Spatiotemporal Gesture Segmentation

Amina Tabassum tabassum.a@northeastern.edu NUID: 002190127
Kassymkanova Zhibek kassymkanova.z@northeastern.edu NUID: 002920573
Manikonda Sai Malleswar manikonda.sa@northeastern.edu NUID: 002926806

February 19, 2023

1 Objective

The objective of this project is to create a classification neural network model to classify phase of gestures (gesture segmentation) based on the position of hands, wrists, head and spine in a video frame.

2 Current state-of-art

Gesture recognition and segmentation is one of the widely solved problems in human computer interaction domain for improving interactive experiences.It is challenging for continuous large domains because of multiple skin tones , cluttered background and other limitations. Different methods have been proposed to solve and improve results. [1] discusses vision-based gesture recognition and unified spatiotemporal gesture segmentation. Three major contributions are spatiotemporal matching which can identify multiple hands even in cluttered background, classifier based pruning framework which removes poor matches of gesture models in early stages of training and subgesture reasoning which can segregate between different gestures. The performance of proposed algorithm was evaluated for American sign language signers. [3] proposes two stream faster R-CNN to segment continuous gestures into isolated gestures based on hand oriented spatiotemporal features. Initially, the extracted hand oriented features from RGB and depth are fused to train SVM classifier. The proposed method outperforms for continuous large scale gesture recognition challenges. [4] discusses gesture segmentation using support vector machines focusing on solving gesture unit segmentation problem which consists of segmenting rest position from gestures in a video frame. Point of interest, position of frame of interest,

time displacement, time domain and frequency domain features are used for classifier. The best results were achieved by training SVM with simple windowed datapoint; window with 46 frames, using hands as points of interest and velocity features.

3 Approach

Our approach is to explore attributes of the dataset [2], pre-process our data , extract features , and use machine learning techniques such as Logistic Regression, LSTM, R-CNN using Sklearn, TensorFlow, Keras or PyTorch,train, test and evaluate models to choose the best performing model for gesture recognition.

4 Dataset to be used

We are planning to use [2]. It is constituted of features that were obtained from seven video recordings and each video's 50 attributes are split into two files: Raw and processed file. Raw file comprises of the user's position of their hands, wrists, head, and spine in each frame, and a processed file that contains the user's hands' and wrists' velocity and acceleration. Depth camera Xbox KinectTM is used to record dataset.

5 Progress timeline

Week of 02/27: Literature Review
Week of 03/20: Further research and implementation of algorithm.
Week of 03/27: Further research and implementation of algorithm.
Week of 04/10: Implementing algorithm and code review. Complete final project report and work on presentation

6 Deliverables

The end goal of this project is to use different neural networks and deep learning approaches learnt in this course to perform gesture segmentation on [2] dataset.

References

[1] Jonathan Alon, Vassilis Athitsos, Quan Yuan, and Stan Sclaroff. A unified framework for gesture recognition and spatiotemporal gesture segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(9):1685–1699, 2009.

[2] Dheeru Dua and Casey Graff. UCI machine learning repository, 2017.

[3] Zhipeng Liu, Xiujuan Chai, Zhuang Liu, and Xilin Chen. Continuous gesture recognition with hand-oriented spatiotemporal feature. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV) Workshops*, Oct 2017.

[4] Renata C. B. Madeo, Clodoaldo Ap. M. Lima, and S. M. Peres. Gesture unit segmentation using support vector machines: segmenting gestures from rest positions. In *ACM Symposium on Applied Computing*, 2013.