Synthesis of Virtual Avatar Animations from Sign Language Videos

Abstract (extended)

Avatar synthesis is one of the most important and challenging tasks when it comes to sign language synthesis. The deaf community feels excluded from the new technologies and the improvements of graphics. Nevertheless, the main reason for the insufficient quality of signing avatars is the lack of sign language data, which is required when working on machine learning or deep learning techniques.

The process of generating an animation from a sign language video can be divided into a two-step process: the first step is a pose estimation problem to obtain the absolute 3D position of the skeleton, the second step can then be tackled in different ways. One possibility is to estimate the angular displacements of all joints over time using inverse kinematics and map it to a virtual sign avatar, proposed by Heike Brock[1]. This approach can only be used when the 3D skeleton obtained from the pose estimation is reliable.

A second approach which involves machine learning techniques can be used by creating a neural network that converts the data from 3D position into rotations to move the virtual avatar (these are specific 4D representations of rotations called quaternions, commonly used in animation). A visual representation of the combination of both parts is shown in *Figure 1.*

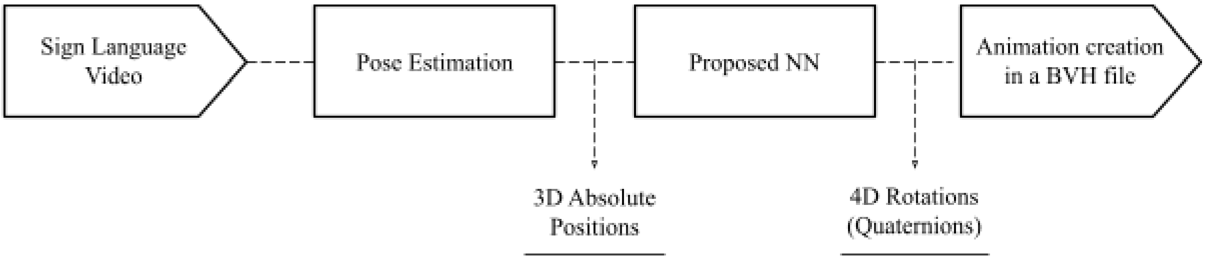

Figure 1. Proposed pipeline for animation synthesis from sign language videos.

In this method no reference nor evaluation was found, which may be due to the lack of datasets to use. In the case of this project, a new approach to create a new dataset will also be presented. The ground truth of the dataset will be the quaternions of the animations, and the training data will be the skeleton positions of the pose estimation method.

In this project I will provide a system to convert sign animations from sign language videos, divided in two steps, a pose estimation method, and a method to create the animations in a BVH (Biovision Hierarchy) file. An evaluation of several approaches for the second step will be presented. Finally, a novel automatic system to generate a dataset will also be proposed and implemented.

[1] Heike Brock, Felix Law, Kazuhiro Nakadai, and Yuji Nagashima. 2020. Learning Three-dimensional Skeleton Data from Sign Language Video. <i>ACM Trans. Intell. Syst. Technol.</i> 11, 3, Article 30 (June 2020), 24 pages. https://doi.org/10.1145/3377552

Abstract (short)

Avatar synthesis is one of the most important and challenging tasks when it comes to sign language synthesis. In this project I will provide a system to generate sign animations from sign language videos. In particular, a novel automatic system to generate a dataset will be proposed. Finally, an evaluation of different approaches to generate realistic sign animations will also be presented.