Towards a User-Friendly Tool for Automated Sign Transcription (TAST): Identification and Annotation of Time Slots, Number of Hands and Handshape

Manolis Fragkiadakis, Victoria A.S. Nyst
Leiden University Centre for Linguistics, Leiden University, Netherlands
Leiden Centre of Data Science, Leiden University, Netherlands

Sign Transcription and Sign Recognition

RECOGNITION

Limitations:

- Language specific (only works for pre-trained lexical items)
- Prone to errors due to the color and motion detection algorithms
- Work on particular type of recordings (e.x. Microsoft's Kinect or signers wearing colored gloves)
- Do not work with videos with more than one signer
- Only work with recordings made in studio conditions

TRANSCRIPTION

Aims:

- Language neutral
- Unbiased for skin color
- Applicabe to all video corpora (incl. real-life & studio recordings)
- Work on multiple signers
- Short training time
- Easy to use

DATASETS

Dataset 1: 7.805 frames (352 by 288 pixels) labeled as signing or not signing from the Adamorobe SL corpus [2] and Malian SL corpus [3].

Dataset 2: 10.120 frames labeled as one- or two-handed signs as well as not-signing sequences (as predicted by the first tool).

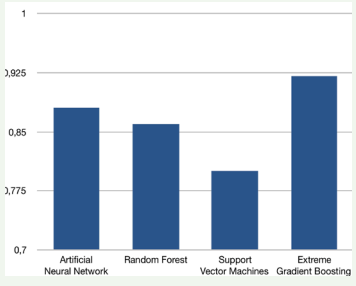
A pose estimation framework (namely OpenPose) was used to extract the positions of the hands, elbows and wrists.

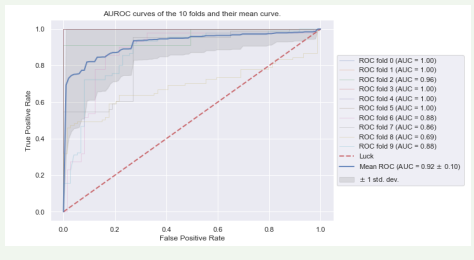
TOOL 1: Signing vs non-signing

Classifiers tested:

- Support Vector Machines (SVM)
- Random Forest (RF)
- Artificial Neural Networks (ANN)
- Extreme Gradient Boosting (XGBoost)

Extreme Gradient Boosting (XGBoost) showed the highest AUC score at 0.92.





TOOL 2: Number of hands

- Random Forest showed the highest accuracy (0.98 AUC).

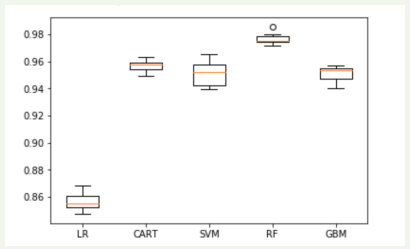


Figure 1: Accuracy of different classifiers tested for the number of hands involved.

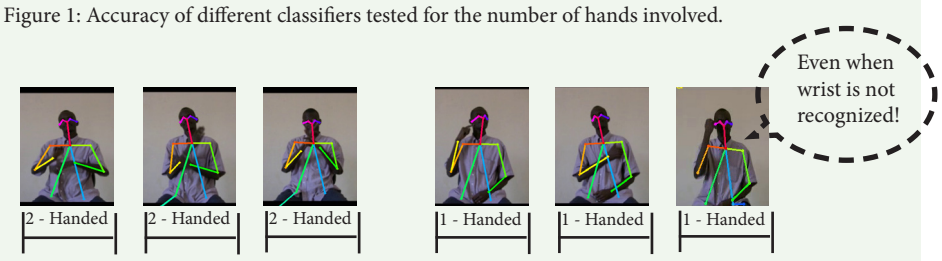
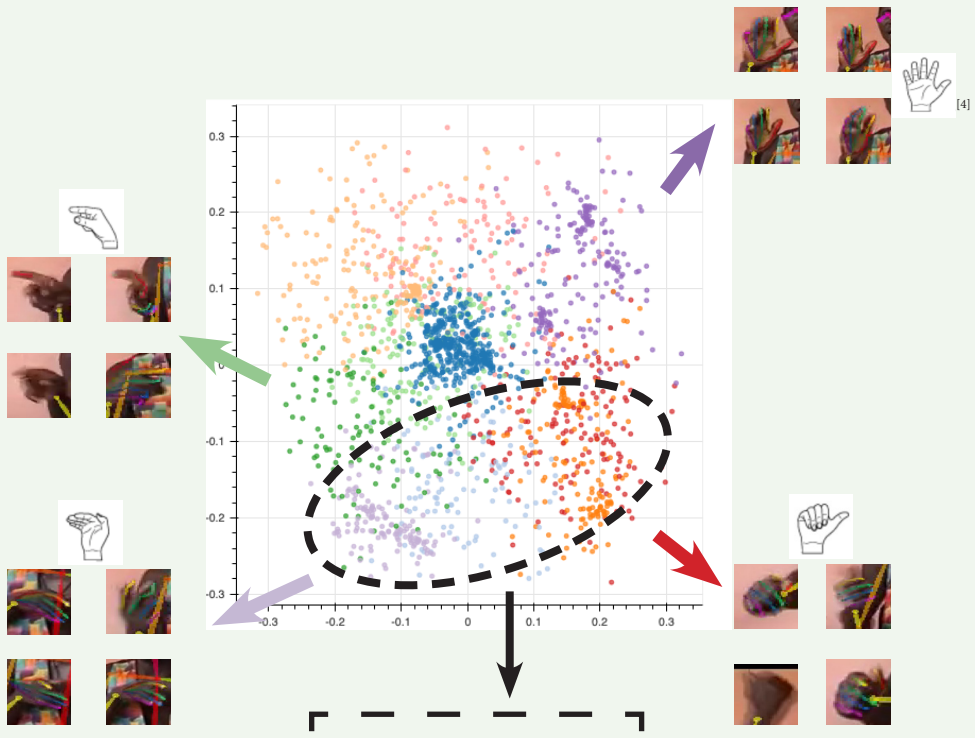


Figure 2: Predicted sequences as '2 Handed' and '1 Handed' using the Random Forest classifier.

TOOL 3: Handshape distribution



CONCLUSIONS

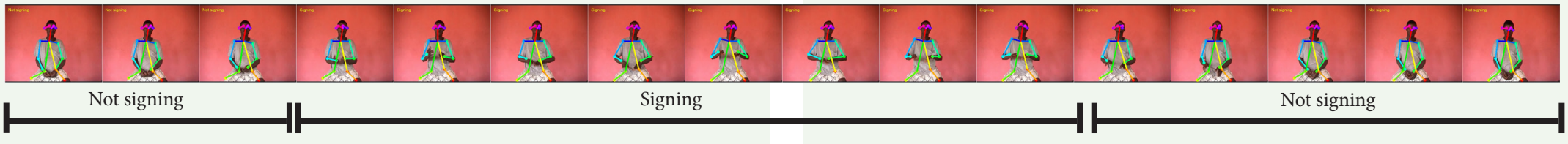
The Tool for Automatic Sign Transcription (TAST) automatically extracts and annotates signing sequences, number of hands involved and handshapes.

Its advantages are:

- Easily retrainable with only 4 minutes of video
- Lifts restrictions on the number of people and the conditions/ quality of the video
- Can be applied in gestural corpora as well.

It can be used for:

- Harmonization of sign language corpora
- Move from simple glossed corpus to ID glossed corpus
- Facilitate phonological analysis.



[1] Ronchetti, Franco, et al. "Handshape recognition for argentinian sign language using probsom." Journal of Computer Science & Technology 2016 | [2] Victoria A.S. Nyst, M.M. Magassouba, and K. Sylla. "Un Corpus de reference de la Langue des Signes Malienne II. A digital, annotated video corpus of local sign language use in the Dogon area of Mali". 2012 | [3] Victoria A.S. Nyst. "A Reference Corpus of Adamorobe Sign Language. A digital, annotated video corpus of the sign language used in the village of Adamorobe, Ghana." 2012 | [4] Kristoffersen, Jette Hedegaard, and Thomas Troelsgård. "The Danish Sign Language Dictionary." Proceedings of the XIV EURALEX International Congress. Leeuwarden: Fryske Akademy. 2010.