CS231A Project Proposal
Sign Language Gesture Recognition with Unsupervised Feature Learning

Justin Chen
Stanford University
justinkchen@stanford.edu

Abstract

This paper focuses on applying segmentation and unsupervised learning algorithms to create an accurate sign language recognition model.

1. Introduction

The problem I am investigating is sign language recognition through unsupervised feature learning. Being able to recognize sign language is an interesting computer vision problem while simultaneously being extremely useful for deaf people to interact with people who don't know how to understand American Sign Language (ASL). I will also investigate the usefulness of using Z-depth information from a 3D camera to get a more comprehensive model of the hand.

2. Methodology

2.1. Data collection

I plan to collect the data by taking videos of my hand, as well as hands from other volunteers, as I perform many sign language gestures. The data will first be taken with a clean background for an initial proof-of-concept, but as time permits, clustered backgrounds will be added in to the dataset to create a more realistic scenario.

2.2. Algorithm

Segmenting out the hands is a non-trivial aspect of the project as I will be exploring different segmentation approaches (color, texture, etc.) in order to cluster the "hand pixels" together. This will then be passed through an unsupervised learning algorithm such as a convolutional neural network to extract features and then use a trained classifier, such as softmax, to recognize specific sign language representations. In order to better classify ASL letters, I am looking into further segmenting the hand into specific fingers using depth and shape information from the 3D camera.

2.3. Background

The CVPR gesture workshop from 2011 provides a great information on modern gesture recognition models as well as how to incorporate different learning algorithms. There is some past work[1] related to my project that I will be looking at such as segmentation-robust modeling for sign language recognition [2] and sign language and human activity recognition [1].

2.4. Evaluation

The final measure of my model performance will be based on the ratio of correct classifications out of a prepared validation set (recognition rate), penalized by some factor of false-positive responses to hand transition phases. Histograms of the error rates of different approaches as well as demo images will be used for comparison.

[1] http://clopinet.com/isabelle/Projects/CVPR2011/

References

[1] D. Metaxas. Sign language and human activity recognition, June 2011. CVPR Workshop on Gesture Recognition.

[2] S. Sarkar. Segmentation-robust representations, matching, and modeling for sign language recognition, June 2011. CVPR Workshop on Gesture Recognition, Co-authors: Barbara Loeding, Ruiduo Yang, Sunita Nayak, Ayush Parashar.

3. Appendix

This project is done in combination with the CS229 Machine Learning final project. The CS231A Computer Vision primary component is the hand and finger segmentation using 3D camera.