

# iGaze-Eye Gaze Direction Evaluation to Operate a Virtual Keyboard for Paralyzed People

Ms. Saily Bhagat<sup>1</sup>, Ms. Tanvi Patil<sup>2</sup>, Ms. Varsha Singh<sup>3</sup>, Mr. Abhishek Bhatt<sup>4</sup>, Ms. Snehal Mumbaikar<sup>5</sup>

<sup>1,2,3,4,5</sup>Dept. of Computer Engineering Ramrao Adik Institute of Technology Navi Mumbai, India

\*\*\*

**Abstract** — Eye Gaze has been broadly utilized and provides an appealing alternative plan to the cumbersome interface devices for human computer interaction(HCI) as an assistive technology. The model proposes a novel pipeline of utilizing eye gaze data for display control, on the basis of the captured picture. This study involves design of iGaze that will assist the paralyzed people to write on screen by gazing at different directions :up, down, left, right based on the targeted key, which is advantageous in terms of cost as well as robustness. The system employs a robust machine learning algorithm for gaze estimation wherein gaze direction is classified on the basis of datasets. An efficient gaze estimator has been trained, by recording a dataset which consists of eye gaze direction frames.The system introduces an economically feasible method of using the convolution neural networks (CNN) for human-machine interaction. The system can effectually enhance the accuracy by selecting the key using eye gaze direction and typing the same. The proposed pipeline shows that, eye gaze direction can be estimated and display content of the screen can be controlled accordingly.

**Index Terms**—CNN, Human Computer Interaction (HCI), gaze direction, HAAR Cascade, Adaboost.

## 1. INTRODUCTION

Communication is regarded as vital and foremost parameter for interaction with other human beings. The disabled or paralyzed people certainly face immense issues while communicating. Eyes being the most influential sense of information can be used to communicate with others adequately and effectively. Our gaze is the fundamental element for interacting with our environment. Latest advancements in computer and image processing technologies have made our gaze the most influential factor and thus allowed it to emerge as a significant branch of Human-Computer Interaction (HCI). Tracking the eye gaze requires successfully detecting the position on the screen where a user is looking at. Many of the eye gaze tracking systems developed so far use invasive techniques to estimate the gaze of a person. Although, such technologies do exist commercially, but are so expensive that it is almost inaccessible to most individuals who need it. In addition to his, some eye trackers require a lot of time to calibrate before using. At the time of calibration, people are not allowed to move their body, and infinitesimal movements may give erroneous results in final estimation. The increased usage adversely affects the performance of the system as the calibrated system dysfunctions and it is required to calibrate the system again. These inconveniences could become a matter of concern for disabled people striving to use these systems in their daily life.

Taking into account the necessities of disabled people the proposed system attempts to resolve the inconsistencies that arise during eye tracking process .The goal of this paper is to introduce the proposed system which involves design of iGaze that will assist the paralyzed and disabled people to write on screen by simply gazing at the keys of the virtual keyboard

This system involves training of gaze estimator, dataset is recorded which consists of eye gaze direction frame. Learning through the trained data, the system reacts to the current inputs feed to it. Eye gaze direction frames of the user are captured, on the basis of the learning, the system identifies the key that the user wants to convey and prints it on the screen and hence helps disabled person in delivering the intended information using the proposed system.

## 2. RELATED WORK

The field of eye gaze tracking has been an area of research and development for quite sometime now, and is still one of the most prominent subjects of study even today. Gaze estimation can be performed through numerous techniques. Some gaze estimation methods emphasize primarily on examining the physical characteristics of human eye. In these methods using corneal reflection, infrared (IR) lights produce a small white reflection point that can be utilized to detect eye features [1]

These methods also use shape based feature extraction which involves inferring gaze directions on the basis of detected eyes shapes such as center of cornea or iris edges [2] [5]. Some other gaze estimation methods involve using head mounted

display and camera mounted on glass [4] that eliminate the need of high resolution cameras and significantly improve accuracy. Gaze estimation can also be performed by training the model in which some feature extraction is performed on the basis of huge sample of training images, and then these unique features extracted are used to train the model. Then eye images are given as input and corresponding gaze direction is predicted on the basis of trained model [3]. Such methods are extensively popular as they can effectively handle low resolution images.

The images used for gaze estimation are captured using a consumer grade camera such as a webcam [1] [3]. From the captured image face Detection can be conducted using HAAR cascade Algorithm which is then used for eye detection process as well [1] [2] [3] [5]. HAAR which is an implementation of Viola-Jones Framework present in OpenCV, is used in which the simpler representations of the images are extracted and integrated into a high-level representation. Binary thresholding can be used to transform colored image into its grayscale form.

### 3. PROPOSED SYSTEM

The system primarily aims to make communication more facile for disabled people with the help of gaze direction frames. The Fig. 1 depicts the pipeline of the proposed system that contains four principal stages. The pipeline is fed with a typical webcam image as an input, from this input image the face area is detected, thus reducing the search space for eye region detection, which can thus be extracted efficiently. In the next step the eye images are normalized. Convolution Neural Network model is utilized in order to estimate the states of the eye image and hence eliminate falsely detected eye images. CNN can be used to learn the mapping between the gaze direction and corresponding eye images which thus predicts the intended key that is to be printed.

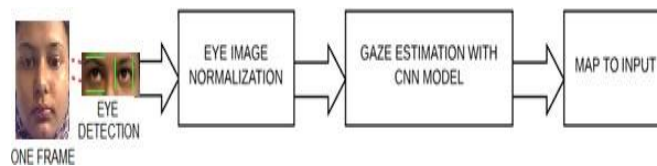


Fig. 1. System Flow

#### 1] Face Detection and Eye Detection

The proposed system uses a moderately scaled dataset consisting of eye gaze images. The images are captured using a 720p HD Webcam. These captured images are then down sampled and converted to grayscale since color channels are extremely redundant. Face Detection from the captured image is done by using HAAR cascade Algorithm which is then used for eye detection process as well. HAAR which is an implementation of Viola-Jones Framework present in OpenCV involves training an algorithm which is given a set of both positive and negative images as inputs. The positive and negative images here indicate face and non-face regions respectively. Thus it assimilates to discriminate features owned by a face region from features owned by a non-face region using a simple threshold function.

After face detection, eye detection is performed from the face region. The images obtained after eye detection is then normalized and rescaled by the use of Histogram normalization method. Binary thresholding is applied on eye images to get eye gaze directions. These images are then further grouped into five different classes based on the state of the eye image. This is then followed by training a CNN for Gaze direction estimation.

#### 2] Gaze Estimation with CNN model

The eye gaze images are classified into five distinct classes namely up, down, left, right which can be distinguished on the basis of gaze directions of the user and a blink class. A sample eye gaze image indicating up direction is shown in Fig. 2. The CNN is used to classify the input image into classes of directions based on the training dataset. The dataset created is utilized to create a Convolution Neural Networks (CNN) which would help in classifying the eye gaze direction. This category of neural networks have been proven to be beneficial in the areas of work such as image recognition and classification and hence could be used to classify the eye gaze direction. The Neural network consists of four major functions namely Convolution, Non Linearity (ReLU), Pooling or Sub Sampling, Classification (Fully Connected Layer).



Fig. 2. Eye gaze image representing Up direction

#### 4. Mapping to input

In the designed virtual keyboard there is a Text Area to display the desired output. iGaze is a pointer movement based system, where in order to print the desired number the user has to traverse from the initial pointer position to the desired key using the gaze directions. The Convolution Neural Network is trained in such a way that eye gaze images can be utilized for efficient pointer movement in their respective directions and the image representing blink can be used to perform selection of the intended key thus printing the desired value on the Text Area. Initially the pointer is set to the key present in second row and second column. The CNN predicts the probable gazing direction on the basis of input frame gaze sequence fed to it. Based on the prediction of the CNN model the pointer movement is done considering one image at a time. If an eye image gazing down is given as input to the system then the CNN classifies it to be belonging to DOWN class, thus the pointer moves down from its initial position and updates its position. This similar procedure is followed for all the directions and when the desired digit becomes the pointer, the user prints it by performing a simple blink.

A virtual keyboard as shown in Fig. 3 has been designed that consists of the numbers 1-9 and a text area to display the desired output. The red highlighted button in the keypad indicates the initial pointer position. The pointer moves in appropriate direction on the basis of the gaze direction estimated by the CNN for the input sequence fed to it.

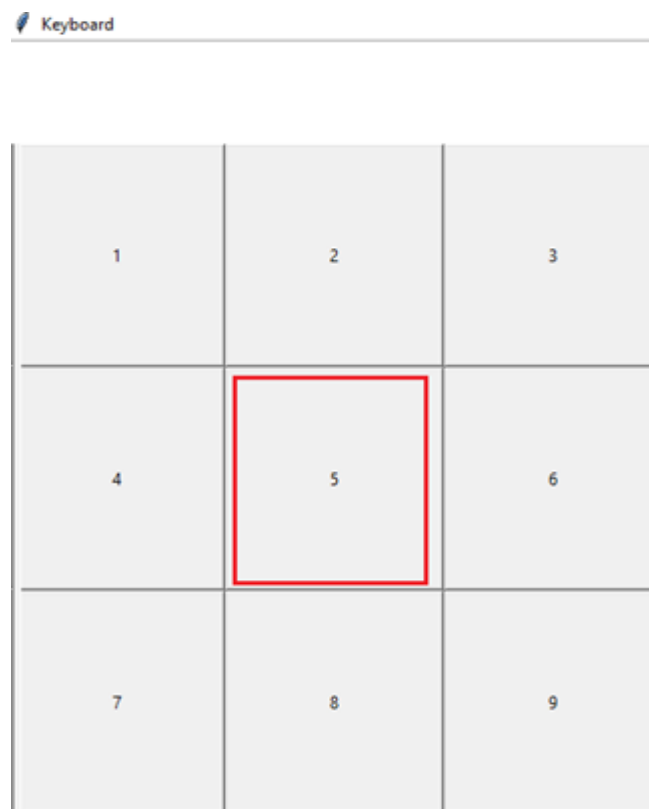


Fig. 3. Virtual keyboard with Initial Pointer position

### 5. RESULTS

An efficient gaze estimation model was trained with the help of CNN using the created dataset. The iGaze system was developed on Windows 10 operating system, Intel i5 processor having 8GB RAM. The project was coded in Python

3.5. As shown in Fig.4, an output is printed on text area considering one from the multiple available paths required to print that particular sequence, an input sequence consisting of eye gaze frames representing left, right, down, up and blink gaze directions is given as input to the system. From its initial position the pointer moves to the left after classifying the frame, then after a blink is encountered the value set for the current pointer is printed on the text area given. This procedure is followed until the end of the input sequence.

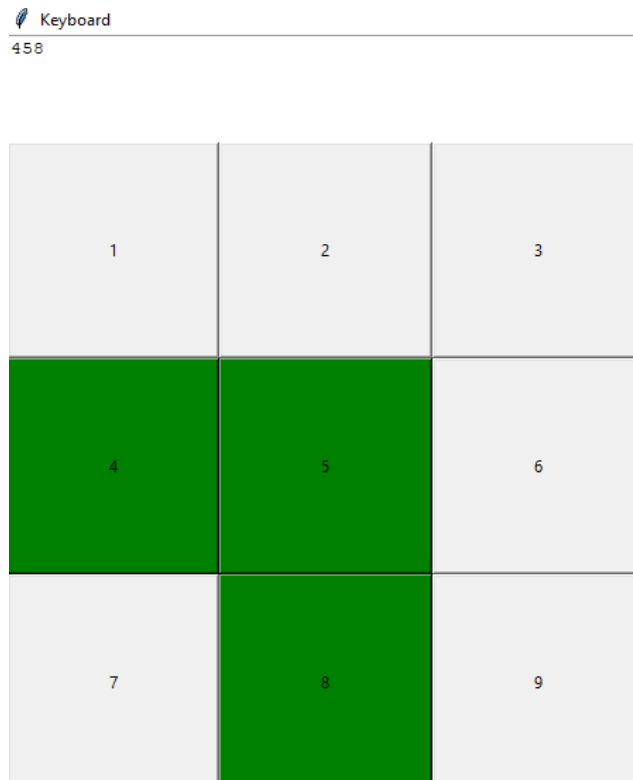


Fig. 4. Key Selection by the system with output in Text Area

The proposed system does not require any calibration process. In addition, the iGaze system also proves to be vigorous against any body movement. The accuracy and efficiency of this model was tested taking into consideration random sequences of numbers having varied length. In addition different paths were considered to display the same sequence. This helped in evaluating the accuracy and complexity of the system as shown Fig. 5.

The graph shown in Fig. 5 depicts the relationship between time and the varying paths that were considered to print the same sequence. The system exhibits a decent performance, is robust to various users. Prediction accuracy of 57 percent for same sequences was determined by considering different eye gaze images in order to print varied length output.

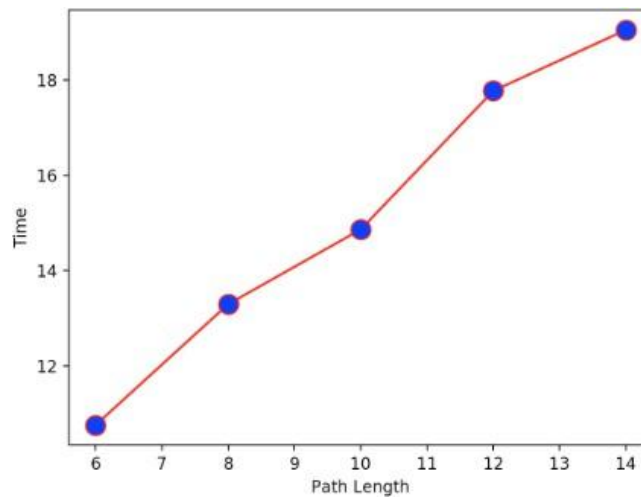


Fig. 5. Complexity Graph

## 6. CONCLUSION

A Vision based text entry system is designed using which the users can input text by eye movement, which will provide comfort to paralyzed and disabled people to convey the message and communicate with the external world easily. A Convolution Neural Network is trained for gaze estimation based on eye gaze images and corresponding key value will be displayed on the text area. Eye gaze images representing left, right, up and down directions are used for pointer movement and blink is used to make selection of the intended key. Use of CNN model focuses to ameliorate the existing system and increase the robustness and accuracy of the proposed system in terms of time management and speed. The proposed system precisely will decrease the writing to typing ratio which will make it more herculean. The system could easily be expanded to support diverse features such as word prediction and various language models and could also be integrated with home automation system and text communication via Email.

## REFERENCES

- [1] Andrea Keil, Georgia Albuquerque, Kai Berger, Marcus Andreas Mag-nor. "Real Time Gaze Tracking with a Consumer-Grade Video Cam-era". International Conferences in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG), 2010.
- [2] Vidya Itsna Saraswati, Riyanto Sigit. "Eye Gaze System to Operate Virtual Keyboard. IEEE". International Electronics Symposium (IES). 10.1109/ELECSYM.2016.7860997, 2016.
- [3] Chi Zhang, RuiYao, Jinpeng Cai. Efficient Eye Typing with 9-direction Gaze Estimation, Multimedia Tools and Applications Journal, Vol. 77, pp 19679-19696, 10.1007/s11042-017-5426-y, 2017.
- [4] Arai, Kohei., Mardiyanto, Ronny. "Eye-Based HCI with Mov-ing Keyboard for Reducing Fatigue Effects", Eighth Interna- tional Conference on Information Technology: New Generations. IEEE.10.1109/ITNG.2011.80, 2011.
- [5] Veena N. Hegde, RamyaS, S. Kumuda. "Low cost eye based human computer interface system". IEEE Annual India Conference (INDI- CON).10.1109/INDICON.2016.7838906, 2016.