Lindsey Heller & Jason Mann

Visual Interfaces to Computers

Project Proposal

**The Program**

The program we will be investigating, and hopefully creating, is a Gesture Library Training Tool. This tool’s purpose is to train a user to use a specific Gesture Library, in order to ensure the user has the ability to correctly perform the set of standard gestures of the system on which user is being trained. We have often spoken in class, and explored in Assignment 1, on the difficulties of detecting specific hand gestures and wish to eliminate more than likely the most common of these. There are many factors to be accounted for when designing a system for recognizing gestures—for example the orientation, direction, rotation, etc. of a person’s hand all need to be considered for an effective recognition system. This project is therefore based largely on the theory and algorithms discussed in the beginning of the course and that were used in Assignment 1. For our purposes we will use the American Sign Language[[1]](#footnote-1) (ASL) alphabet to provide an example set of static hand gestures as it expresses good examples of similar and dissimilar pairings of gestures. The overarching goal of our project then is to through experimentation with the training and recognition of hand gestures, find the limitations of differentiation, but also provide a useful tool for any gesture based programs to come.

**Limitations**

We expect to encounter two main classes of limitations in developing our project: external and internal environment restrictions. The main external environment restriction we expect to encounter and account for at the start is background. For successful tests, we will work to eliminate background noise and effects from the start, to make detecting the hands in varied environments possible; however, we do realize that this may be out of the scope of the project, and could instead restrict the background. In addition, since the goal of the project is to create a training toolbox for any language, not just the ASL alphabet, we will attempt to make our project as generic as possible while still allowing for specific measurements of progress and success/failure. As for internal environment restrictions—this will depend on the programming language and Image Processing Toolbox we will use. As of now, we plan to be using the OpenCV computer vision library with the 2011 standard library of c++ and possibly the QT Project library for the GUI.

**References**

Kulkarni, Vaishali S., and S. D. Lokhande. "Appearance Based Recognition of American Sign Language Using Gesture Segmentation." *International Journal on Computer Science and Engineering* 2.03 (2010): 560-565.

Yoruk, Erdem, et al. "Shape-based hand recognition." *Image Processing, IEEE Transactions on* 15.7 (2006): 1803-1815.

bengal. "OpenCV Hand Detection." 9 August 2012. Git Hub. 10 April 2013

<https://github.com/bengal/opencv-hand-detection>.

Starner, Thad, Joshua Weaver and Alex Pentland. "Real-Time American Sign

Language Recognition Using Desk and Wearable Computer Based Video." IEEE Transactions On Pattern Analysis and Machine Intelligence 20.12 (1998): 1371-1375.

**Methods and Results**

The program will use live video capture to analyze static hand gestures (possibly including movement, if time permits.) Below is a sketch of the steps we anticipate in implementing this Gesture Library Training Tool, including the methods and results we will be basing our measures of progress on:

**Step 1:** **Equivalence Classes And Background Elimination**

In this step we will create different classes of gestures in the set based on similarity in shape and position, and then develop algorithms to differentiate between them. At this step we are exempting the letters of ASL that use motion as part of the gesture (namely the letters “J” and “Z”). Time permitting, we hope to expand the tutorial to include motion vectors as well although it is likely outside the scope of this project. Also we hope to also limit the effect of the background on recognition by exploring various methods (Codebook, histogram back-projection, etc.), and will decide before the next whether we will use a more constrained environment.

**Step 2: Distinguish Individual Gestures**

After the each gesture is placed into an equivalence class, we will then work to distinguish more features of specific gestures that will allow us to differentiate the entire static portion of the ASL alphabet. We hope to achieve a 90% acceptance rate under reasonable degrees of rotation and position, for at least 10 subject hands of differing shape, size and color. It is at the completion of this step we believe will be our largest milestone and the work done on the differentiation algorithms or methods that will be the true judge of our program’s performance and usability. To help in completing this task we will explore the use of Haar classifiers; potentially building our own classifier models, but perhaps using one of the many hand image databases and their requisite Haar models that already exist. These classifiers are useful in achieving quick recognition of various similar hand gestures, though they are only as good as the amount and quality of data on which they are based.

**Step 3: Binary Answers**

At this step, the recognition of the user’s gesture will be paired with a comparison to the expected gesture feedback with prompting images will be provided to the user, displaying the correct gesture side by side with the live video. Through this feedback the user will be able to continuously change their position/rotation until the tool reports that the gesture is correct. This binary feedback is not the ultimate goal for our program, but should be a good bail out point if the earlier steps take more time than anticipated— i.e. the complexity of recognizing the very similar hand gestures such a “M” and “N” in ASL.

**Step 4: Create Overlays**

This step will first require measurements of the users hand to be captured. The user’s hand dimensions will be recorded through a series of training commands, which will capture measurements specific to the current user. These measurements will be used to accurately model the dimensions of the user’s specific hand model used in the interpretation of gestures for the ASL set described above. The purpose of creating this model will be to *overlay* this expected gesture or hand model over the user’s actively pronounced hand gestures in preparation for the final step of our project.

**Step 5: Active Feedback**

This is the step where the work we will have done is put to use in a real world scenario. We will turn the above methodology into a training tool for a non-fluent user of ASL (but more generally any gesture set). This will not be a linguistic tool but purely as coaching software to enhance precision and will give a good example of simulating training for any hand gesture set. We plan to accomplish this by taking the overlays or hand models that were developed in Step 4, and overlay them on the real-time video input of the users hand. This is to allow the system to provide active feedback on what they specifically need to change to correct themselves. These models will be tailored specifically to the user’s hand dimensions in order to provide accurate and colorful overlays that will allow the user to adjust in real-time. This is crucial for seeing the minute differences between what may be a gesture similar to but not an exact representation of the correct gesture. We feel this will be useful training for any system using gestural input whether it be hand or other object visual input.

**Evaluation Metrics & Measures of Success**

The first challenge we envision is eliminating the background from the picture by expounding upon an earlier derived algorithm in Jason Mann’s project from Assignment 1 and in doing so we hope to be able to distinguish skin regions from the majority of reasonable backgrounds. This may require us to calibrate what is background beforehand through some sort of environmental training step. This will be our first milestone and point of evaluation of the success of our program. The second major challenge we envision is distinguishing between very similar gestures, such as “M” and “N” in ASL. Successfully accomplishing this task or having a reasonable percentage of successful classifications will be good metric for assessing the program as a whole. In addition, distinguishing between similar gestures while allowing a certain degree of allowance for rotation/orientation will also be a major milestone if accomplished. Our last major challenge is to be able to provide active feedback using the generated gesture overlay. Judging both the helpfulness and correctness of the overlay & feedback will be true measures of success.

We expect the main point of failure in our project to be the differentiation of similar hand gestures. If this proves true we hope to expound upon the exact bounds of similarity in a specific gesture set recognized by purely visual input. In doing this we still hope to keep our algorithms general enough to be able to be extended as a library into other projects. In exploring this we can set a standard not only for our own training program but for other future programs to come.

We hope at every milestone to test using more than 10 subjects of different hand shapes, sizes and colors to prepare and reinforce the general scope of the algorithms. We will work closely with the subjects to see how the system is running and how they are reacting to the system—for example whether or not the feedback was helpful; we will also use these tests to ensure that algorithms are correctly working and correctly generalized.

**Management Plan**

After dividing the gestures into equivalence classes jointly, each member will be responsible for certain subgroups/classes—developing the specific metrics and algorithms. In many ways this can bleed into other milestones as well while the same set of algorithms will likely be used throughout the program. Because our backend is very integrated with the “front-end” of the application, we do not foresee being able to split the work on those bounds, but if we reach our final milestone we plan on evolving the project into a more marketable user interface at which point the GUI side and the specific implementation of the training algorithms can be split once more. Another proposed split at the beginning of the project would be for one member to work on background filtering while the other begins work on refining the input/output interface and other tasks such that the interface will be ready for testing before more complex algorithms are implemented. This will allow us to use our time more wisely and also eliminate the dead time that is frequently encountered when waiting on linear completion of program components.

1. ASL is the predominant sign language of deaf communities in the United States. [↑](#footnote-ref-1)