



UPPSALA UNIVERSITET

Intelligent Interactive Systems - Project Proposal

Hand Gesture Recognition to Speech

Mohammad Ashraf

Arslan Hameed

Raheel Ali

Jatin Anand

Vivek Vivian

Group - 12

May 8, 2020

Introduction

People with the inability to speak use different modes to communicate with others, one of them being sign language.[1]. The purpose of this project is to use the camera to capture gestures in the form of a video and detect different alphabets from the raw data set and make use of a virtual agent to convert the text to speech. Each gesture can be detected with a video input of the gesture and since we have 6 gestures in the data set we would be training the model to detect only the first 6 alphabets.

Objectives

Our main objectives of this project is to develop a model that accurately detects each alphabet that we input as a hand gesture and a virtual agent recognises it and speaks the alphabet it detects. To do this we would need to:

- Process the data set which was prepared from assignment-1.
- Perform feature selection.
- Implement the different classifiers to identify the different alphabets.
- Compute the performance metrics of the trained model and compare it with different models to improve accuracy.

Techniques & libraries

- The data set on which the model will be trained is provided from the assignment-1.
- Training a model that performs feature extraction with classification techniques.
- K-nearest neighbors and Convolution neural network for classification.
- pandas for reading the data from a file.
- Open CV for images and videos.
- keras.layers for deep learning and image processing.
- sklearn or tensorflow for sequential model selection.

Deliverables

We will be submitting the program code that detects the input which is a video of a gesture which makes the virtual agent say the alphabet it detects from the inputs provided.

By week-4 we can submit all the working codes and data files.

Evaluation

- Confusion Matrix.
- Precision and Recall.
- Classification Accuracy.

Specialization

After the foundational part of the model is completed we would like to use a combination of these 6 gestures to implement more alphabets other than the initial six alphabets. For example, combining the open palm followed by fist palm as a gesture to detect the alphabet 'G' and similarly we can combine the other gestures for more alphabets.

Workflow

Week - 1 Feature Selection

- Principal Component Analysis (PCA)
- Support Vector Machine (SVM)

Week - 2 Classification

- Convolutional Neural Networks (CNN)
- k Nearest Neighbour(kNN)

Week - 3 Model Selection & Training

- Multilayer Perceptron for supervised learning
- Training and Validating the model.

Week - 4 Performance Testing & Final Report

- Evaluate the performance metrics of different methods.
- Fix the bugs.
- Finalize and prepare the report.

References

- [1] Book: *Rajaganapathy, Aravind, Keerthana, Sivagami. et al.*, Conversation of Sign Language to Speech with Human Gestures ,*2nd International Symposium on Big Data and Cloud Computing (ISBCC'15)*