

Flipping pages using facial gesture

Sean Yang Computer systems lab 5/13/2019

Introduction

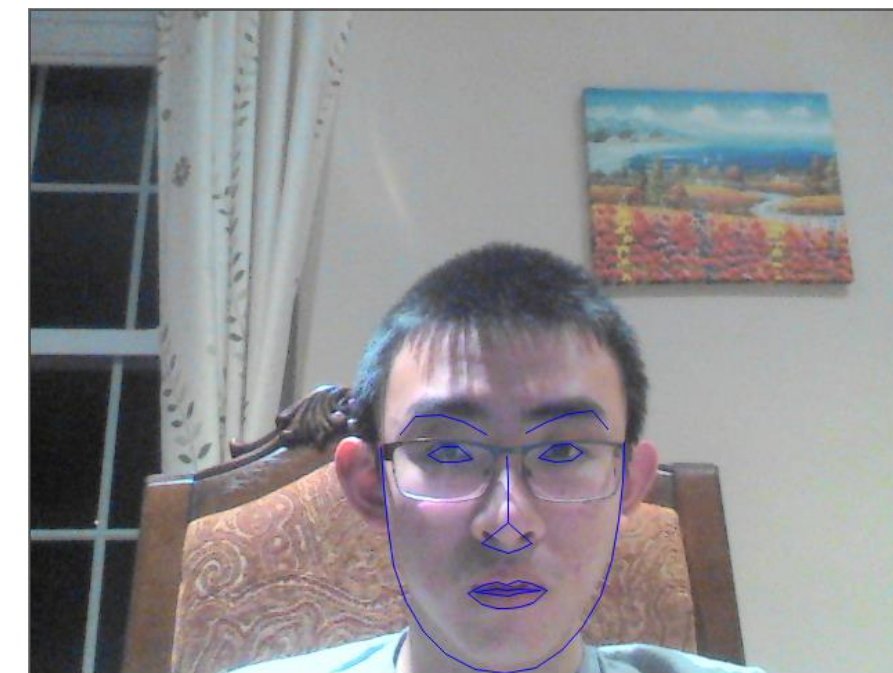
I found having to keep my hands on the keyboard while reading on my computer inconvenient. I can imagine it being a bigger issue for the people who have Cerebral Palsy with their hands. My software is designed to help people who have disabilities with hands have an easier time reading books or navigating on the computer.

Process

My program first uses Dlib library to perform facial detection and land marking detection. It uses HOG + SVM to detect the face. Then it takes the detected face and performs a 68 points facial landmark detection based on ensemble of regression trees. Then, it generates a feature vector using the relationship among the vectors compared to the center point of the face. The program has a set of recorded labeled faces to create vectors for comparison. Each image of the face has a feature vector. In the end, it uses KNN to compare the gestures to find the closest feature vectors in the labeled set. It looks for consecutive frames of non neutral gesture to minimize the effect of the misdetections.



(Figure 1) An example of an e-book app (Figure 2) An example of a facial gesture



(Figure 3) An example of the detected key points

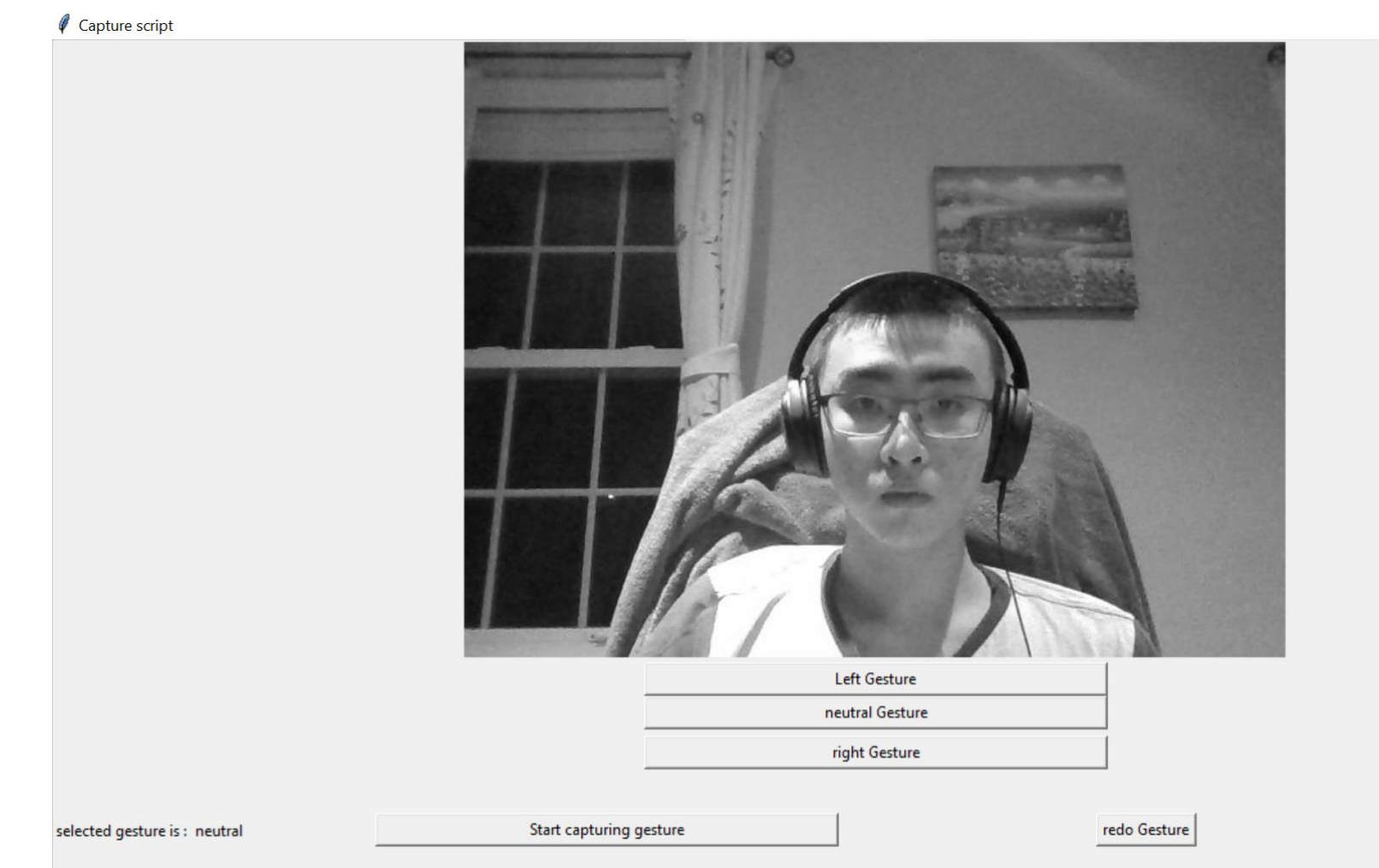
Why not CNN?

- Didn't get great result when I tried to finetune Resnet-34 with around 220 images in total for 3 classes
 - Requires much more data than KNN
 - Not as flexible
 - Training locally without a GPU can take a long time
 - Users who experience overfit or underfit might not know what to do
- HOG + SVM : First divide the image into small cells and find the histogram of oriented gradients of each cell. Then, apply bounding boxes and perform classification with support vector machine.
 - Ensemble of regression trees: Using gradient boosting to repeatedly using new classifiers to learn the error and offset that
 - KNN: k nearest neighbor classifier perform classification by finding the vectors with the closest distance in the labeled training set

Description of the algorithms

Final software

- A GUI to record training data
- An executable file to run detection and keyboard simulation to flip pages
- Runs at a good speed and each flip usually takes less than 2 seconds.



(Figure 4) My GUI for capturing gesture

Conclusion

I successfully finished what I planned at the beginning of the year. I have a fully working software that flips pages using the facial gestures. My program works the best with a training set of one's own face since it's comparing the current keypoint-vectors with the labeled ones. My program is also fairly flexible. The gestures recorded on my own face has a decent chance of working on other people based on my testing.