Ivy Zheng, Jerry Li, Yiyao Fu
EECS 349
Prof. Downey
May 23, 2015

**Hand in the Dark**

**Project Status Report**

Our task is to tell what kind of hand gesture a picture shows based on an existing dataset of categorized hand gestures. This technology is important because it can be potentially applied on a wide range of commands, UI's, and communication programs. For example, people can use gesture to send simple commands to the computer via a camera without touching the keyboard.

For the dataset, we use a set of 100x100 hand gesture pictures by capturing frame-by-frame pictures when taking videos:

Note that we preprocess the pictures when capturing them so that the backgrounds of the pictures are black – this is because we blackened everything without human skin colors. The only attribute we use is the grayed picture, encoded as a list of pixels. Each picture is labeled with a gesture ID, which is also what the program is going to predict.

So far, we have 1722 pictures of 11 gestures, with an average of 157 for each gesture. The gestures are taken via a camera, and a script runs to save some frames of the video and to filter them to the 100x100 pictures for learning. Right now, all the pictures are taken under fairly restrictive conditions and by a single person.

We have partitioned the data so that we can use the 10-fold cross validation for a more accurate estimate.

Based on our research, it seems that k-nearest neighbor is one of the most commonly used algorithms for our attributes in gesture recognition. For every two pictures, we compute their

difference in every pixel, square them separately, and sum them all together. The pictures with least sum are considered most close. We now consider the majority of the three nearest neighbors, but this number is subject to change.

After running the algorithm with 10 fold cross-validation, we reach 0.998 for precision. However, considering the restrictiveness of our data set, it is very likely that a more mixed data set (for example, hands taken on different lighting, positions and direction) will have lower precision.

First, we need to make improvements so that our algorithm is robust for more diverse and mixed conditions. We may consider using algorithms other than nearest neighbors, or improve the hand capturing process so that it normalize hands taken in various conditions.

We also aim to improve the time performance of the algorithm. Specifically, the classification process needs optimization for larger dataset.

If we can have robust, accurate predictions on recognizing gesture pictures, we may also try to extend the project to video recognition. Given the fact that videos are essentially sequences of pictures, we expect the approach to be pretty similar.