Mouthing Recognition with OpenPose in American Sign Language

Maria Del Carmen Saenz
June 24, 2022
The American Sign Language Avatar Project, DePaul University

Mouthing in Sign Language

Having no facial expressions or mouthing/mouth gestures in sign language, according to Baldassarri et al., “is like speaking in a monotonic voice: more boring, less expressive and, in some cases, ambiguous” (2009).

Sign languages use mouth and face gestures to modify verbs, adjectives, or adverbs; these are known as non-manual components of the sign. To have a translation system that the Deaf community will accept, we need to include these non-manual signs.

Analyzing Mouthing with OpenPose

Motion capture is one way of data collecting to analyze sign languages. Much of this data is, primarily focused on the hands and how they move. Another way of studying sign languages is by using images or videos of native signers that are already available. OpenPose is a pre-trained Neural Network that analyzes video and images for a “real-time multi-person system to jointly detect human body, hand, facial, and foot keypoints.” (Cao et. al, 2021). With 135 keypoints overall and 70 face keypoints, we will be analyzing videos of native signers which are publicly available.

Goal

This research looks at how well an algorithm can be trained to spot certain mouthing points and output the mouth annotations with a high degree of accuracy. With this, the appropriate mouthing for animated signs can be applied to avatar technologies.

Experiment Details

The National Center for Sign Language and Gesture Resources (B.U., 1999), has a significant corpus of ASL videos of native signers. It contains multiple synchronized video files showing views from different angles and close-ups of the face. The corpus is a collection of 2,617 videos in MP4 format that has been compressed from 60 frames per second to 30 frames per second.

Tools used: Python and OpenPose

Machine Learning Models: Pre-trained Neural Network and Random Forest Classifier

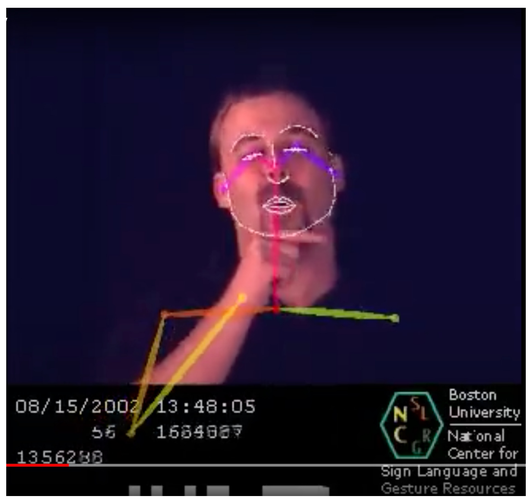
To coincide with each video, DePaul University has created an ELAN (also known as EUDICO annotation format) formatted file that groups different areas of the signer’s mouthing and mouth gestures.

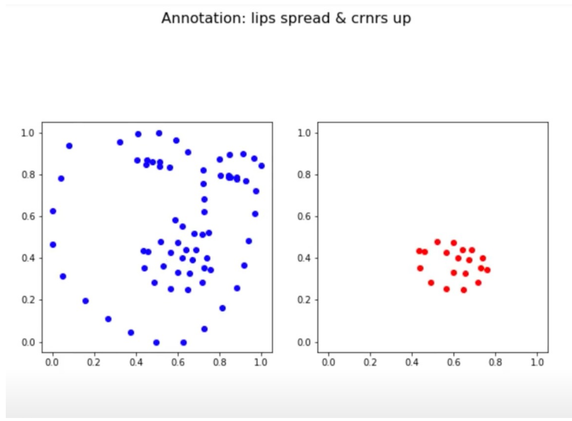
The ELAN formatted file offered many mouthing annotations, but we focused on 9 annotations with a minimum of 35 examples as a requirement.

The 9 annotations we focused on were:
- Open and corners down
- Intense
- Raised upper lip
- Lips spread and corners down
- Lips pursed: mm
- Open (as in mouth open)
- Onset (mouth movement start)
- Offset (mouth movement end)

Applying OpenPose

- OpenPose has 70 face keypoint estimations for use on the video dataset. This study focused on points 48, 54, and 60-67 which pertain to the mouth.
- After running OpenPose on all 2,617 videos the output visually shows the facial keypoints being mapped to the video.
- The next step then joined the video JSON output with its respective ELAN annotation file by converting both into data frames and joining them via timestamp keyframe.





- The joined ELAN data and video file were then converted to matplotlib animator to analyze appearing in more than 35 videos.
- The videos analyzed, 2,617 had instances of the selected annotations.

Applying Random Forest Classifier

- To take advantage of this classifier required an oversampling method called SMOTE to improve the random oversampling.
- A Grid Search was performed to obtain the best hyperparameters.
- Grid Search was done on both sampled and re-sampled data; producing the same hyperparameters.
- Validation accuracy was increased significally with resampling of the data.

Random Forest Classifier Results

Dataset	Validation Accuracy	Test Balance Accuracy
With Resampling	0.96 (+/- 0.01)	0.666
Without Resampling	0.43 (+- 0/03)	0.439

Overall, the dataset showed a higher accuracy with the resampled data as opposed to the non-resampled data in the test balanced accuracy of the model and the validation accuracy of the annotations on the facial points themselves.

Future Work

- Add ELAN notation to other publicly available corpora.
- Integrate other languages into analysis from other sign languages; German, Mexican, or French.

Paper Reference

Please see the published paper with references. Thank you!

DePaul University