International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2020): 7.803

# Human Computer Interface for Physically Impaired

**Aayush Bhansali[1], Shivani Vhatkar[2], Shamoil Arsiwala[3], Aditi Srivastava[4], M. K. Nivangune[5]**

[1, 2, 3, 4] Student, Computer Engineering, Sinhgad Academy of Engineering, SPPU, Pune, Maharashtra, India

[5]Assistant Professor, Computer Engineering, Sinhgad Academy of Engineering, SPPU, Pune, Maharashtra, India

**Abstract:** ***Human Computer Interface (HCI) is a trend-in technology. HCI refers to means of communication between a user and a computer via various input/output devices. Working in this field, we've developed a system to produce an answer to the limb disabled individuals to navigate the computers. Different algorithms are enforced in this paper, to estimate the gaze of the user to acknowledge the reference key. Due to the recent advancements within the field of Computer Vision and Object detection and tracking, computers are currently ready to understand the world with a perspective corresponding to humans. The head involuntarily follows the field of vision, i.e., you progress your head wherever your eyes are centred. Exploiting this biological feature of the autonomous and involuntary movement of the head, any of the facial expressions (eyes, nose, jaw, etc.) is sufficient to confirm what the user desires to concentrate on, and also the mouse indicator would be stirred consequently. The system, once enforced and fine-tuned consistent with the user, would alter a swish navigation of the mouse indicator across the viewport of the screen, altogether eliminating the requirement of the mouse.***

**Keywords:** HCI, Computer Vision, Facial Landmark Detection, SSD framework.

## 1. Introduction

The field of computing has achieved wonders in the past two decades. From the explosion of the internet to handheld smartphone devices, computers have revolutionized our day-to-day life. Today with the help of computers we are able to achieve things that were once imagined to be next to impossible. As transistors are becoming smaller, allowing for a high transistor packing density, computing speeds are growing steadily. With such massive computing power, computer scientists are now able to implement and improve various algorithms which were once thought impractical due to technological limitations. Thanks to the advancements in Artificial Intelligence, computers can now mimic some limited aspects of humans such as object detection, pose estimation, finding patterns in data, playing games such as chess. One of the branches of AI is the field of computer vision, which enables computers to analyze static/real-time images and detect objects and draw inferences.

The system we have developed uses computer vision to determine the gaze of the user. This information can then be used to navigate the cursor of the computer. The system would enable users, who are misfortunate to have limb impairments, to use the computers with ease as the existing interfaces provide little to no help for them. Using the system, such individuals can use a camera device in order to navigate through the screen just by head movements, thereby providing them with a faster means of communication with the computer. The performance parameter of such a system is also taken

into consideration as a real-time system would easily hog up the limited computationalresources. It would be impractical for the system to require high computational resources in order to function.

## 2. Motivation

Traditional human computer interfaces do provide access to impaired individuals to a certain extent. Speech recognition allows users to interact with the computers using just their voice, which the computer then interprets and performs the required actions. However, such systems are still primitive and their accuracy isn't at the peak. These systems require internet connectivity as the samples are processed online for a better accuracy. The system we have developed aims to overcome various disadvantages of current interfacing devices. The proposed system uses a webcam to obtain the images of the user from the video stream and process them in real-time to translate it into cursor movements. The system is capable of working offline and can theoretically eliminate the need of a mouse pointing device entirely. With the help of this system, we aim to overcome the drawbacks of the existing accessibility features and allow the disabled audience to interact with the computers with ease.

## 3. Literature Survey

Before we start developing, we need to study the previous papers of our domain which we are working on and on the basis of study we can predict or analyse the shortcomings and start working with the reference of previous papers.

In paper [3], the system implemented required a web camera and microphone. Image processing algorithms were applied to the video stream to detect the user's face and tracked points to determine their head movements. The audio stream analysed by the speech recognition engine determined relevant voice commands. This interpreted logic in turn received relevant parameters from the signal-processing units and translated these into on-screen actions by the mouse pointer. The positive side of the system was the fact that users could control the mouse pointer using head

movements, while their hands remained free to perform other tasks.

The drawback includes that people who are unable to speak and hear can't use such a system for interaction with the computer.

In paper [1], the system implemented has a 4-key keypad created on the computer system. Whenever the user looks at the key to be pressed, the position of the pupil centre varies with the reference. At that time, the eye images are captured by video camera and transmitted to the Computer through USB cable. The pupil corneal reflection method is used to recognize the position of the eye and to determine the gaze of the user.The image processing software developed in computers compares the position of the user eye with reference keys present in the database to recognize the appropriate key. For the typing process, the blink detection method is used.Thepositive note of the system is that it is nearly accurate and cost effective for disabled people but the drawback includes the system having better quality pictures for capturing the user's eye position as well as blink detection method for typing could be an impossible task for typing infinite characters. As well keeping in mind the duration for typing would be very tiring.

In paper [4], the author compares traditional machine learning approaches based on shallow learning to deep learning approaches in the history of face recognition. The paper proposes Convolutional Neural Network based architecture with an additional normalisation of two of the layers in the system. The CNN architecture is employed in the system to extract distinctive facial features to obtain a powerful recognition system with better performances in terms of accuracy and speed.

The prominent features of the paper is that it employs batch normalisation for the outputs of the first and the final convolutional layers in the training stage itself which makes the network reach higher accuracy rates. In the end the system uses SoftMax Classifier to classify the faces in the fully connected layer of CNN.

In paper [5], the author describes how the traditional CNN-Architectures perform well on a large sized dataset and also emphasizes how difficult it is to obtain the large dataset specific to our desired goals. Making our own dataset is also a tedious task as labelling of such huge datasets is not easy. The paper proposes dataset augmentation to overcome this problem and achieve a comparatively higher accuracy with a limited size dataset. The dataset is augmented in 4 ways - Horizontal flip, scaling, shifting and rotating. Using this we can retain the image labels and augment the dataset up to 1000 times. The paper takes advantage of orientation robustness of the CNN and proposes a methodology to use the detection power of the CNN on limited size datasets, on which the traditional CNNs perform very poorly.

In paper [6], the author investigates both CNN based approaches i.e., regression and heat map, their advantages and disadvantages. Using the approaches, the author develops a variation of the heatmap approach - the pixel-wise classification model (PWC). The system also designs as a hybrid loss function and a discrimination network for strengthening the landmarks' interrelationship implied in the PWC model to improve the detection accuracy without modifying the original model architecture.

## 4. System Architecture

The system captures a stream of images from a camera device or a webcam and passes them on to a ResNet SSD network which has been trained to classify human faces. The ResNet generates a bounding box around the detected face. This bounding box is then passed to another convolutional neural network which has been trained to identify the facial landmarks of the given face. Use the nose point as a reference, the initial position of the user head is set and any further movements would be captured relative to the initial position. This eliminates the need for a mapping function as the cursor can also implement relative motion. Based on the magnitude of displacement of the head from its initial position, the new mouse coordinates are calculated and the cursor is moved accordingly.

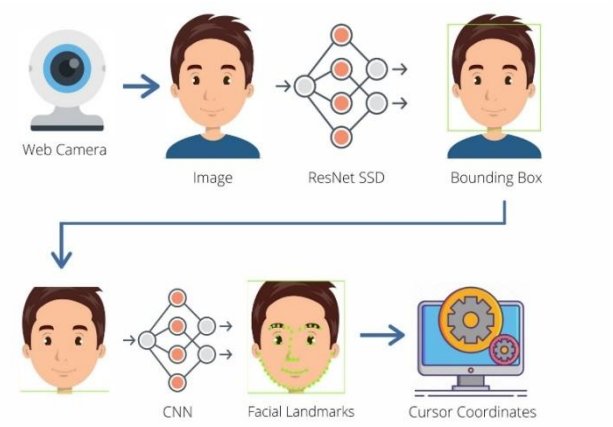


**Figure 1:** System Architecture

## 5. System Implementation

### 5.1 Initial Phase

The user interacts with the system using GUI. Once the user adjusts the initial settings and clicks start, the application tries to access the camera and fetches the image stream from the video camera. The image obtained is of resolution 640 * 480 pixels. The image is then scaled down to 300x300 resolution in order to make it compatible with a pre-trained ResNet SSD model which extracts faceboxes from a given image. The images from the video stream are then passed on to the ResNet classifier. The ResNet classifier accepts an image and outputs confidence values indicating the probability of face in the image as well as it's bounding box. The pre-trained ResNet model is inbuilt in the OpenCV library and has been optimized for intel-based CPUs thereby allowing us to use it on systems that don't have a dedicated GPU and still work at decent FPS.

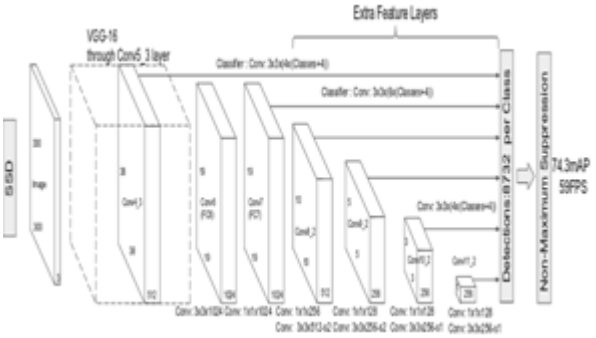


**Figure 2:** ResNet10 300x300 SSD Architecture

### 5.2 Facial Landmark Detection

After the bounding boxes and confidences are obtained from the ResNet architecture. The bounding boxes are then forwarded to another CNN (designed using TensorFlow) which has been trained to detect facial landmarks given the bounding boxes. This CNN outputs 68 facial landmarks which would be used for further manipulation of the cursor. This CNN architecture contains 8 convolutional layers, 4 pooling layers and 2 dense layers. An already trained model by Yinguobing has been used which is well suited for the requirement. The input to this CNN is an image of size 128x128 pixels with three colors channels each and it outputs a set of 68 landmarks in 2D space. This architecture provides decent FPS when working with the above mentioned ResNet architecture. It is important for the system to be able to work on systems with lower specs so that an average user would not have to spend much on a higher end computer to be able to use the functionality, and ensure smooth flow of head movements to cursor movements.

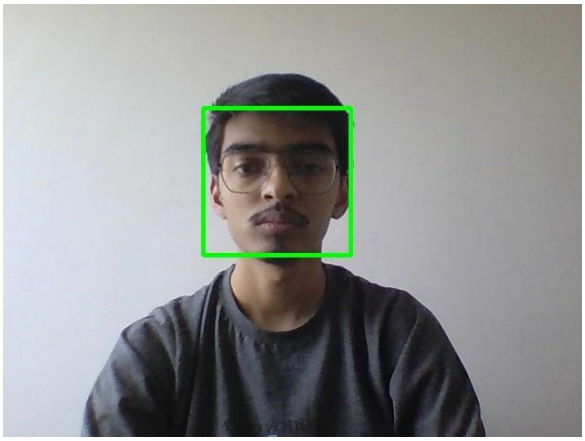

**Figure 3 (a):** Bounding Box

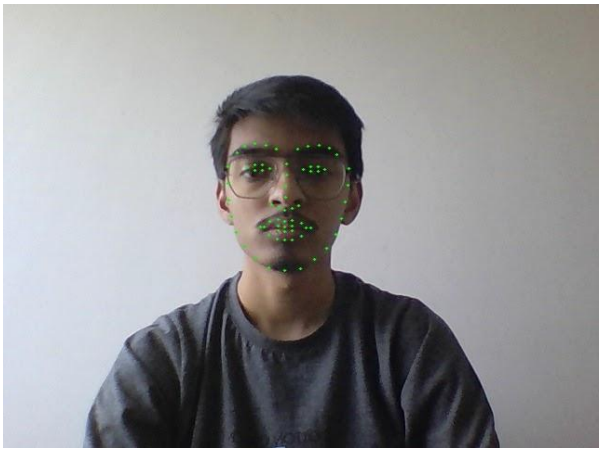

**Figure 3 (b):** Facial Landmarks

### 5.3 Normalizing the reference point

The reference point used for manipulation of the cursor pointer is the nose. This point was selected as it has a comparatively higher displacement compared to other points thereby less prone to noise. The point obtained from the above reference model is set as an initial reference point. This reference point is then subtracted from the further obtained coordinates, so that the new coordinates obtained would be with respect to the reference point instead of the image obtained from the video frame. To simplify, we would get relative displacement of the nose from the starting position instead of an absolute position with respect to the video. The human body cannot stay still, and this may cause a jitter in the cursor even when the user does not intend to do so. In order to overcome this problem, a threshold variable is used, and only the displacement above that threshold value would be used for further calculations thereby providing stability to the cursor. This would also be helpful to bring the cursor to a complete halt easily.

### 5.4 Cursor Manipulation

After the initial setup of the reference point, the system captures a continuous stream of images from the video camera and obtains the coordinates of the reference point. These coordinates are then used to move the cursor relative to its original position. The speed of movement of the cursor is directly proportional to the displacement of the nose from the initial reference point. This speeding can be controlled by the user as per his/her convenience. When the cursor is stationary, the user has to nod in order to register a left click, and the reverse of this action, an up nod would register a right click. Sufficient care has been taken to eliminate the false positives that can be obtained due to noisy data. The registering of the nod is passed through a few filters before the cursor actually clicks.

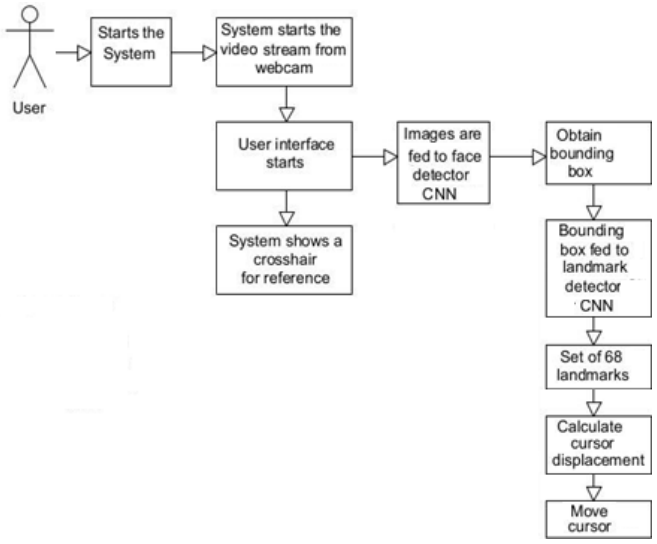


**Figure 4:** System Interaction

**Performance Analysis**

The GUI and the predictor logic have been separated into two independent codebases as a single thread was not able to handle both the graphical processing as well as image processing. Separating them and running them independently allows us to run the GUI and the pre-trained models as separate processes thereby leveraging parallelism for better performance. The two processes communicate using a common json file which uses locking mechanism to prevent data inconsistencies arising from concurrent writes to the same file. This file holds all the key variables which facilitate a smooth flow of data from GUI to the predictor logic and vice versa.

The system was tested on an average end laptop with an i3 processor, integrated graphics - Intel HD Graphics 520, 4GB RAM and 1TB HDD. the system performs at 10 fps and would perform even better if the system has higher specifications. The camera used for capturing the images is a stock camera which most laptops are equipped with before shipping. However, the current system on which it is tested does ensure a smooth flow of cursor movements and clicks. Initially the user would face difficulty as he is not accustomed to such an interface. However, with usage, the brain would adapt to this interface, allowing him/her to move the cursor as seamlessly as a physical mouse would allow.

## 6. Conclusion

The system is one of the many existing ones which aims to use computer science and machine learning for the betterment of society. The main goal of any innovation is to help humanity progress further. Using this system, we hope that more people are encouraged to explore Computer Vision and other related fields of Artificial Intelligence in order to provide new innovative ways of interacting with the computer as well as exponentially improve the performance of the existing systems.

## 7. Future Scope

The system is at a primitive stage and a lot of new functionality can be added as well as the existing functionality can be improved. The system currently limits the usage to left and right clicks, and further mouse functionality can be added to it such as scrolling, click and hold, double click. These systems can eliminate the need for a physical mouse altogether. Further, this app can be extended to use speech-recognition and text-to-speech translation in order to eliminate the necessity of a physical keyboard as well.

## References

[1] Human Computer Interaction For Disabled Using Eye Motion Tracking, Uma SambrekarDipaliRamdasi.
[2] New Features for Eye-Tracking Systems: Preliminary Results 1st Audi I. Al-Btoush ,2nd Mohammad A. Abbadi, 3rd Ahmad B. Hassanat , 4th Ahmad S. Tarawneh ,5thAsadHasanat, 6th V. B. Surya Prasath.
[3] Motion-Tracking and Speech Recognition for Hands-Free Mouse-Pointer Manipulation Frank Loewenich and Frederic Maire Queensland University of Technology.
[4] Face Recognition Based on Convolutional Neural Network. Musab Coşkun,AyşegülUçar, ÖzalYıldırım, Yakup Demir
[5] Human face recognition based on convolutional neural network and augmented dataset. Peng Lu ,Baoye Song & Lin Xu.
[6] A Detailed Look At CNN-based Approaches in Facial Landmark Detection. Chih-Fan Hsu, Chia-Ching Lin, TingYangHung , Chin - Laung Lei , and Kuan-Ta Chen
[7] Online webcam-based eye tracking in cognitive science: A first look Kilian Semmelmann1 & Sarah Weigelt
[8] Face Detection and Segmentation Based on Improved Mask R-CNN, Kaihan Lin, HuiminZhao, JujianLv, Canyao Li, Xiaoyong Liu, Rongjun Chen, and Ruoyan Zhao."
[9] Mouse Control Using Image Processing. 1. Shridhar Kolap, 2. Madhu Khaparkhande, 3. Pooja Latane.
[10] Johnston, B., Chazal, P. A review of image-based automatic facial landmark identification techniques. J Image Video Proc. 2018, 86 (2018).
[11] Liu, Wei &Anguelov, Dragomir & Erhan, Dumitru &Szegedy, Christian & Reed, Scott & Fu, Cheng-Yang & Berg, Alexander. (2016). SSD: Single Shot Multi-Box Detector. 9905. 21-37. 10.1007/978-3-319-46448-0_2.