

Hand Gesture Recognition in Real Time

Joe Holt, UW-Madison Computer Science

Fall 2020

0.1 The Problem

For this project, I propose creating a computer vision application that can recognize static hand gestures from a webcam feed in real time and then use said gestures to perform different actions on your computer. The "static" (ie a thumbs up, but not clapping) gestures will be predefined and relatively small in number (no more than 10). I want to create this application because I think it would be a good way to learn about the entire computer vision application workflow, from the image capture, to the image preparation, to the feature extraction, to the model. On top of this, I believe that an application that can track hand gestures has a lot of cool applications and would be used by a lot of people if it were to work well. Furthermore, I think this project could be expanded in the future to provide more purpose. For example, we could take the hand gesture recognizer and potentially use it to read ASL in real time.

0.2 Steps

To solve this problem, I have defined 4 distinct steps. Within each step I included an estimated time for completion.

0.2.1 Step 1 - Preprocessing Image

The first step is setting up the image pre processing pipeline. The first part of this will be to grab an image from the webcam. Initially, I will ask for images/setups that have a solid background of different color than the hand (for simplicities sake). After this, I want to apply a binary filter to the image (so that, ideally, the hand appears black and everything else in the frame appears white). At this point, I want to extract the boundary lines between the background and the hand. These boundary lines and their orientation

is what I will use as features to train my neural network. I expect this step to take 1-2 weeks and I hope to have it done by mid October.

0.2.2 Step 2 - Train Model

The next step I decided on was to architect and train a CNN that takes the boundary lines of a hand as input and outputs as a classifier to one of the N gestures. To train this model, I will use one of the many static hand gesture recognition datasets already out there (running each image of the dataset images through the same preprocessing). Furthermore, there are already some CNN's out there that do similar tasks to gesture recognition, so I may also attempt to use one of the existing models with transfer learning. I expect this step to take 2 weeks and I hope to have it done by the end of October.

0.2.3 Step 3 - Finetune Model

The thirist step I defined was to run actual test data through this model in real time. This will consist of getting the image preprocessing + NN execution working in less than 1/30th of a second. At this point, I hope to have the ability of running my webcam and having a constant output stream of gesture predictions. I will also finetune the model during this step if it does not appear to be of acceptable accuracy standards. I expect this step to take approximately 1 week.

0.2.4 Step 4 - Build Tool around Model

Finally, I will build a Python application around my completed model that can use the model with a webcam to modify settings on my laptop based on the gesture feed. One piece of functionality I will add, for example, is the ability to cause a thumbs up to increase the volume. Another piece of functionality I think would be cool to add is to have the app automatically update BBCCollaborate status based on your gestures. I expect this final step to take 2-3 weeks and I want to complete it by the end of November.

0.3 Final Remarks

Overall I am super excited on this project and am ready to get started. Please let me know if any part of this proposal seems unfeasible / too big of a time commitment to complete within a semester.