HCI Based Input Device for Differently Abled

Brammi J [a,1], Dhivya R [b], Padmavathy T [b], Anitha R [b]

[a,b] *Dept of CSE, Sri Venkateswara College of Engineering, Chennai*

Abstract. Tremendous advancements in the technology has lead to the monumental growth of various branches of computing such as Computer Vision and Human-Computer Interaction (HCI). Computer's input has data about different properties of users, objects or places. For instance, mouse and keyboard works by the movement performed by the end user's hands. These approaches are not appropriate for differently abled people. Away to create an application which replaces the input devices such as mouse and keyboard by using face of the user is proposed. This paper introduces how head motion of the user can be used to control the mouse cursor and how gaze tracking can be used to control the keyboard. A face detecting system precisely records the motion parameters from video at real-time using a typical webcam. While the pace reduces while using the virtual mouse and keyboard, the performance of the system is intact for differently abled people whose only means of communication are the head movements and gaze.

Keywords. HCI, Computer Vision, Gaze Tracking, Eye-Aspect-Ratio, Dwell time, Mouth-Aspect-Ratio

1. Introduction

Computer vision focuses on assisting the computers to identify objects by seeing. Computer vision's goal is to make use of observed image data to infer something useful and meaningful. It is a discipline which can be said as the sub-field of machine learning and AI that makes use of general learning algorithms and exclusive techniques. Computer Vision tries to make sense of digital images. This includes developing techniques to replicate the potential of vision of humans. To understand digital images, we must be able to obtain information from an image, that can be an entity, a 3D model, a textual explanation etc. This is a very demanding work. We are very farther away from creating a versatile "seeing machine". In this paper, Human-Computer Interaction (HCI) in Computer Vision is discussed in detail. Human Computer Interaction (HCI) studies how technology can be used and designed to act as interface between humans and computers. Researchers contemplate the methods in which computers and humans interact with each other and develop methods that allows people to communicate with machines in unconventional methods.

---

[1]Brammi J , Dept of CSE, Sri Venkateswara College of Engineering,Chennai;
  Email id: brammi01@gmail.com

When media studies, design, and various other study disciplines. Eye-based Human–computer interaction (HCI) can be taken advantage by physically disabled people. It enables effective communication for disabled people with computers and also helps to avoid eye damage caused by external hardware wearable devices. considered as a research discipline, HCI can be placed in behavioral sciences,computer science,

The applications of object tracking of both stationary and movable objects like engineering and health care makes use of a conventional technology that is used to develop systems that tries to disentangle human life. Concepts like virtual/augmented reality, image processing, human-computer interaction etc. Are used by these systems to create devices using which humans can effectively communicate with the computers. Haar Cascade method can be applied to detect certain features of the face and Integral Projection can be used to obtain the positions of the movement of the eyes. The scanning keyboard is expandable, as more buttons can be included. Menu selection with numbers and shortcuts to use at times of emergency increases the system's possibilities. Gaze isn't as precise as mouse. Inaccuracies originate due to technology and also due to the features of the human eye. Fovea's dimensions and also the inability of the camera to resolve the fovea position limits the preciseness of the measured gaze point to 0.5 degrees. False detection can be minimized using high resolution images, allowing more head orientations and colored backgrounds. Even though a device which tracks the eye is precise in the beginning, with prolonged exploitation, the actual point of gaze and the measured point of gaze do not match. To type the text in the virtual keyboard, dwell time method can be implemented. The subject has to gaze at any of the buttons in the virtual keyboard. The corresponding button is highlighted and gets sent automatically to the text box or message box. If dwell time exceeds threshold value, then corresponding command is enabled. Measure of intention is one of the major problems in HCI that deals with eye tracking. It is difficult to calculate because of the various involuntary eye movements that causes involuntary selection of items. One way to overcome this is by taking into account explicit actions like blinking of eye as an indicator of user's intention to run a command. Gaze is not used for control but only for selection. The virtual keyboard works on a tree selection method. The tree can have two levels. In the first level, the sections are dedicated to the selection of digits, punctuation marks and letters. Selection of one of these commands opens the tree's second level in which the commands are the tree's underlying eight characters. The traditional keyboard layout can be inefficient for its use with eye tracking devices because closeness of buttons increases the confusion of the procedure of gaze detection. The virtual keyboards are hard to assess as the performance is dependent on the amount of texts that is spelled out during the experiments, the subject using the system and the user's experience and motive.Hand's free control for Human-Computer Interaction (HCI) is becoming indispensable nowadays. For a solution, utilizing the body movements of humans in real time as video input and making use of tracking is an important solution.

2. Related Work

The object tracking (stationary and moving) is an older discipline and its applications are varied occurring in different fields like health care and engineering. Concepts like virtual/augmented reality, image processing, human-computer interaction etc. are utilized by these systems, to create devices using which humans can effectively communicate with the computers. HaarCascade method can be applied to detect certain features of the face and Integral Projection can be used to obtain the positions of the movements of the eye stated by authors [2]. The scanning keyboard is expandable, as additional buttons can be included as suggested by enat Al-Kassim [1] et al. They further claim that false detection can be minimized by using high resolution images, allowing colored backgrounds as well as allowing further head orientations. Menu selection with numbers and shortcuts increases the system's possibilities [3].Gaze isn't as precise as the mouse. Inaccuracies originate due to technology and also due to the features of the human eyes.

Fovea's dimensions and the inability of the camera limits the preciseness of the point of gaze measured to 0.5 as suggested by Sara Bilal [4] et al. Even though a device which tracks the eye is precise at the beginning, with prolonged exploitation the actual point of gaze and measured point of gaze do not match. To type the text in the virtual keyword, the dwell-time method can be implemented. The user has to gaze at any key in the eye tracking keyboard. Then the key is highlighted and sends the text in the key to the message box or display box. If the dwell time goes past the specified threshold level, the corresponding action is enabled.

Measure of intention is a crucial problem in HCI that deals with eye tracking. It's difficult to calculate because of the various involuntary eye movements that causes involuntary selection of items. One way to overcome this is by taking into account explicit actions like blinking of eye as a sign of user's willingness to execute an action.Gaze is not used for control and it is used only for selection. The virtual keyboard works on a tree method for selection. The tree can have two levels. In the first level, the sections are dedicated to the selection of digits, punctuation marks and letters. Selecting one of these commands open the tree's second level in which the commands are the tree's underlying eight characters [5]. The traditional keyboard layout can be inefficient for its use with eye tracking devices because closeness of buttons increases the confusion of the procedure of gaze detection. The virtual keyboards are hard to assess as its performance is dependent on the amount of text that is spelled out during experiments, the subject using the system, and the user's experience and motive.

Since the invention of computers, there is a long history to use mouse as human-computer interface. Several researchers have attempted to develop techniques to help the differently abled people to communicate with computers making use of signals such as facial muscles signals (EMG), electroencephalography (EEG) from the brain and electro-oculogram (EOG). Other state-of-the-art techniques to track eye movements to control computers are costly and not cost-effective for people in need.

Farida Mohamed [8] et al., used the laptop or a PC along with webcam and did not make use of additional hardware like infrared, electrodes, or any other source of light for tracking the eyes. By taking consecutive photos of the user from the webcam, the program processes these frames at individually at extremely high speeds and compares the iris shift in each frame with respect to the initial frame. Before detecting and tracking the eyes, the frames undergo several levels of processing. On obtaining the processed image,the iris shift is calculated and the program prompts the cursor to move to the respective location on the screen. In order to do this, initially the face detects the landmarks of the facial features like mouth and eyes are mapped on the face using the dlib library. dlib is mainly used for face detection. Histogram of Oriented Gradient (HOG), Image descriptor together with Linear Support Vector Machine (SVM) is used to set up highly accurate image classifiers. Initially, HOG descriptor is used to extract feature matrix, then the SVM classifier is trained using these features. HOG uses the benefits of both bi-class and multi-class HOG based detectors to develop three stage algorithms with low computational cost. In the first stage, the multi-class classifier with coarse features is used to estimate the orientation of a potential target object in the image; in the second stage, a bi-class detector filters out the false positives respect to the detected orientation with intermediate level features; and in the final stage, a bi-class detector corresponding to the detected orientation using fine features is employed to attain accurate detection with low rate of false positives. In this manner, features are drawn out from a detected face. After extracting the features, it is fed to linear support vector machine for classification.

A Haar Cascade is basically a classifier that detects the object for which it's been trained for, from the source. As stated by Ashlesha Singh in [9], haar cascade structure primarily slides the image over one by one, the pixel values that are masked in the black portions are summed up, finally the added values are composed and accordingly the threshold values are set. Once HaarFaceDetector finds a face within the captured image, the eye-models isolate eye-regions. It is possible to track user's gaze information for many decades. But, systems that tracks the eye gaze are still expensive.

Manu Kumar [6] proposed many face and eye detecting strategies to reduce the cost of eye gaze tracking systems. The Gaze Tracker software make use of the OpenCV library. Also, the software uses the HaarFaceDetector to detect faces in the captured image. To detect eye-regions, classifier set is trained by using 3000 sample images.

Ashlesha Singh [9] et al., made it possible to find the location of the pupil center (P) and the corneal reflection (CR) which can be used to estimate the point of regard. In most of the desktop, the head motion occurs in horizontal plan. Chairat Kraichan [7] et al., proposed a mechanism to extract the eye region (i.e.) by dividing the image of the face into three equal horizontal areas. The eyes are located in the upper area. The robustness of the system at different roll and pitch angles of head pose for detecting face and eye center is examined. Finally, eye center co-ordinates are mapped in X-axis and Y-axis according to the camera co-ordinates to computer screen and creates mouse control based on face and eye tracking. The image of one eye is extracted then it is normalized in order to remove the background noises and then it is converted into binary image to enhance contrast. Activation and deactivation of the computer mouse is controlled according to the Mouth-Aspect-Ratio.

Wang Jun [10] et al. provided a model that selects the center point of mouth as a reference point to estimate mouth rotation angle with respect to X-axis and Y-axis. When face mouth is rotated from one position to another (Example: From frontal face to left 45 degree in x direction) corresponding to the captured image, the mouth position is also moved from horizontal center to a corresponding position linearly. We can utilize the relationship between mouth position and face boundary to calculate the angle of the mouth for measuring the Mouth-aspect-ratio. After the face, eye and mouth detection, the threshold value is compared with eye- aspect-ratio and mouth-aspect-ratio. Since both left and right eye blinks synchronously the aspect ratio of both left and right eye is averaged. Mouth-Aspect-Ratio is tweaked from the Eye-Aspect-Ratio.

3. Proposed Work

The proposed system involves capturing of the head movements and tracking of the eyes. This is achieved with the capturing of the mouse cursor movements along with the character selections of keyboard. The system includes mechanisms to detect face, track the features of the face, extract eye blinking and mouth movements and to capture the keystrokes.
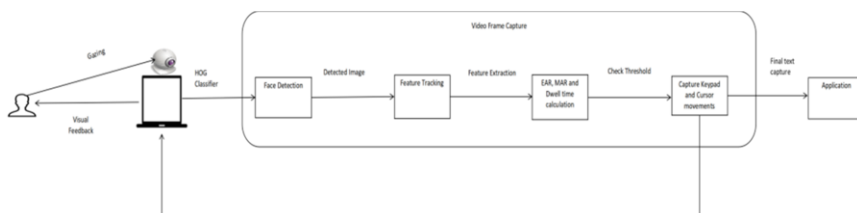

Figure 1. Architecture of HCI Input Device

Initially, a typical web camera is set to monitor the recorded image containing the object which is a face. For accurate and rapid image processing, OpenCV library is used. All the pixels of the image recognized as a face is categorized into one of the types, namely skin and non skin pixels. To examine if the detected area is face or not face, connectivity analysis is used to analyse the established skin area. If a face is identified, dlib is used to map the facial features' landmarks. Facial landmark is an inbuilt HOG SVM classifier which is used to determine the position of 68(x,y) co-ordinates which maps the facial structure. The indices of the 68 co-ordinates are marked. Once the landmarks of eye and mouth are determined, EAR and MAR are computed. The mouse events are performed if in case EAR and MAR values reaches the specified threshold value. Opening of the mouth is used to control the activation and deactivation. Cursor movements are controlled by moving the head in the respective directions such as up, down, right and left. Wink actions are used to control the clicks of the mouse. Right wink is used to perform right click and left wink is used to perform left click. In addition, the scroll movements of the mouse is introduced in

this model. Squinting of the eyes will activate and deactivate the scroll movements. A visual feedback is continuously provided regarding the position of the mouse.

Table I. Mouse Movements

Action	*Functions*
	Activate/Deactivate Mouth Control
	Right Click
	Left Click
Head Movements (Pitch and Yaw)	Cursor Movement

In virtual keyboard, the image processing is executed in the same way as the computer mouse cursor controller. Basically the GUI of the virtual keyboard is made up of two components - First, the center of the screen where the selected text is displayed and second is the edges of the screen where the character sets are displayed. Since the traditional keyboard layout is inefficient, a tree selection method is used with 6 sections. There are 3 levels. In the first level, the 6 sections are scanned by using dwell-time method. In the second level, one of the 6 section is elaborated for selection of alphabets "ABCDE, FGHIJ, KLMNO, PQRST, UVWXYZ".

For example, if a person selects the box "ABCDE" the partitions s1, s2, s3, s4 and s5 become "A," "B," "C," "D," "E". The third level is dedicated to numbers and commonly used predefined sentences. The predefined sentences helps the users to communicate with ease at times of emergency. A typed character can be deleted by closing the eyes for a short period of time. Furthermore, to display the key chosen by the system, a feedback is provided for the chosen key to the user in a visual form. In the beginning, the color of the button chosen is light blue. When the user gaze a particular section, color varies with respect to dwell time t. the button becomes dark blue.
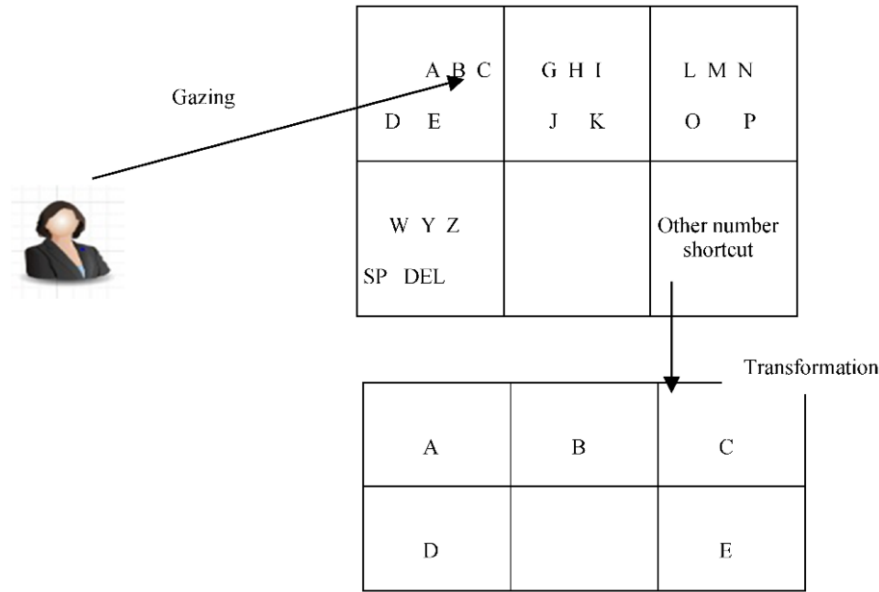
Figure 2. Virtual Keyboard Layout

Both mouse cursor controller and the eye tracking keyboard work well in sub optimal lighting conditions. People wearing spectacles could also use the system with ease. Since an in-built webcam is used rather than an external webcam, the proposed system is cost effective.

4. Experimental Results

The mouse cursor is activated/deactivated by opening the mouth. The mouse cursor movements are performed by moving the head up, down, left and right, right click is done by performing right wink and left click is done by performing left wink. The scroll mode is activated/deactivated by squinting the eyes. All the mouse cursor movements are performed without any error. The system scans the six sections of the keypad effectively and captures the keys accurately. The system was tested under various lighting conditions. It works efficiently even in sub optimal lighting. It is also found to be taking accurate inputs when the subjects are wearing spectacles. Reducing the number of sections in the virtual keyboard to six helps us to reduce the false positives.
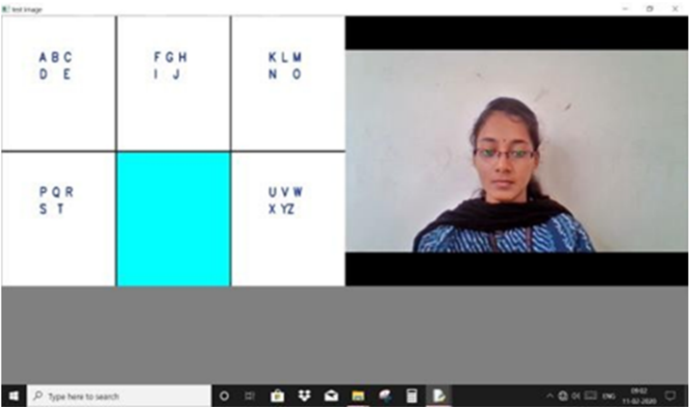
Figure 3. Demonstrates Virtual Keyboard Layout

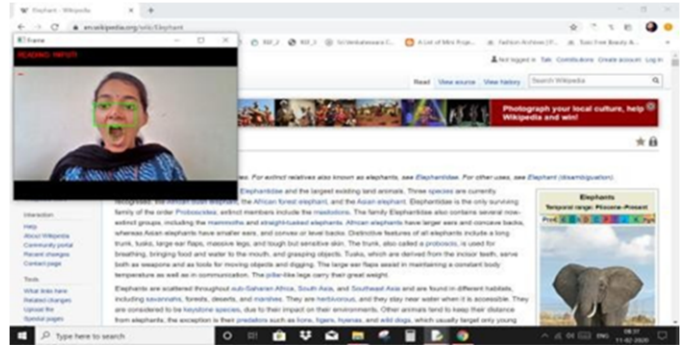**Figure 4. Demonstrates Key Selection based on Tree Selection Method**

Figure 5. Demonstrates Mouse Cursor Activation/Deactivation

5. Conclusion

An eye tracking keyboard that makes use of tree method and mouse cursor controller using the head movements of the user is proposed in this paper. The suggested model provides an alternative solution to external input devices. This model uses face/head motion to control the movement of mouse cursor such as right eye wink for right click, left eye wink for left click, head rotation for mouse movement and gaze tracking for using a virtual keyboard. The system provides an alternative solution for future computing devices, such as PDA, Tablet, Laptops, and GPS. It inspires various PC based application like gaming, entertainment, robotics, etc. The suggested system improves the living conditions of disabled people since the system works efficiently even under sub-optimal lighting and with the use of spectacles.

References

[1] Al-Kassim Z, Memon QA. Designing a low-cost eyeball tracking keyboard for paralyzed people. Computers & Electrical Engineering. 2017 Feb 1, pp.20-29.

[2] V. I. Saraswati, R. Sigit and T. Harsono .Eye gaze system to operate virtual keyboard .2016 International Electronics Symposium (IES), Denpasar, 2016, pp. 175-179.

[3] Cecotti H .A multimodal gaze-controlled virtual keyboard .IEEE Transactions on Human-Machine Systems. 2016 May 3, pp.601-606.

[4] Sara Bilal, Mohamad Hassrol Bin Mat, and Rasheed Nassr .Design a Real-Time Eye Tracker .Proceedings of the 2018 the 2nd International Conference on Video and Image Processing, 2018, pp. 187–191

[5] Cecotti H, Meena YK, Bhushan B, Dutta A, Prasad G. A multiscript gaze-based assistive virtual keyboard. In2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) 2019 Jul 23, pp.1306-1309, IEEE.

[6] Kumar M .Reducing the cost of eye tracking systems .CSTR, Stanford Tech Report, 2006 Apr 8.

[7] Kraichan C, Pumrin S, Face and eye tracking for controlling computer functions. In 2014 11th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 2014 May 14 pp.1-6, IEEE.

[8] Nasor M, Rahman KM, Zubair MM, Ansari H, Mohamed F .Eye- controlled mouse cursor for physically disabled individual . In 2018 Advances in Science and Engineering Technology International Conferences (ASET), 2018, pp. 1-4. IEEE.

[9] Khushaba RN, Kodagoda S, Lal S, Dissanayake G .Driver drowsiness classification using fuzzy wavelet-packet-based feature-extraction algorithm. IEEE transactions on biomedical engineering, 2010 Sep 20, pp.121-131.

[10] Naizhong Z, Jing W, Jun W .Hand-free head mouse control based on mouth tracking . In 2015 10th International Conference on Computer Science & Education (ICCSE) 2015 Jul 22, pp.707-713, IEEE.