Sign Talk - A Sign Language Recognition System

Project by Ahmed Khan, Saim Malik, Sergio Agraz
Supervised by Dr. Rong Zheng

April 2018

1 Abstract

Many people with a hearing disability rely on sign language to communicate. However, only a fraction of the world population understands this form of communication. Sign Language Recognition (SLR) is a method that allows such people to communicate with the society. The goal of this project is to develop an American Sign Language (ASL) recognition system by using two separate Machine Learning models and comparing their accuracies. One model is trained using surface Electromyography (sEMG) signals from the user's forearm and the other is trained using RGB-D (RGB with depth dimension) data. Due to limited resources, this project is only targeting 10 letters of the English alphabet as the gestures for these letters are very distinct and can be done with only one hand. Following are the letters used.

A, B, D, G, I, L, N, P, V, W

The data collection is done using two separate hardware. The sEMG data is collected using the XTREMIS Board, designed by WiSeR [5]. The RGB-D data is collected using the Tango Tablet via its Infra-Red (IR) sensors and Fish Eye camera. The two input sources are processed separately and the accuracy of each is compared.

Overall, the results of this project are expected to give an insight on advantages and disadvantages for the two input sources mentioned. The accuracy will be based on the number of correct predictions from the test data set.

2 Introduction

According to World Health Organization, as of 2015, there were over 360 million people with a hearing disability [4]. Majority of this population has to rely on sign language to communicate. However, only a small percent of people actually understand sign language.
In this project, a Sign Language Recognition (SLR) system is developed using sEMG signals from the user's forearm and RGB-D data from the Tango Tablet. Both input sources are fed into the appropriate Machine Learning model and are then classified into the corresponding English alphabet.

3 Hardware

3.1 XTREMIS Board

WiSeR (Wireless Systems Research Group) developed a Printed Circuit Board (PCB) that is able to detect millimetre-precise finger movement called XTREMIS. It is an ECG/EMG/EEG board that is capable of collecting data at high sampling rates. It is based on the ADS1299 chip, and is inspired by the OpenBCI open source project [5].

3.2 Tango Tablet

Tango is an Android tablet that is equipped with Fisheye-lens cameras and Infrared (IR) sensor. Its API allows to collect RGB pixel data, as well as depth data as a Point Cloud. The two sources are then combined as RGB-D data.

4 Implementation

4.1 Data Collection

Raw sEMG and RGB-D data is collected from different people. This includes some people who do, and some who don't know how to speak using sign language. For the sEMG signal, five electrodes are attached to the user's forearm and one on the wrist bone for reference. The user is given three seconds to perform a sign for one of the English alphabets in front of the Tango Tablet camera to simultaneously record both sEMG and RGB-D data. Then, the entire three seconds of data is saved for both input sources. This process is repeated multiple times for each letter.
In total, 1000 samples will be collected from 10 individuals (10 samples per letter per person) for training and testing purposes. 75% of this data set will be used for training and the rest will be used as the test set.

4.2 Preprocessing

The raw data contains a lot of noise and unwanted signals. To obtain cleaner features, the raw data will go through a preprocessing stage. The Zero-Mean filter will be applied using the following equation to both raw sEMG and RGB-D data.

$$X_{\text{zero}} = X - mean(X)$$

4.3 Feature Extraction

Once the raw data is cleaned, several features will be extracted for each of the input sources. For sEMG data, the following features will be extracted.

$$Mean\ Average\ Value = \frac{1}{N}\sum_{i=1}^{N}|X_i|$$

$$Simple\ Square\ Integral = \sum_{i=1}^{N}|X_i|^2$$

$$Root\ Mean\ Square = \sqrt{\frac{1}{N}\sum_{i=1}^{N}X_i^2}$$

$$Standard\ Deviation = std(X)$$

$$Max = max(X)$$

$$Min = min(X)$$

For RGB-D data, one of the features extracted is the Histogram of Oriented Gradients (HOG). HOG is an image processing algorithm used for edge detection. It operates by dividing the image into regions and computing the gradient directions for each region. Areas with higher intensity changes are usually the edges of the object in the image. This allows to detect the hand and finger positions in our application.

4.4 Classification

Support Vector Machine (SVM) will be used as the classifier. Labelled data will be input into the SVM for training purposes. Once trained, the test set will be used to verify the accuracy of the system for both the input sources.
Couple other classifiers that were looked at include Independent Component Analysis (ICA) and a Convolutional Neural Network (CNN). ICA is commonly used for signal processing applications such as the "Cocktail Party" problem. It allows to extrapolate independent signal sources from a noisy signal. Since we have 8 different electrodes recording finger muscle movement, ICA could be useful for indentifying patterns between individual channel signals.

CNN is commonly used for image related machine learning applications. This classifier will be useful for the RGB-D data. The network will contain 10 nodes in the input and output layers, with 3 fully connected layers in the middle. The classifier will be trained with labelled data and the accuracy will be tested using the test set.

5 Conclusion

To summarize, the results of this project will give insight on the accuracy of sEMG vs. RGB-D based SLR system. However, further research is required to validate the correctness of these results since only a subset of the English alphabets were used in this project. Furthermore, the results may vary substantially for an SLR system made for recognizing words and sentences instead of letters.

References

[1] Ferat Sahin Celal Savur. “Real-time American Sign Language Recognition with Convolutional Neural Networks”. In: (2015).

[2] Yingli Tian Chenyang Zhang. “Multi-Modality American Sign Language Recognition”. In: (2016).

[3] Bo Li Qiang Li. “Online Finger Gesture Recognition Using Surface Electromyography Signals”. In: (2013), p. 102.

[4] Celal Savur. “American Sign Language Recognition System by Using Surface EMG Signal”. In: (Feb. 2015), pp. 1–10.

[5] Ala Shaabana. *XTREMIS*. URL: `https://github.com/shibshib/XTREMIS_CAPSTONE`.

[6] Stavros M Panas Vasiliki Kosmidou Leontios Hadjileontiadis. “Evaluation of surface EMG features for the recognition of American Sign Language gestures”. In: (Aug. 2006).

LICENSE TO MCMASTER UNIVERSITY

This Project Report will be written by

Saim Malik
[Full Name (s)]

Ahmed Khan
[Full Name (s)]

Sergio Agraz
[Full Name (s)]

[Full Name (s)]

[Full Name (s)]

Undergraduate course number CS 4ZP6 at McMaster University under the supervision/direction of

Dr. Rong Zheng

In the interest of furthering teaching and research, I/we hereby grant to McMaster University:

1. The ownership of one copy of this work;
2. A non-exclusive license to make copies of this work (or any part thereof), the copyright of which is vested in me/us, for the full term of the copyright, or for so long as may be legally permitted. Such copies shall only be made in response to a written request from the Library or any University or similar institution.

I/we further acknowledge that this work (or a surrogate copy thereof) may be consulted without restriction by any interested person.

Saim
Signature of Student

Ahmed
Signature of Student

Sergio Cosman Agraz
Signature of Student

Signature of Student

Signature of Witness
Supervisor

Signature of Student

April 26, 2018
Date