Gesture Recognition using new Intel Real Sense coded light camera

Shane Clancy

CS 461

Fall 2019

Abstract

Communication is a large part of our day to day lives. Some people lack the ability to verbally communicate, and because of the process that they have to go through, communication tends to become a slower process. However, this can be addressed by being able to recognize gestures and process them at a much faster rate in real time than a translator might be able to do. This project aims to be able to recognize gestures using the new Intel Real Sense light coded camera, a machine learning model for gesture recognition, and an interface to display the text that these gestures translate to. This will accelerate the communication ability of someone who cannot interact verbally with people who might not be able to understand sign language at a high accuracy.

Description of Problem

There are many people throughout the world who lack the ability to verbally communicate, and because of this, many interactions are much slower. This is due to the fact that translation needs to be verbal for another person to be able to understand. This is mostly done through a human translator, who translates the sign language into verbal communication to the non-verbally impaired person. This can be a long process, since it is very difficult for a human translator to understand the sign language and verbally communicate this sign language to another person at the same time. However, if the communication can be done in real time, those who have impaired verbal communication will be able to have more complex and faster conversations that feel more interactive.

Proposed Solution

Due to new technology, and our allowed use of the Intel Real Sense light coded camera, we can do human gesture recognition much easier than it might have been possible in the past using regular cameras. Because of the light coded camera, we can develop a machine learning model that analyzes the images that we can obtain from the camera, and build a classification model around this.

Signal processing will play a key role in being able to do fast image processing. Depending on some of the specifications and the information that the Intel Real Sense light coded camera provides for us, this can make signal processing simple, but might also make us have to design a system to process the information that the device gives us. The camera is one of the key parts of the system, and making sure that the light coded camera can communicate with the other parts of the system is extremely important to the success of the project.

There are many types of classification models that we can choose to use on this information, however there are classification models that already work well with other tasks, such as image recognition and word processing, that we can start from. One of the models that has been proven to work best for these topics is a convolutional neural network, which is also known as a deep neural network. This model trains a computer to look for patterns within images, which can help find patterns that exist within different gestures. This type of model is currently being further developed and used because of how effective it can be at finding patterns by itself, which is more of a recent breakthrough in machine learning. This model can take a long time to train, but with the use of newer hardware and optimization, we should be able to train the model on an extensive set of data in order to be able to build a functional project that can work in the real world.

We plan to spend a lot of time gathering data for our model to train on, as we want the model to be as accurate as possible for all users of the program. Accuracy is very important for translation reasons, as the goal of having a human translator is because of accuracy and speed. If the model can reach an accuracy level that a human translator can have, while also being much faster, it would be a large advancement.

A usable interface is also a very important feature that needs to be a part of the project, as without a usable interface, the project will not be useful enough to the general user. We plan to create an interface that is easy to read and customizable by the user. We also are interested in implementing different types of text processing on the interface, as this system can be used for single word and letter processing, as well as sentence processing. We don't want the text to be too fast to be readable, but also want to measure precision of individual worlds and letters.

Performance Metrics

Performance metrics are extremely important for this project, as the product needs to be as accurate as possible to a human translator in order for this device to become usable. We will be able to measure this on a set of data that we will set aside for our machine learning model to test on. A good accuracy that should be aimed for is above 95% as a minimum, but hopefully the model will be effective enough after to achieve above 99% accuracy. This is definitely possible for convolutional neural networks, as current image processing models can classify any image with certainty that it has been trained on. The more accurate this project is, the more likely it will be used, which is a key part of this project.

The project will also have to be considered cost effective, since if many people would like to use this product to enhance their lives, they will need to be able to afford the product. The product needs to be profitable to an extent as well, so making sure that the product is optimized and economical is an important factor in the success of the project. The product also needs to be usable, and controllable by the user, as without this functionality, the user will not be able to use the product as it is intended and customize it to their liking.