A Direct Manipulation and Voice Interface for CAD

Will Whitney and Cristina Lozano

One of the major barriers to entry in CAD design is how nonintuitive the interfaces for current CAD software are. One major source of this difficulty is the mismatch between the two-dimensional input devices (mouse) and the three-dimensionality of the finished product. This leads to a great deal of time and effort being spent simply in manipulating the model into the proper plane for input, and this manipulation itself is highly inefficient, involving repeated mouse drags on the model, scrolling, and keyboard input.

Instead of using the mouse and keyboard to move the three-dimensional object in space, we propose a system for using a 3D camera system, such as the Kinect or Leap Motion Controller, to control the position of the part in a way that is similar to a direct manipulation style. It differs from a purely direct manipulation system in that instead of the user reaching into the virtual space and having to actually grab the part to move it, the user simply makes a grasping gesture, then can control the part as if it were in his or her hand. That is, when they “grab” anywhere in the tracked space, the system treats the current state as the mapping between the physical and virtual spaces, such that their closed hand “becomes” the part. Then however they move their hand, the object on the screen moves the same way in all axes.

We plan to apply the techniques developed in the gesture recognition papers we studied in class, mainly focusing on the hidden markov model technique. We will only be focusing on recognition of hand gestures and the set of these gestures will be fairly limited in scope. This project would be an extension of some of the techniques we explored in project three, but it would additionally include the need to segment gestures and to be able to track a continuous gesture from frame to frame. We plan to implement segmentation by declaring any large changes in relative position of the fingers to the hand as a new gesture. Any gesture that is intended to be continuous, such as rotation, will keep the hand in a relatively consistent pose and thus be recognizable as a one gesture. For tracking continuous gestures from frame to frame, we plan to compare the frames and find the nearest transformation between the two forms.

To test our project we will need to come up with a collection of samples of each gesture to train our model. We would ideally like to collect as many examples of the gestures as possible so as to have a more robust model. We will start off by collecting examples of ourselves making the ideal motion for the gesture and then we will invite friends to make the gestures as well so we can see how other people with different body shapes and movement patterns would make the provided gestures. We will then gauge the accuracy for each gesture by comparing a test gesture with the training set as well as testing how the gestures work in series.

The resources we will need to complete this project are a Kinect and a computer. We may be able to get access to a Leap Motion Controller, which we would potentially use to get finer-resolution finger data.

Depending on how deep the scope of this project should be for two people, we could potentially foresee having some cool applications of voice control as well. Having voice control would allow us to modify our objects instead of just manipulating them. An example of this might be pointing to a location on the object and saying “drill”. This addition of voice would add another interesting modality to our project and would increase the functionality of the service. The voice recognition would be manageable since we would only be responding to a fixed set of commands which would be limited to only a dozen or so words.