investigation of HCI systems for users with limited mobility

RIT
Rochester Institute of Technology

Ruta Kulkarni | rk6930@rit.edu   Advisor: Dr. Thomas Kinsman

Abstract

Paralysis is the loss of voluntary muscle control in which a person loses the ability to communicate using his or her limbs. In some cases of paralysis, such as pseudocoma, a person loses his or her abilities to speak also. This project examines the issues and approaches in developing a human-computer interaction system for people who can control the interface using their neck muscles. An initial design of the system is built using the facial landmark points extracted using Dlib and OpenCV. Different classifiers are compared like Decision Trees, Random Forests, K—Nearest Neighbors and Neural Networks, using explicit and automated feature selection techniques. This paper makes a small contribution in the independence of paralyzed people.

Background and Introduction

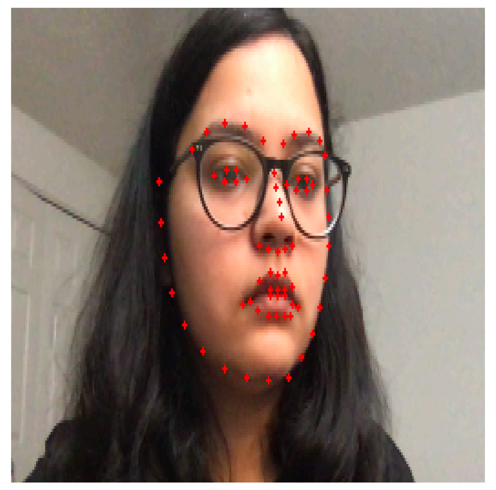
This problem space includes people with neurological damage and ways of communication they could use. Eyes and face muscles can be tracked by either mounting external sensors on the user's body or building contact-less vision-based systems. Eye tracking refers to tracking the eye movements or the Point Of Gaze (POG). A casual overview of eye-tracking methods includes methods like blink detection, eye-gaze detection and morse-codes using blinks.

Another way of communication is the Brain Computer Interfaces (BCI) systems which take brain signals, analyzes them and translates them to carry out the desired action. P300 is a wave that occurs in the EEG of the brain when the user is trying to identify the target. One of the oldest applications of P300 is in lie detectors.

Stephen Hawking, a longest survivor of ALS used the Assistive Context-Aware Toolkit (ACAT) which used infrared sensors on his and detected cheek muscles. Such systems are the motivation behind this project.

Methodology

Initial design of the HCI system



1	2	3
4	5	6
7	8	9

fig. 1 — User looking at the box 6

Using OpenCV, 225 data points are collected from the user i.e. 25 sets of 9 points for each box, each data point having 68 facial landmark points. Two features are manually computed using this data, such that they compensate for optical magnification. These explicit features are then used to train three models : 10-NN, Decision Trees and Random Forest with 50 trees.

System flowchart

Facial landmarks detection using Dlib
68 points with (x,y) coordinates

↓

Camera Calibration and Data Collection
Numerical and Image data

↓

Explicit feature selection method
left-right and top-bottom ratios

↓

Training the classifiers on these features
10—NN, Decision Tree, Random Forest

↓

Evaluation using 10 fold cross-validation
Random Forests 100% accuracy

Experiment with Autoencoders

An autoencoder is a type of neural network that learns low-dimensional manifolds of the given input in an unsupervised manner. A vanilla autoencoder is trained on the same numerical dataset using Tensorflow and Keras.

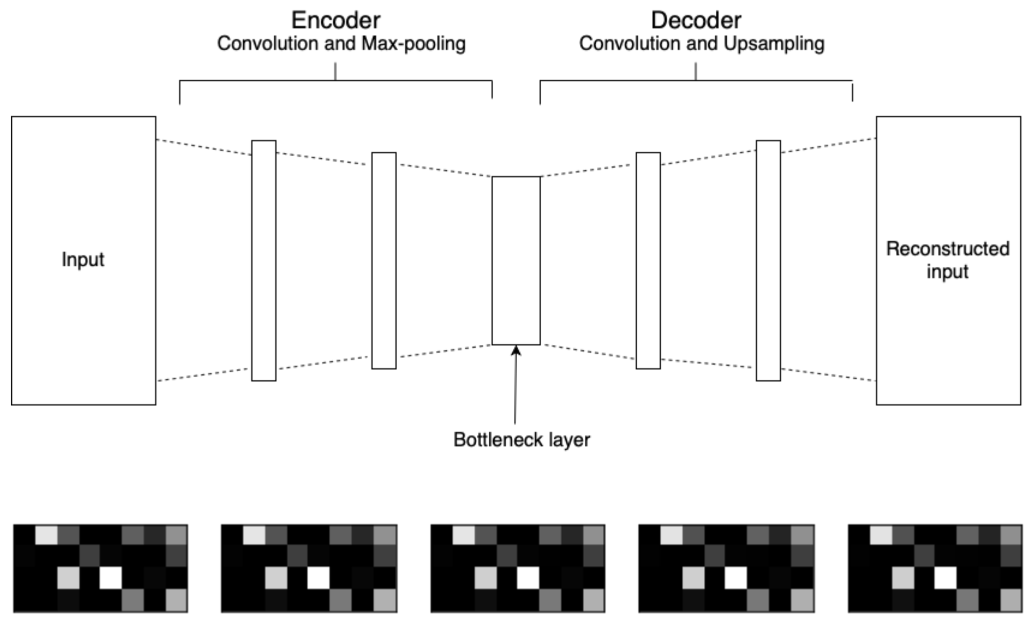


fig. 2 — Encoded features from numerical data

Second experiment is performed using the image data with facial landmarks points on a deep convolutional autoencoder, having multiple convolution and pooling layers.
Four classifiers, including neural network are then compared for the task of predicting where a subject is looking on a 3-by-3 grid on the screen.

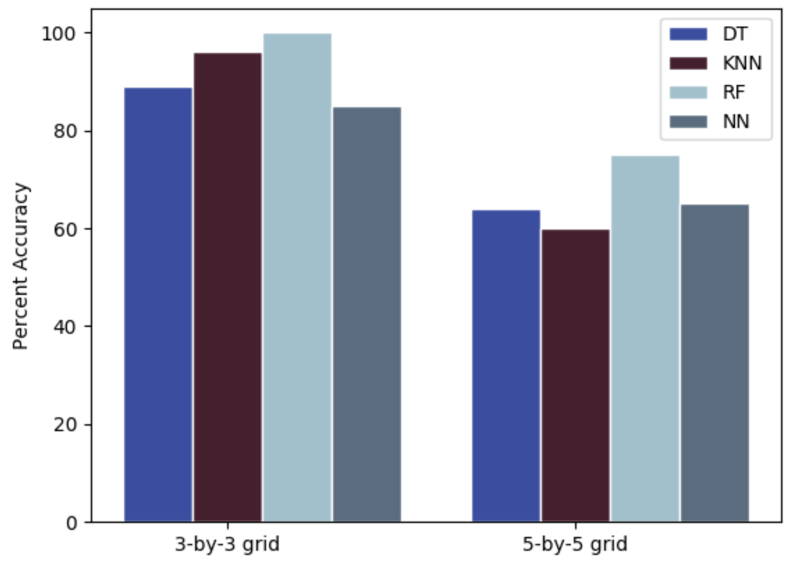


fig. 3 — Accuracy comparison for four classifiers

Results

Confusion matrix for 3*3 grid

2	0	0	0	0	0	0	0	0
0	2	0	0	0	0	0	0	0
0	0	2	0	0	0	0	0	0
0	0	0	2	0	0	0	0	0
0	0	0	0	2	0	0	0	0
0	0	0	0	0	3	0	0	0
0	0	0	0	0	0	3	0	0
0	0	0	0	0	0	0	3	0
0	0	0	0	0	0	0	0	3

Encoded and reconstructed images

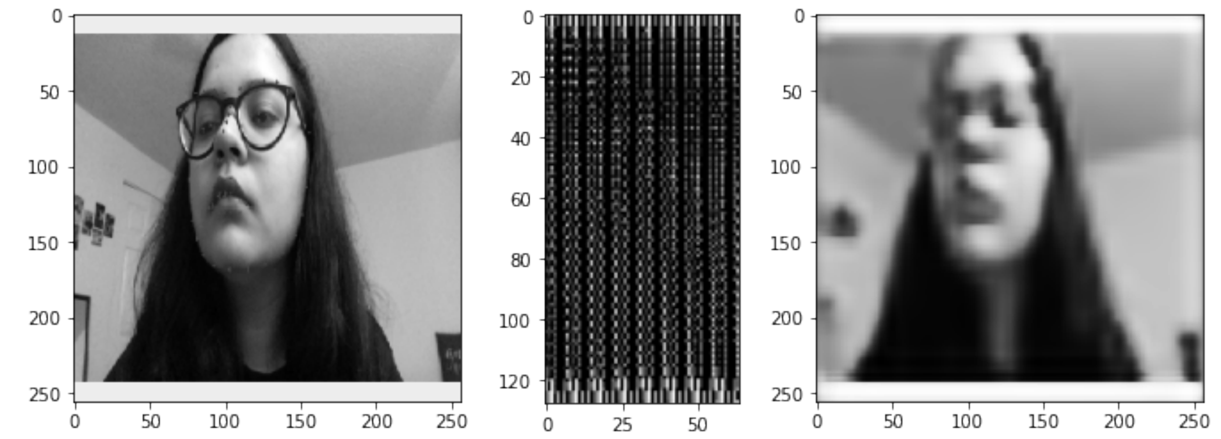


fig. 4 — Input, encoded representation, reconstructed input

Outcomes and Conclusion

- Built an application to create a 3-by-3 grid and collect head pose data using OpenCV. Wrote a Dlib installation and usage guide for the students.
- Compared four different classifiers, out of which Random Forest resulted in 100% accuracy when predicting where a subject is looking at on a 3-by-3 grid.
- Experimentation on autoencoders revealed that the encoded representations are not human-readable and did not improve the classification.
- A key result is that a simple algorithm with appropriate features beats a neural network trained with inappropriate and uncleaned features.
- Successfully built a head-controlled user interface.

References

1. Adrian Rosebrock, “https://www.pyimagesearch.com/2017/04/17/real-time-facial-landmark-detection-opencv-python-dlib/“, 2017.
2. Intel, Assistive Context-Aware Toolkit (ACAT), “https://01.org/acat/“, 2015.
3. J. J. Shih, D.J.Krusienski, and J.R. Wolpaw”. ”Brain Computer Interfaces in medicine”. Mayo Clinic Proceedings, March 2012.
4. Daniel Roy Barman, "Intelli-Cursor: A Webcam Assisted Multi-Monitor Intelligent Cursor”, 2020.
5. R. Kaushik, T. Arora, Sukanya, and R. Tripathi”. ”Design of eyewriter for ALS patients through Eyecan”. International Conference on advances in Computing, Communication Control and Networking (ICACCCN), pages 991–995, 2018.