Sign Language Avatar Generation

Gürkan Demir, Burak İkan Yıldız
Department of Computer Engineering, Boğaziçi University, Advisor: Lale Akarun

Introduction & Motivation

- The scantness of directly matched Sign Language data restrains conducting projects and addressing several problems about Sign Language. In this work, we will not only succeed augmenting data in an automated way in the field of Sign Language but also pave the way for future projects that are focused on hearing impaired people.
- We have built a pipeline in which we are feeding the *SMPL-X*[1] library with JSON keypoints of human which are extracted by using *Openpose*[2] in order to generate realistic avatars of *BosphorusSign* data set videos.
- We have augmented the reliable sign language data by modifying parameters of constructed 3D avatars, or by rotating avatars.
- To test reliability and consistency of newly generated avatars, we have trained a model and test it with different RGB, Avatar and rotated videos.

Datasets

As a dataset we have used subset of *BosphorusSign*. There are 5 different actions that acted by 6 different users ***(User2, User3, User4, User5, User6, User7)***.

Actions in Turkish:

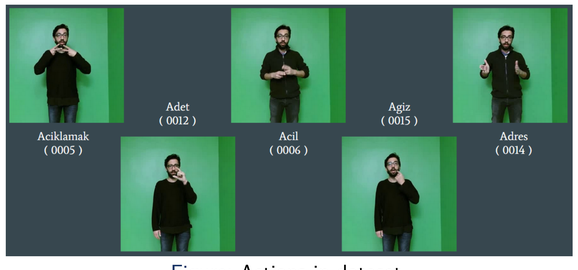


Figure: Actions in dataset.

We have used users with ids **2-3-5-6-7** for training purposes; and User4 for evaluation and test purposes.

Methods

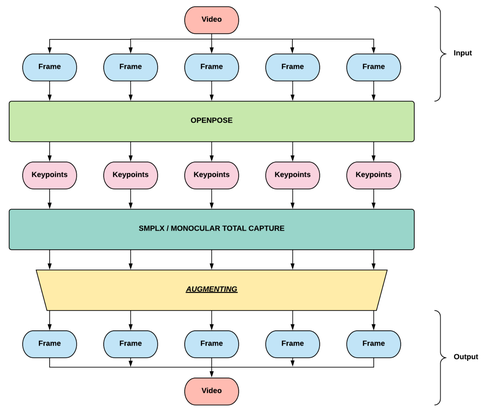


Figure: Algorithm flow.

1. Feed the Openpose[2] with generated frames in order to extract keypoints of hand, body, and face as JSON format.
2. Use the intermediate keypoint outputs of frames to feed SMPL-X[1]. SMPLX tries to encode 3D orientations of all body parts in the common 2D image space by using PAF vectors and then fit those orientations to SMPLX model.
3. Augment data by changing predefined shape parameters, or altering camera rotations etc. before rendering.

Experiment ID	Train Set	Test Set	Accuracy
1	Original BSIGN videos	Rotated Human videos	73.02%
2	Avatar videos (not rotated)	Rotated Human videos	30.16%
3	Original BSIGN videos + Rotated Avatars	Rotated Human videos	58.73%
4	Rotated Avatars	Rotated Human videos	44.45%

Table: Table of accuracy for different train sets.

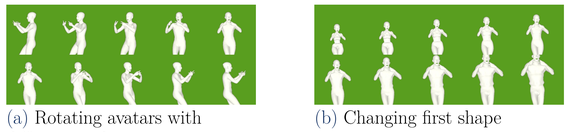
We experimented to analyze the effects of rotated avatars on action recognition. In all experiments, we used different train sets but, the test set, consisting of videos where actions are performed at different angles, were the same.

We can say that rotated avatars badly affected performance. It can be due to the fact that avatar has no clothes. In order to improve general performance of our model, dressing the avatars and changing some layers of model can be applied.

Results

Figure: Body capturing for different action frames. From top to bottom: original input, avatar, symmetric avatar.

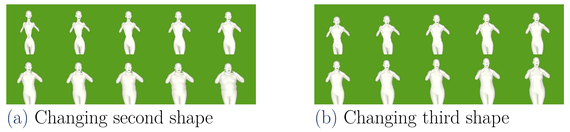
Results of body capturing in 3D coincide with the original input most of the time. However, there are some failures when there exists insufficient resolution or blur, so output differs from input.



(a) Rotating avatars with different angles.

(b) Changing first shape parameter.

Figure: Results of augmenting data - 1.



(a) Changing second shape parameter.

(b) Changing third shape parameter.

Figure: Results of augmenting data - 2.

The first shape parameter of the SMPLX model is mostly related with upper body of the generated avatar.

The second shape parameter is mostly about fatness of the generated avatar.

The third shape parameter of SMPLX is related mostly with the lower body of the avatar.

Conclusion & Future Work

1. Face, body, hand detection using Openpose[2]
2. Generating an avatar using SMPLX[1]
3. Data augmentation using avatar.
4. Testing newly generated data using deep learning.

We have successfully generated avatars and managed to increase the number of data using computer vision techniques. However, as a result of the experiments, we found that the deep learning model has difficulty in establishing a relationship between avatar and human. For this reason, techniques such as dressing clothes to avatar and changing skin color of avatar can be applied in the future.

Acknowledgements

Firstly, we would like to thank Lale Akarun for giving us the wonderful opportunity to complete our senior project under her supervision. Furthermore, we would also like to acknowledge with much appreciation the crucial role of A. A. Kındıroğlu and O. Özdemir for their continuous support. Lastly, we would like to mention the significant work of all signers who participated in BosphorusSign videos.

References

[1] Georgios Pavlakos, Vasileios Choutas, Nima Ghorbani, Timo Bolkart, Ahmed A. A. Osman, Dimitrios Tzionas, and Michael J. Black.
Expressive body capture: 3d hands, face, and body from a single image.
In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[2] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh.
OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields.
In *arXiv preprint arXiv:1812.08008*, 2018.

Contact Information

- gurkan.demir@boun.edu.tr
- ikan.yildiz@boun.edu.tr