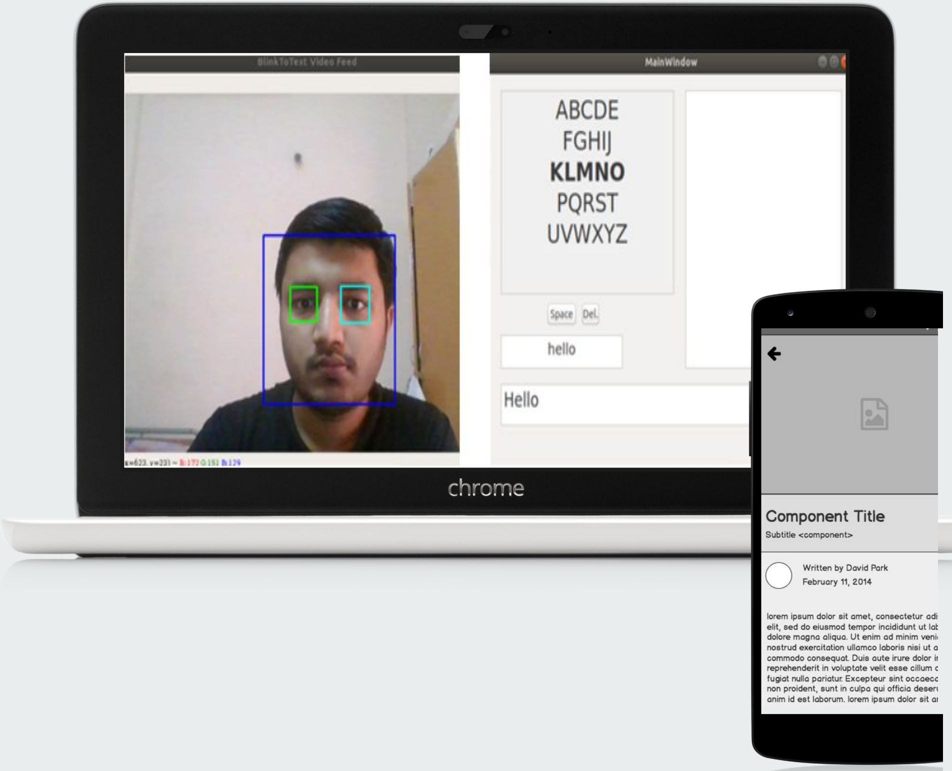
BTP Presentation, June 2020

Virtual Keyboard

Vision based Communicator

Avish Kabra 160108044
Chandan Agrawal 160108043

Under the guidance of Dr. M.K. Bhuyan



Outline

Detection Methods

Gaze estimation

Communicator Controls

Observations

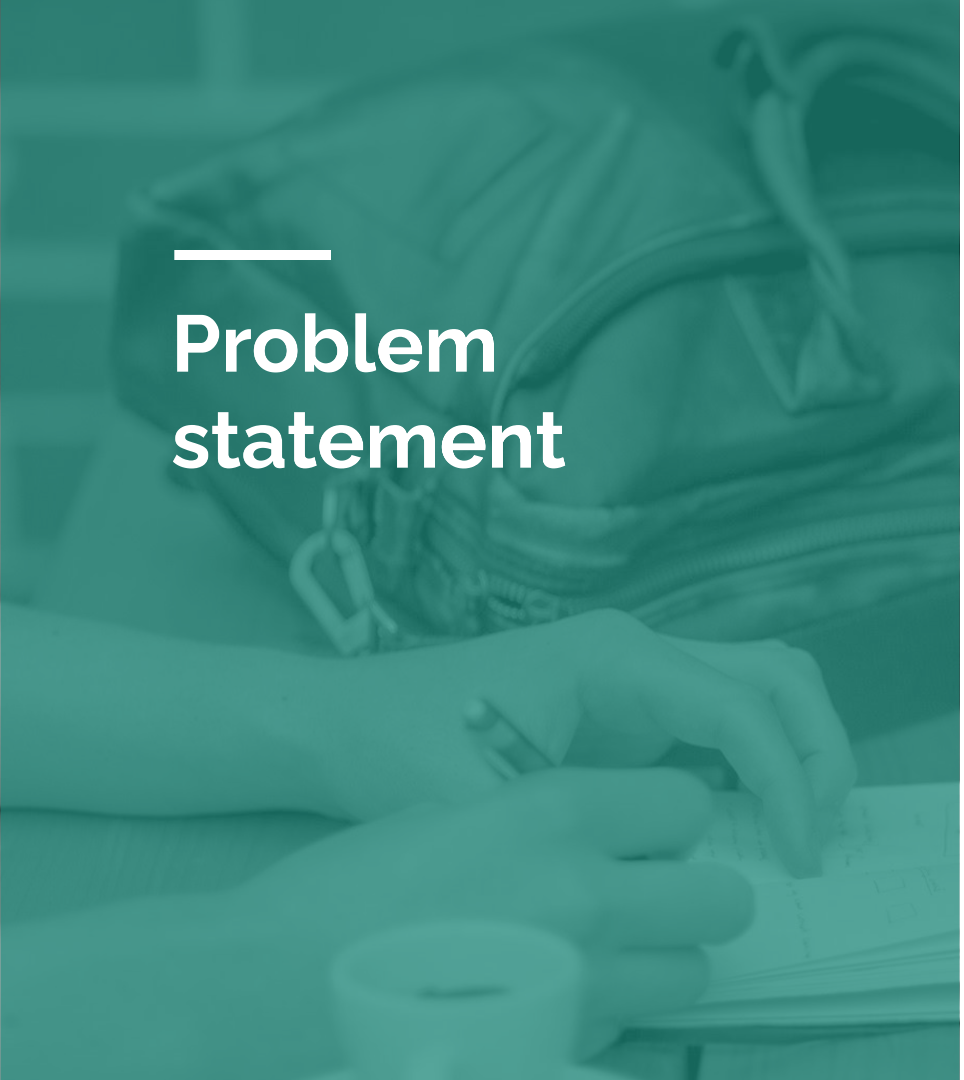
Conclusion and Future Work

The Problem

- One out of 50 people suffer from paralysis
- In severe cases, an affected person cannot perform a majority of everyday tasks, and the inability to communicate with others is one of them

Finding a way

- Nerve connections that control blinking connect directly to the brain not the spinal cord
- Therefore we use these voluntary blinks and Gaze direction as an input mechanism for our communicator



Problem statement

In severe cases of paralysis, an affected person cannot connect with others

Developing a mechanism which can be used by someone with limited mobility to communicate with people.

What do we offer

A variety of ways to type using eye & facial movements

- ➢ Type using Eye blinks
- ➢ Cursor Control
- ➢ Scroll using face movement
- ➢ Gaze detection
- ➢ Heatmap to understand our users

Survey on Eye Tracking

01

Oculography is a method of documenting the location and movements of the eyes.

Electro Oculography:

location of eye found by detecting variations in the skin potential around the eye

Sceleral Search Coils:

Small wire coils are embedded in a contact lens to measure the human eye movements

Survey on Eye Tracking

02

Oculography is a method of documenting the location and movements of the eyes.

Infrared Oculography:

The intensity of reflected infrared light is measured and disparity is used to find pupil

Video Oculography:

Video-based eye tracking

- ➢ Single Eye Camera Tracker
- ➢ Multi Camera Eye tracker

Haar Cascades

- ➔ Machine learning-based approach that is used for object detection.
- ➔ Proposed by Paul Viola and Michael Jones
- ➔ Considers neighboring rectangular regions in a detection window at a specific location, sums up the intensities of pixels in each region, and measures the difference between these quantities.

Integral Images

- ➔ Calculating features for all possible cases takes a lot of time, and large amount of computation
- ➔ The sum of the subset grid of any rectangular grid can be efficiently and quickly calculated using Integral Images

41	67	34	0	69	24
78	58	62	64	5	45
81	27	61	91	95	42
27	36	91	4	2	53
92	82	21	16	18	95
47	26	71	38	69	12

41	108	142	142	211	235
119	244	340	404	478	547
200	352	509	664	833	944
227	415	663	822	993	1157
319	589	858	1033	1222	1481
366	662	1002	1215	1473	1744

Adaptive Boosting

- ➔ Most of the features calculated are irrelevant
- ➔ Selects best features from all features we have
- ➔ Creates a "strong" classifier as a linear combination of simple weighted "weak" classifiers

Cascade Classifier

- ➔ Used to form a strong classifier from a large number of Haar features
- ➔ Mainly interested in true positives therefore rejects negative samples very fast
- ➔ Detection stops at false-negative therefore it’s rate must be slow
- ➔ Faster than linear SVM and HOG

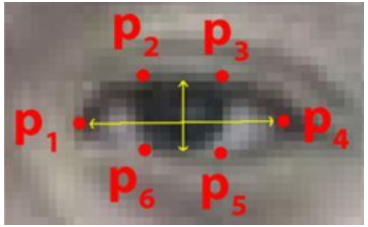
Landmark Detection

- Identifying and monitoring main features o the face
- Two main steps
  - Detection of face in the image
  - Detecting the principal facial structures
- Positions of facial characteristics are found out using intensities of respective pixels by training a group of regression trees

Blink Detection

Using Sudden Fall in Eye Aspect Ratio

EAR=(||p2-p6|| + ||p3-p5||) ÷ (2||p1-p4||)



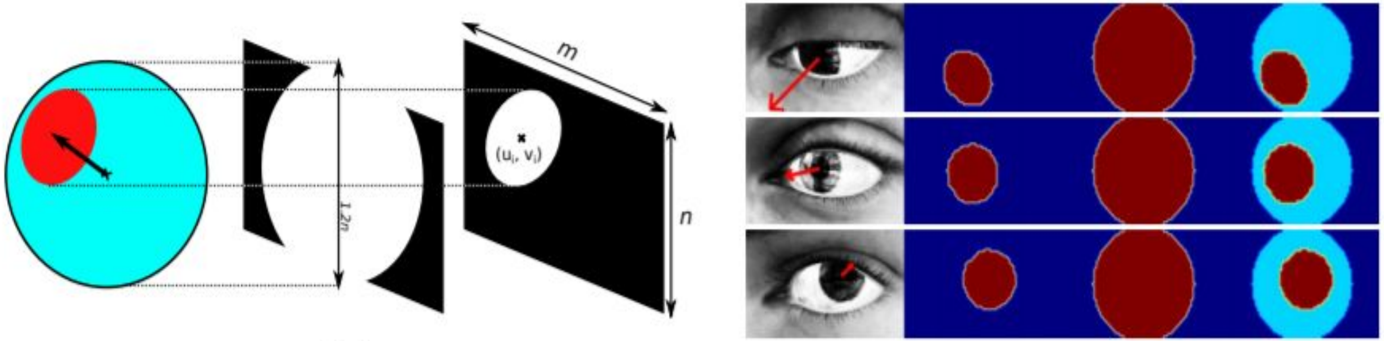
Eye localization.

Finding the white parts of the eye.

Tracking if the white area of your eyes disappears for a period of time which indicates that a person has blinked

Gaze Estimation

- ➔ Inspired by the estimation of human pose using confidence or heat maps
- ➔ Using gazemaps as pictorial representation of iris and pupil at the center of the eyeball
- ➔ Spherical eyeball as a circle and circular iris as an ellipse is projected onto the image plane
- ➔ Vector connecting the ellipse and the larger circle’s center defines direction of gaze



(a)

(b) Example gazemaps from UnityEyes

Gaze Estimation

Pictorial Representation of 3-D Gaze

- Introduces intermediate representation of the input eye image
- Gaze estimation consists mainly two steps:
  - Reducing input eye image into gazemaps
  - Estimation of gaze using previously estimated gazemaps
- This reduces learning complexity of gaze estimation

Neural Network Architecture

➔ Interpreting gazemaps from input eye image needs a fully convolutional architecture for regression

➔ Stacked hourglass architecture is used for reducing input eye image into gazemaps

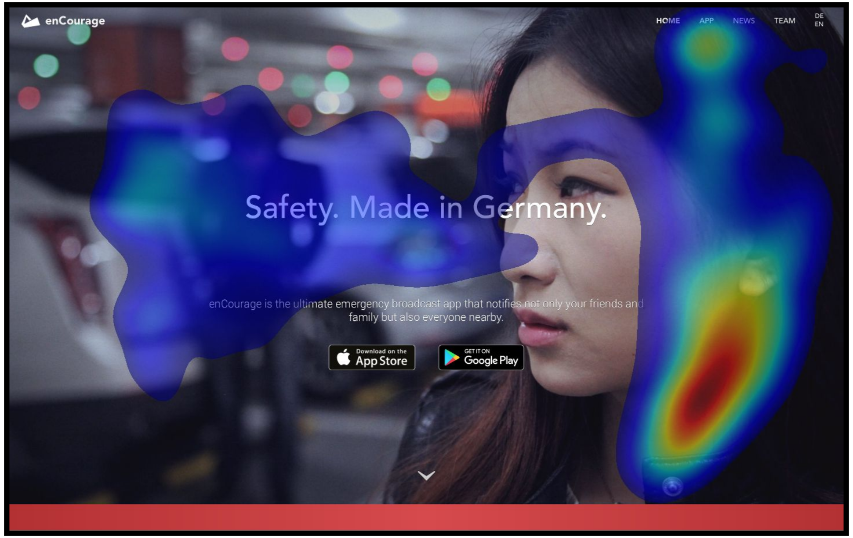
➔ Multi-scale refinement is performed repeatedly in hourglass architecture for feature maps

➔ Allows multiple-scale processing of the input image over several layers, with the required features matched with the final result of gazemap representation

Gaze Point Heatmap

➔ Heat maps are visualizations of the data that can effectively explain aspects of visual behavior

➔ In our case, The study involves changing keyboard layout to increase understanding and to facilitate a better rate of response to the action performed for a particular key click.

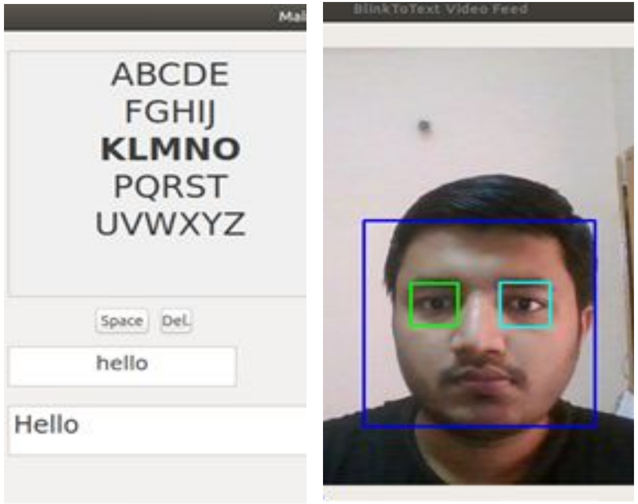


Communicator controls

- ➔ The main window of our communicator consists of the 26 alphabet tabs, Space, backspace key, a word predicting panel, and white space
- ➔ It is presented as a matrix with groups of consecutive alphabets sequentially broken into rows
- ➔ A row determining sliding bar keeps moving across
- ➔ When a blink is detected, that particular row is selected, and a column determining bar start sliding across all the letters of the selected row

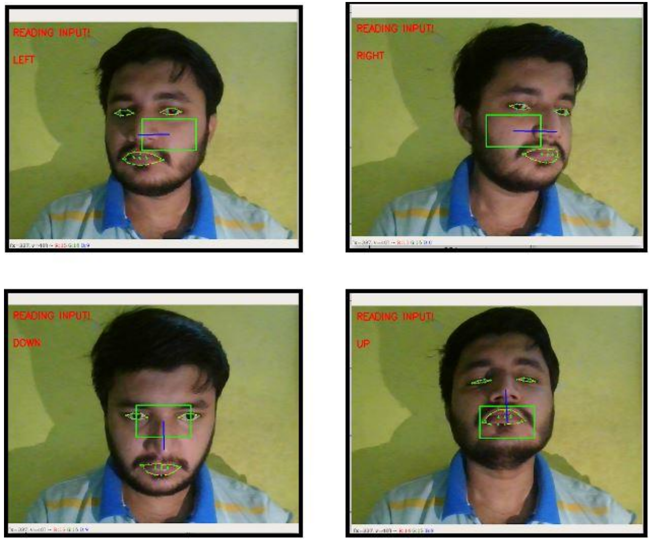
Communicator control window

On detection of the second blink, the corresponding letter is selected and displayed in the display bar

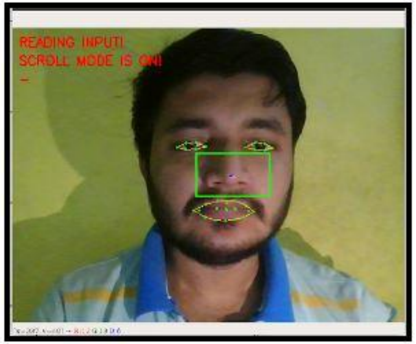


Cursor Pointer and Scrolling

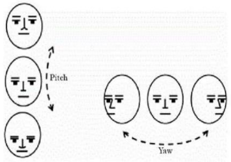
- ➔ Providing multiple functionalities using only eye blinks is difficult and might lead to confusion
- ➔ We provided a way by which a user can decide which operation he has to perform and can enable that particular functionality at that point in time.



Using Scrolling Mode

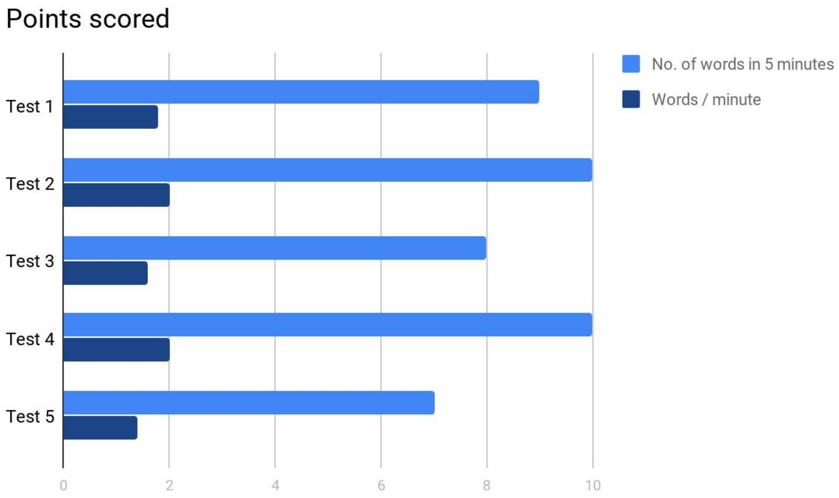


Control Gestures

Action	Function
 Opening Mouth	Activate / Deactivate Mouse Control
 Right Eye Wink	Right Click
 Left Eye Wink	Left Click
 Squinting Eyes	Activate / Deactivate Scrolling
 Head Movements (Pitch and Yaw)	Scrolling / Cursor Movement

Observations

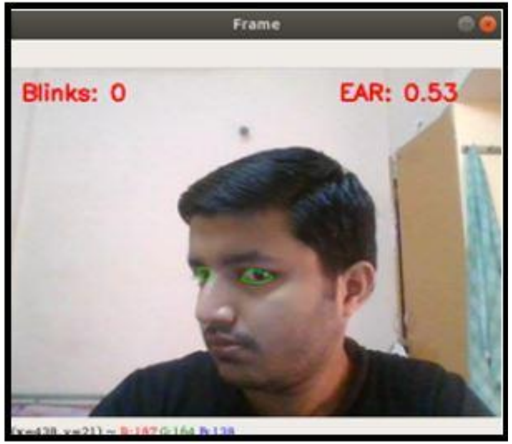
- To calculate the number of words a user can write per minute we placed the laptop in front of the user's face at a distance of 60 c.m.
- We started the software and timer for 5 minutes and noted the number of words a user could write in the given interval.



Observations

➔ It was observed that using Landmark detection provided better results than other algorithms.

➔ It was able to detect blinks for a wide range of face orientations.



Conclusion

- It can provide an easy way of communication with disabled people who are unable to move voluntarily
- Doesn’t need any high-end hardware or costly devices to operate, so it could be made available to people with a poor financial background also.
- Research findings have shown the possible predictive strength of eye tracking in autism treatment, and related information can be found using heatmap produced from user data
- Eye-tracking is commonly used in psychological experiments such as implicit connection test

Future Work

- ➔ Algorithms to detect blink of the eye could be more robust and accurate
- ➔ A working commercial platform with features like sending written sentences to other devices and many more could be launched in future
- ➔ The underlying tech can be used to create a diverse range of tracking platforms each catering to the needs of various kinds of customer base.
- ➔ Looking at the business point of view eye tracking can reveal:
    - What people see on a computer or in the actual world
    - When such visual elements obtain the attention
    - How long every attachment lasts
    - The sequence wherein these visual elements are received by the looker
    - When a person's attention returns to a visual item that was explored before

Thank You