EYE-INTERACT: A LOW-COST EYE MOVEMENT CONTROLLED COMMUNICATION SYSTEM

Amir Sadeghian, Chol-Ho Yim, Adrian D. C. Chan, James R. Green
Department of Systems and Computer Engineering, Carleton University, Ottawa, ON

Abstract-In this paper we introduce Eye-Interact, a low-cost, eye movement controlled communication system for persons with high-level physical disabilities. This system consists of a computer equipped with a webcam. The Hough transform and a K-Nearest Neighbor (KNN) classifier are used to determine the eye position and gaze direction, respectively. The Eye-Interact user is presented with a dynamic user interface, consisting of a computer screen segmented into a 3×3 grid, with each of the 9 grid areas displaying different user inputs. The user can select an input by gazing at the desired grid area. Activation of the input can be determined by the gaze sequence, duration, or blinking.

INTRODUCTION

Eye-Interact is a low-cost, non-contact eye movement controlled communication system, which is mainly targeted to persons with high level physical disabilities. Indeed, eye movements are often the only communication alternative for persons who have suffered an acute neurological trauma or those who have been hospitalized with certain progressive and chronic diseases. Figure 1**Error! Reference source not found.** depicts a usage scenario of the Eye-Interact system. The Eye-Interact interface presents the user with a 3×3 grid, with each of the 9 grid areas displaying different user inputs (e.g. letters, words, or symbols; Figure 2**Error! Reference source not found.**). The user selects an input by gazing at the desired grid area. The user's gaze is determined from eye images capture using a webcam. Activation of the input can be determined by the gaze sequence, duration, or blinking. The user interface is dynamic as the input options will change depending on the user's sequence of input activations.

There exist a number of commercially available eye movement controlled communication devices. Eye gaze in these systems is typically determined by using corneal and pupil reflections from an infrared light source. The H.K. EyeCan VisionKey system [1] is one such commercial system, which consists of a user interface that is worn by the user, similar to a pair of eyeglasses. The up-close measurement of the eye helps to achieve reliable gaze detection and eliminates the need to compensate for head movements. The system has the negative effect of occluding the vision of one eye.

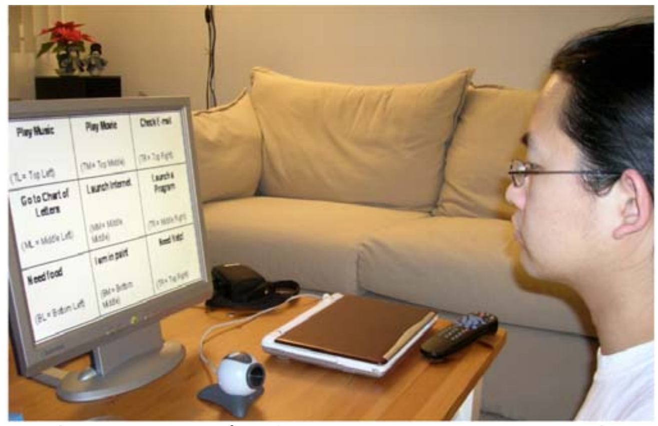

Figure 1 Example of Eye-Interact usage scenario

The Tobii MyTobii [2] is a system similar to the proposed Eye-Interact system. It includes a non-contact eye tracking system that can provide estimates of gaze direction accuracy up to 0.5°. This system is costly, however, and is dependent on its own custom hardware, which consists of a monitor display system, integrated with an eye imaging camera.

The Eye-Interact system is implemented on a PC equipped with a webcam; such systems today are often readily available making the system cost low. In addition, the system can be made portable by using a laptop or a tablet PC.

Play Music (TL = Top Left)	**Play Movie** (TC = Top Center)	**Check E-mail** (TR = Top Right)
Go to Speller (ML = Middle Left)	**Launch Internet** (MC = Middle Center)	**Launch a Program** (MR = Middle Right)
Need food (BL = Bottom Left)	**I am in pain!** (BC = Bottom Center)	**Need Help!** (BR = Bottom Right)

Figure 2 Example of Eye-Interact user interface

METHODOLOGY

The implementation of the Eye-Interact system can be divided into three main stages: Eye Detection, Gaze Detection, and Menu Activation. Figure 3 provides a system flow diagram.

An image of the user's eye is captured from a video stream using the webcam. The Eye Detection stage locates the position of the eye within the image and a 80×60 pixel subimage of the eye is extracted. The Gaze Detection stage uses this eye subimage to discern the user's gaze, which is used for input selection. Menu Activation determines when an input is activated.

The system is divided into two modes of operation: Calibration Mode and Operation Mode. In Calibration Mode, training images of each gaze direction are obtained. This is done once during initial setup of the system. In Operation Mode, the training images are used in a pattern recognition system to discern the gaze direction while the system is in use. These two operation modes are indicated in the system flow diagram (Figure 3).

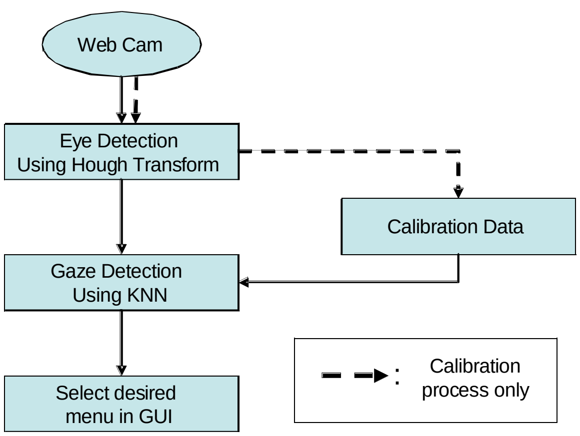

Figure 3 Eye-Interact system flow diagram

Eye Detection

The Eye-Interact system determines the eye position within a search area defined by the previous estimate of the eye position. It is assumed that any movements of the head with respect to the webcam will be small and slow, which is a reasonable assumption given that this system is targeted to persons with high-level physical disabilities. A rough initial eye position is provided during the calibration process by having the user click on the center of their pupil in a captured frame. While this solution is suitable for the prototype system, ultimately, the calibration stage will not require manual intervention.

The Eye-Interact system is assumed to be operated under ambient lighting conditions, with levels similar to office illumination. With overhead lighting, the eye will often be shadowed by the forehead. Therefore, the grey scale image pixel intensities are adjusted to improve the contrast of the iris (Figure 4b). This compensation is accomplished by increasing the dynamic range of the image values using a constant scaling factor. A Sobel filter is then applied to perform edge detection (Figure 4c).

The Hough transform is applied to the edge-detected image in order to determine the position of the center of the iris. Each pixel in the eye subimage is tested as a potential circle center and an accumulation score is generated by summing the number of pixels which fall on the hypothetical circumference. The accumulation score includes a range of radii, ranging from 10-14 pixels, which was empirically determined to be an appropriate operating range. The position with the highest accumulation score should correspond to the center of the iris (Figure 4d). The 80×60 pixel eye subimage is re-centered at this location prior to the Gaze Detection stage.

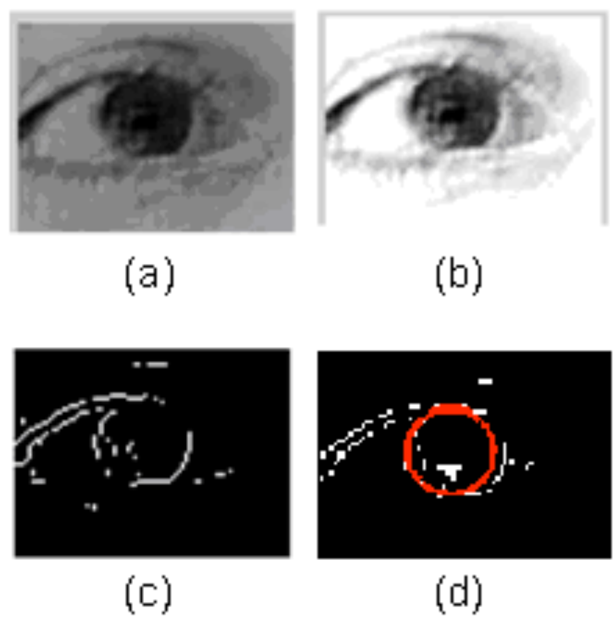

Figure 4 Eye image: (a) original (b) illumination compensation (c) edge detection (d) iris detection

Gaze Detection

Gaze detection is determined by applying pattern recognition to the eye subimage. A KNN classifier is used, along with the calibration data, to classify the eye subimage as belonging to one of the 10 eye cases

(9 gaze directions and eyes-closed). Input to the KNN classifier is simply an absolute difference score, obtained by summing the absolute difference between the test image and the calibration image.

Menu Activation

Activation of a selected menu item can be determined by the gaze sequence, duration, or blinking. The H.K. EyeCan VisionKey system [1] uses a gaze sequence for menu selection, which requires a two step process. First, the user must first look at a peripheral item on the display to initiate activation mode. Second, the user moves their gaze to the desired menu item. This allows eye gaze to wander centrally without any erroneous menu selection. A similar method will be used in the Eye-Interact system.

Menu selection could also be performed using gaze duration; however, it would be difficult to determine a practical duration length that would avoid undesired activations, while also ensuring timely responses for desired activations. Using eye blinks for menu selection may also be problematic as eye blink detection is often unreliable, including discerning natural eye blinks from intentional ones.

Data Collection

Central to the operation of the Eye-Interact system, is the ability of the system to accurately and reliably discern gaze direction. To that end, we have focused our efforts on the development of this portion of the system.

To help the initial design and testing of the Eye-Interact system, data were collected using a 19” (48.26 cm) LCD computer monitor that was set at a distance of 70 cm from the eye. A 640x480 pixel webcam was set between the user and the monitor at a height of 9 cm from the desktop and a distance of 40 cm from the user. There are a total of 10 eye cases: the 9 gaze directions and an eyes-closed case (Figure 5). An automated data collection program was created to highlight one random state every 5 seconds, with a rest period of 3 seconds between states. Participants were instructed to focus their gaze on the highlighted state. For each of the 10 eye cases, 40 images were collected (400 images in total); 20 of these images were used as calibration data and 20 used as testing data. Data were collected from six subjects with varying skin tone and eye shape (3 Asian, 2 Persian, and 1 Indian subject). To simulate a user with a high level disorder, subjects used a chin rest during the study.

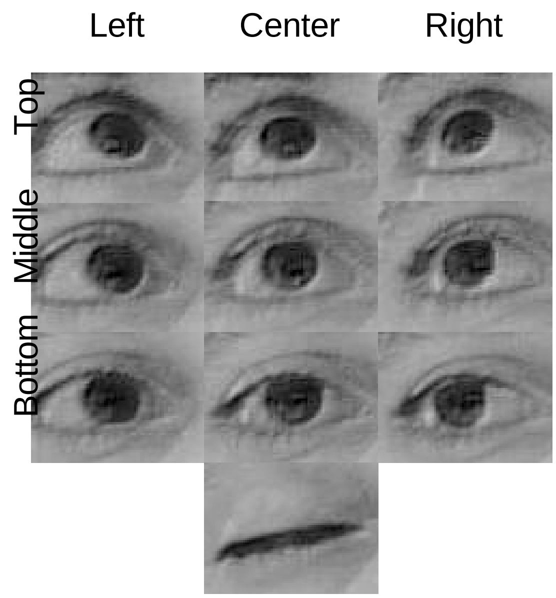

Figure 5 Example images of the 10 eye cases: 9 gaze directions and eyes-closed.

RESULTS AND DISCUSSION

Figure 6 shows the classification accuracy of the KNN classifier as a function of K, averaged over the six subjects. The highest performance was obtained for K = 1, which corresponds to a *best template matching approach*. For K = 1, the classification accuracy ranged from 91.0% to 98.0%.

An examination of the classification process revealed that the majority of gaze classification errors were due to errors in the eye detection stage (i.e. subimages not correctly centered on the true iris location). The two principle causes of this error type were: 1) when multiple candidate circles are detected and the wrong circle is chosen, and 2) when eyelid occlusion leads to a rectangular iris. Problem 1) may be corrected by checking the original grey scale values of all pixels contained within each candidate circle. The darkest circle is assumed to be the iris. A prototype implementation of this correction step looks very promising, correcting all iris location errors when tested on one of the six subjects. Problem 2) may be addressed by modifying the Hough transform to search for different iris shapes depending on the ethnicity of the user. In general, the use of a two-dimensional cross-correlation between the test image and each calibration image to find the best alignment of the images (a form of image registration) before computing the difference may mitigate errors introduced at the Eye Detection stage.

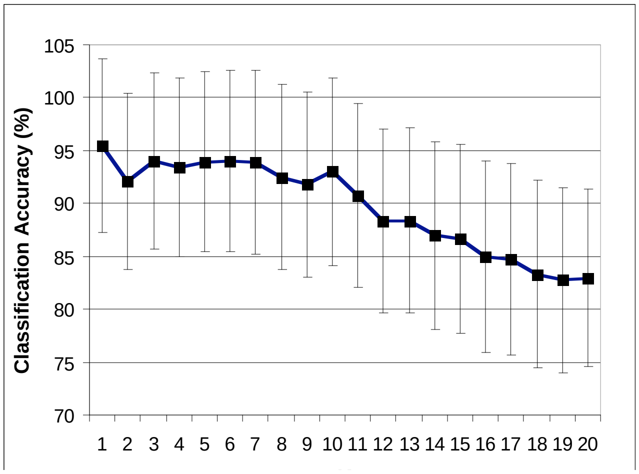

Figure 6 Average classification accuracy (and standard deviation error bars) of the KNN classifier as a function of K.

CONCLUSIONS

The Eye-Interact is a low-cost communication device for persons with high level physical disabilities, enabling a human computer interface via eye movements alone. The system consists of a webcam and a computer running software that will determine eye gaze from the incoming webcam images, which will drive a dynamic menu system. The preliminary results presented in this paper have demonstrated the ability to derive eye gaze from webcam images, which is the key to the success of the system. An average accuracy of 95.5% was achieved across six subjects, which is much higher than the *a priori* accuracy (i.e. 10% from randomly guessing); however, this accuracy is still well below the desired accuracy for a communication device.

Since we have focused on creating a low-cost solution, the Eye-Interact gaze detection system will likely have limitations relative to commercial systems. The system may ultimately only work in a constrained environment, including ambient lighting conditions. Certainly, automatic contrast and brightness adjustments can help mitigate these effects. Currently, the system also limits the movements of the user, which may be resolved using face tracking algorithms. Indeed, effective face tracking has already been demonstrated with current low-cost webcam devices [4]. Such systems can also automatically compensate for varying distances between the webcam and eye through zooming functionality, enabling increased freedom of the user and easier system setup.

As discussed above, the accuracy of the gaze detection system can be improved with a number of different methods. One simple enhancement is the simultaneous analysis of the two eyes; whereas, the current system is only operating on the right eye. In addition, the current implementation of the Eye-Interact system is using a 3×3 menu grid. Examining Figure 5 one can see that it the task of discerning the horizontal gazes (i.e. left, center, and right) is easier than discerning the vertical gazes (i.e. top, middle, and bottom). An alternatively grid, such as a 2×2 grid or 3×2 grid, would improve the accuracy of the system with the tradeoff of a decreased number of menu inputs.

While the Eye-Interact gaze detection may not be able to outperform commercial gaze detection systems, the system is expected to achieve a performance that will be acceptable to the majority of users. Under a reasonably constrained environment, the performance difference should be insignificant. The Eye-Interact system has the benefit of significantly lower cost and hence increased accessibility.

Currently, the Eye-Interact system is employing a template matching pattern recognition scheme to perform gaze detection. Future work will seek to replace this with a heuristic approach based on the relative position of the iris with respect to the rest of the eye to determine the eye gaze. The advantage of this approach is that it will eliminate the need for calibration data thereby simplifying system setup.

REFERENCES

[1] D. A. Kahn, J. Heynen, G. L. Snuggs, “Eye-controlled computing: The VisionKey Experience”, Proceedings of the 14th International Conference on Technology and Persons with Disabilities (CSUN’99), Los Angeles, CA, 1999

[2] Tobii Ltd., “Eye tracking products from Tobii technology”, http://www.tobii.com/, 2006.

[3] M. Nixon, A. Aguado, Feature Extraction & Image Processing, Newnes, 2002.

[4] Logitech, “Video Technologies” http://www.logitech.com/ index.cfm/products/videotechnologies/welcome/US/EN, 2006