |

### F. Multiple Points

Although the current system does not track multiple points, some initial experimentation was done to test such tracking. In the future, multiple-point tracking may be utilized, for example, to compute the distances between the nose and pupils for gaze detection. This would result in a more specialized tracking system. The current version's strength is its generality—any feature can be selected for tracking.

## VII. EXPERIMENTAL RESULTS

### A. People Without Disabilities

The effectiveness of the Camera Mouse was tested with a group of 20 computer users in two main experiments. Each user was given a brief introduction to how the system worked and then allowed to practice moving the cursor for one minute. After the practice period, each user was asked to play a video game. The one-minute practice period was perceived as sufficient training time by the users, who prefered to "train on the job" while playing a computer game.

In the computer game, "aliens" appear at random locations on the screen one at a time, as shown in Fig. 9. In order to "catch an alien," users must point the cursor at the alien. No mouse clicks are generated. Each user was asked to "catch ten aliens" three times with the mouse and three times with the Camera Mouse. The type of mouse that was tested first was chosen randomly. For each test, the user's time to play the game was recorded. Table I reports the times to "catch one alien" with the regular mouse and the Camera Mouse as averages for the 20 users. The average time to play the video game with the regular mouse was

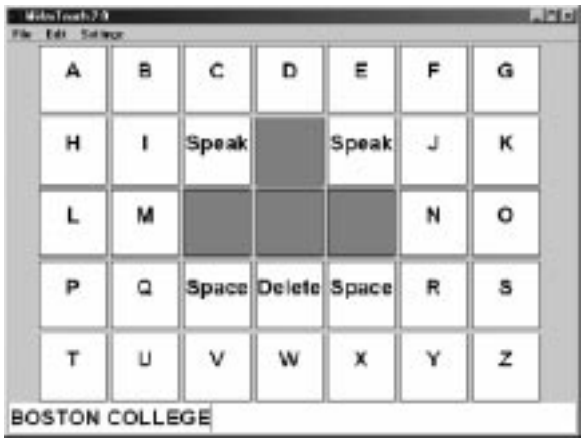


Fig. 10. Spelling board: The user controls what is typed at the bottom of the interface by selecting the letters with the standard mouse or the Camera Mouse. "Boston College" was typed here.

faster than the Camera Mouse by a factor of about 1.6. The difference is statistically highly significant ($P < 0.001$) [28]. Practicing the video game seems to improve input speed, since on average the game was played faster in each consecutive test—both with the standard mouse and with the Camera Mouse.

In a second set of experiments, each user was asked to type with a spelling-board program. The program's interface, shown in Fig. 10, contains 26 cells to type the letters of the alphabet, two cells to type the space symbol, a cell to delete misspelt letters, and two cells to activate the speech synthesizer, which then speaks the just-typed words. To activate a cell, a mouse click is required. If the Camera Mouse is used, a mouse click is simulated after the pointer has occupied the cell for more than 0.5 s. In our experiments, we fixed this "dwell time" for comparison purposes. In general, the dwell time can be adjusted for a particular user. The process of clicking a regular mouse button takes 0.05 s on average. The interface also contains four cells without functionality, which can be used to rest the cursor. These cells are important to avoid the "Midas touch problem" [29] that everything the mouse pointer touches is selected.

In the spelling-board experiment, each user was asked to test the program with the standard mouse and the Camera Mouse three times to spell out and speak "Boston College." The decision of which type of mouse to test first was made randomly. Each user was able to spell the message with the Camera Mouse, which shows that it provides a viable method of controlling the cursor. The timing results of the experiments are summarized in Table II. The regular mouse provides a faster interface with the spelling board than the Camera Mouse. The difference is statistically highly significant ($P < 0.001$).

Spelling with the regular mouse was on average more than twice as fast as the Camera Mouse. In particular, in the third test, the factor is $2.1 = 25.24\ \text{s}/11.90\ \text{s}$. The dwell time required for the Camera Mouse is an important element in this comparison. When the difference in dwell time for the Camera Mouse and click time for the regular mouse is accounted for by excluding these times from the overall lengths of the communication tests, the regular mouse was still faster than the Camera Mouse, but only by a factor of $1.6 = 18.24\ \text{s}/11.20\ \text{s}$. This is the same factor obtained for the "alien game" that did not require any mouse clicks to be generated.

A short dwell time will increase the rate of communication. However, if the dwell time is too short, chances increase that cells are unintentionally selected. Further studies are needed to determine the dwell time that provides the best rate of communication for a particular user. We expect this to be dependent on the user's abilities to control his or her body movements.

### *B. People With Disabilities*

A dozen people who cannot speak and have very limited voluntary muscle control have tried using the Camera Mouse to access the computer. Ten have cerebral palsy. Two have traumatic brain injury: one from an automobile accident and one from a motorcycle accident. Nine of the people are continuing to use the Camera Mouse. Six can use the Camera Mouse to spell using an onscreen keyboard system. Three of the people did not have sufficient muscle control to use the Camera Mouse. They are using EagleEyes to control the mouse pointer by moving their eyes. One of the people who was not successful was close and will be given additional opportunities to try the Camera Mouse in the future. The results are summarized in Table III.

Most users tested the Camera Mouse system at the "Campus School" at Boston College. The Campus School provides a combination of educational and therapeutic services to children ages 3 to 21 who have multiple disabilities including cerebral palsy and traumatic brain injury. Various systems such as switch-based and EOG-based systems [13] provide these children computer access that is reliable but not without barriers. The Camera Mouse has proven to be a useful alternative. It does not require direct physical contact, allows direct selection of choices, and is easier to operate for both supervisor and user. The children were able to master the Camera Mouse system within 2 h of use. For those who have muscular control of a body feature like the head, foot, or hand, the Camera Mouse has been very effective.

There is an important story to be told about each of the dozen individuals listed in Table III. One user is an eight-year-old child with cerebral palsy. He had been using EagleEyes for one year but has had problems with the system due to his perspiration and dislike of the physical contact with the electrodes. Once he tried the Camera Mouse, it was immediately clear that he understood that he was controlling the cursor. It took two hours, spread over a week, to determine the best tracking point for this user. The feature of choice for him is his foot. Using his foot, he is able to consistently "catch ten out of ten aliens" using the game described in Section VII. His foot can be tracked without requiring frequent supervisor intervention.

A three-year-old girl with cerebral palsy has used the system at home regularly. Her mother reports: "When she was 16 months old we had the opportunity to try a computer that was accessed [using EagleEyes]. At this time we had no idea if or how much [she] understood. To our amazement she followed every direction that was given her. Can you just imagine the joy we had watching our daughter exploring her environment? . . . The possibilities seem endless now. Since then [she] has become more sophisticated with this computer. . . . [The

TABLE II
TIMING COMPARISON BETWEEN STANDARD MOUSE AND CAMERA MOUSE USAGE. SAMPLE MEAN AND STANDARD DEVIATION $\sigma$ OF THE LENGTHS OF THREE COMMUNICATION TESTS WITH THE SPELLING-BOARD INTERFACE PERFORMED BY 20 HEALTHY SUBJECTS

| | Spelling Board Experiments | | | | | |
|---|---|---|---|---|---|---|
| | Regular Mouse | | | Camera Mouse | | |
| | Mean Time Including Mouse Clicks | Mean Time Excluding Mouse Clicks | $\sigma$ | Mean Time Including Mouse Clicks | Mean Time Excluding Mouse Clicks | $\sigma$ |
| Test 1 | 15.29 s | 14.59 s | 2.88 s | 28.67 s | 21.67 s | 3.58 s |
| Test 2 | 12.48 s | 11.78 s | 2.23 s | 26.67 s | 19.67 s | 3.32 s |
| Test 3 | 11.90 s | 11.20 s | 1.55 s | 25.24 s | 18.24 s | 2.70 s |

TABLE III
SUMMARY OF RESULTS FOR THE FIRST DOZEN PEOPLE WITH DISABILITIES TO TRY THE CAMERA MOUSE

| Age | Gender | Condition | Results | Continuing to Use ? |
|---|---|---|---|---|
| 2 | M | Cerebral Palsy | Obtaining a system for home. | Yes |
| 3 | F | Cerebral Palsy | First regular user with home system. | Yes |
| 6 | F | Cerebral Palsy | Spelled name. Obtaining a home system. | Yes |
| 8 | M | Cerebral Palsy | Spells naughty words and laughs. | Yes |
| 11 | M | Cerebral Palsy | Obtaining a home system. | Yes |
| 14 | M | Cerebral Palsy | Spells words. Obtaining a home system. | Yes |
| 15 | M | Cerebral Palsy | Close, but could not control reliably. | No |
| 19 | M | Cerebral Palsy | Does not have sufficient muscle control. | No |
| 23 | M | Traumatic Brain Injury | Does not have sufficient muscle control. | No |
| 31 | M | Traumatic Brain Injury | Spelled "TAKE OFF DAD." | Yes |
| 37 | M | Cerebral Palsy | Spelled "MERRY CHRISTMAS." | Yes |
| 58 | M | Cerebral Palsy | Spells, explores internet on home system. | Yes |

Camera Mouse] has given her a way to communicate her thoughts, it gives the school that she is attending a way to adapt the curriculum so that [she] can participate in a REGULAR preschool, it puts [her] in a situation where people can see her ABILITIES rather than her disabilities. . . . When [she] uses the Camera Mouse, she is alert, attentive and responsive. She controls the mouse with her chin and plays with educational software."

Since obtaining the Camera Mouse, the girl has transitioned from an early intervention program to a preschool regular education program. Despite her inability to vocalize and her very limited muscle control, she is functioning at grade level and above in most areas of cognitive and social development. Our goal now is to help her access the regular schoolwork that is afforded her peers.

Another user is 19 years old and has a traumatic brain injury. This user learned how to use the camera mouse quickly and had fluid control of the cursor after ten minutes of use. He does have less control of horizontal movements, though, due to a lack of complete muscular control.

The Camera Mouse system has also been used by a 23–year-old man with traumatic brain injury who has some control of his thumb. By placing his hand on the white tray of his wheelchair and focusing the camera on his thumb, he can move the mouse pointer with small movements of his thumb.

The gentleman who is 58 with cerebral palsy spends his time in bed at home in New Jersey. He had no expressive language ability and no voluntary movement below the neck. We were approached by his 80+-year-old father who wanted his son to be able to communicate before he died. We set up a prototype system in the home. The gentleman is now using the Camera Mouse to communicate with his father and access a variety of software.

## VIII. CONCLUSION

The experiences with the Camera Mouse system are very encouraging. They show that the Camera Mouse can successfully provide computer access for people with severe disabilities. It is a user-friendly communication device that is especially suitable for children. The system tracks many body features and does not have any user-borne accessories, so it is easily adaptable to serve the special needs of people with various disabilities. To meet the current demand, additional Camera Mouse systems are being installed. A single-computer version of the system is being developed. Future work will incorporate a detection component into the visual tracking algorithm.[4]

## ACKNOWLEDGMENT

The authors thank P. A. DiMattia, Director of the Campus School at Boston College, for his support.

## REFERENCES

[1] W. J. Perkins and B. F. Stenning, "Control units for operation of computers by severely physically handicapped persons," *J. Med. Eng. Technol.*, vol. 10, no. 1, pp. 21–23, 1986.

[4]The Camera Mouse technology has been licensed from Boston College by CM Solutions, Inc., of Austin, TX. J. Gips serves as Chairman of the Board of Advisors.

[2] R. C. Simpson and H. H. Koester, "Adaptive one-switch row-column scanning," *IEEE Trans. Rehab. Eng.*, vol. 7, no. 4, pp. 464–473, 1999.
[3] G. Vanderheiden, "Human interface design and the handicapped user," in *Proc. Computer-Human Interaction Conf. (CHI)* , 1986, pp. 296–297.
[4] O. Takami, N. Irie, C. Kang, T. Ishimatsu, and T. Ochiai, "Computer interface to use head movement for handicapped people," in *Proc. IEEE TENCON'96, Digital Signal Processing Applications*, vol. 1, 1996, pp. 468–472.
[5] D. G. Evans, R. Drew, and P. Blenkhorn, "Controlling mouse pointer position using an infrared head-operated joystick," *IEEE Trans. Rehab. Eng.*, vol. 8, no. 1, pp. 107–117, 2000.
[6] Y. L. Chen, F. T. Tang, W. H. Chang, M. K. Wong, Y. Y. Shih, and T. S. Kuo, "The new design of an infrared-controlled human-computer interface for the disabled," *IEEE Trans. Rehab. Eng.*, vol. 7, pp. 474–481, Dec. 1999.
[7] R. B. Reilly and M. J. O'Malley, "Adaptive noncontact gesture-based system for augmentative communication," *IEEE Trans. Rehab. Eng.*, vol. 7, no. 2, pp. 174–182, 1999.
[8] L. Young and D. Sheena, "Survey of eye movement recording methods," *Behav. Res. Meth. Instrum.*, vol. 7, no. 5, pp. 397–429, 1975.
[9] T. Hutchinson, K. P. White Jr., W. N. Martin, K. C. Reichert, and L. A. Frey, "Human-computer interaction using eye-gaze input," *IEEE Trans. Syst., Man, Cybern.*, vol. 19, no. 6, pp. 1527–1533, 1989.
[10] G. A. Rinard, R. W. Matteson, R. W. Quine, and R. S. Tegtmeyer, "An infrared system for determining ocular position," *ISA Trans.*, vol. 19, no. 4, pp. 3–6, 1980.
[11] C. H. Morimoto, D. Koons, A. Amit, M. Flickner, and S. Zhai, "Keeping an eye for HCI," in *Proc. XII Brazilian Symp. Computer Graphics and Image Processing*, 1999, pp. 171–176.
[12] J. R. LaCourse and F. C. Hludik Jr., "An eye movement communication-control system for the disabled," *IEEE Trans. Biomed. Eng.*, vol. 37, no. 12, pp. 1215–1220, 1990.
[13] P. DiMattia, F. X. Curran, and J. Gips, *An Eye Control Teaching Device for Students Without Language Expressive Capacity EagleEyes*. Lampeter, U.K.: Edwin Mellen, 2001.
[14] Z. A. Keirn and J. I. Aunon, "Man-machine communications through brain-wave processing," *IEEE Eng. Med. Biol.*, pp. 55–57, May 1990.
[15] M. Pregenzer and G. Pfurtscheller, "Frequency component selection for an EEG-based brain to computer interface," *IEEE Trans. Rehab. Eng.*, vol. 7, no. 4, pp. 413–419, 1999.
[16] J. Gips, P. Olivieri, and J. J. Tecce, "Direct control of the computer through electrodes placed around the eyes," in *Human-Computer Interaction Applications and Case Studies*, M. J. Smith and G. Salvendy, Eds. Amsterdam, the Netherlands: Elsevier, 1993, pp. 630–635.
[17] H. R. Arabnia, "A computer input device for medically impaired users of computers," in *Proc. Johns Hopkins National Search for Computing Applications to Assist Persons With Disabilities*, 1992, pp. 131–134.
[18] H. Sako, M. Whitehouse, A. Smith, and A. Sutherland, "Real-time facial-feature tracking based on matching techniques and its applications," in *Proc. 12th IAPR Int. Conf. Pattern Recognition*, vol. 2, Conference B: Computer Vision & Image Processing, 1994, pp. 320–324.
[19] O. Takami, K. Morimoto, T. Ochiai, and T. Ishimatsu, "Computer interface to use head and eyeball movement for handicapped people," in *IEEE Int. Conf. Systems, Man and Cybernetics Intelligent Systems for the 21st Century*, vol. 2, 1995, pp. 1119–1123.
[20] T. Ohtani, H. Ichihashi, T. Miyoshi, and N. Tani, "A pointing device using coordinate transformation of neurofuzzy GMDH," in *Proc. 1998 2nd Int. Conf. Knowledge-Based Intelligent Electronic Systems (KES'98)*, vol. 2, 1998, pp. 108–115.
[21] J. Gips, M. Betke, and P. Fleming, "The Camera Mouse: Preliminary investigation of automated visual tracking for computer access," in *Proc. Rehabilitation Engineering and Assistive Technology Society of North America 2000 Annu. Conf. (RESNA)*, Orlando, FL, July 2000, pp. 98–100.
[22] J. Gips, M. Betke, and P. A. DiMattia, "Early experiences using visual tracking for computer access by people with profound physical disabilities," in *Universal Acccess in HCI Toward an Information Society for All*, C. Stephanidis, Ed. Mahwah, NJ: Lawrence Erlbaum, 2001, vol. 3, pp. 914–918.
[23] C. Brown, "Assistive technology computers and persons with disabilities," *Commun. ACM*, vol. 35, no. 5, pp. 36–45, 1992.
[24] B. K. P. Horn, *Robot Vision*. Cambridge, MA: MIT Press, 1986.
[25] M. Betke, E. Haritaoglu, and L. S. Davis, "Real-time multiple vehicle detection and tracking from a moving vehicle," *Mach. Vis. Appl.*, vol. 12, no. 2, pp. 69–83, September 2000.
[26] M. Betke and N. C. Makris, "Recognition, resolution and complexity of objects subject to affine transformation," *Int. J. Comput. Vis.*, vol. 44, no. 1, pp. 5–40, Aug. 2001.
[27] J. Gips and J. Gips, "A computer program based on Rick Hoyt's spelling method for people with profound special needs," in *Proc. Int. Conf. Computers Helping People With Special Needs (ICCHP)*, Karlsruhe, Germany, July 2000, pp. 245–250.
[28] D. G. Altman, *Practical Statistics for Medical Research*. Boca Raton, FL: Chapman and Hall/CRC, 1991.
[29] R. J. K. Jacob, "What you look at is what you get," *Computer*, vol. 26, no. 7, pp. 65–66, July 1993.

**Margrit Betke** (S'93–M'95) received the S.M. and Ph.D. degrees in computer science and electrical engineering from the Massachusetts Institute of Technology (MIT), Cambridge, in 1992 and 1995, respectively.

Her research area is computer vision. She is particularly interested in real-time applications to rehabilitation engineering, human–computer interaction, and biomedical imaging. After completing her Ph.D. dissertation on "Learning and Vision Algorithms for Robot Navigation" at MIT, she spent two years as a Research Associate with the Institute for Advanced Computer Studies, University of Maryland (UMD). She was also affiliated with the Computer Vision Laboratory of the Center for Automation Research at UMD. She was an Assistant Professor at Boston College from 1997 to 2000, and has been a Research Scientist at the Massachusetts General Hospital and Harvard Medical School since 1999. In July 2000, she joined Boston University as an Assistant Professor in computer science.

**James Gips** (M'97) received the S.B. degree from the Massachusetts Institute of Technology, Cambridge, in 1967 and the M.S. and Ph.D. degrees in computer science from Stanford University, Stanford, CA, in 1968 and 1974, respectively.

His research interests generally are in human–computer interaction, especially in developing technology to allow people with profound disabilites to interact with the computer. He is currently a Professor in the Computer Science Department, Boston College, Boston, MA.

Prof. Gips was a Finalist, Discover Magazine Technological Innovation Awards, for the EagleEyes Project in the area of Computer Hardware and Electronics in 1994.

**Peter Fleming** received the B.A. degree in computer science from Boston College, Boston, MA, in 2000.

He currently is with Accenture, Wellesley, MA.