* General topic, motivation, background

In the 1920s, a Russian physicist invented an instrument that was made up of two metal antennas that sensed the relative position of the user’s hands. The instrument used oscillators to control frequency with one hand and amplitude with the other.

Currently, hand gesture recognition is a common problem for Computer Vision. Often, hand gesture recognition systems are used to help people such as the visually impaired or to help people in therapy. In our research, we found a few interesting papers related to hand gesture recognition. we saw that work had been done to create a hand gesture recognistion system for the physically impaired (by whom??). We noted that Solanki and Desai from Gujarat Technological University had even attempted to make a remote control for home appliances (can this in some way be connected to physically impaired people?). Work had also been done to have hand gesture recognition systems for physically impaired. In Leeds Metropolitan University, a researcher had previously done gesture interaction for electronic music performance (again, any connection to physically impaired people?). This led to the inspiration for our project.

Our main motivation was to experiment with way to connect audio sounds with hand gesture recognition. If we could connect hand gesture information to audio output, then this could add a new dimension to recognition systems. As well, it can be used in fields such as occupational therapy. We looked to the Theremin instrument for inspiration (because why? Maybe explain what the Theremin is a little here). We found papers such as Svilen Dimitrov’s “Analyzing Theremin Sounds for Touch Free Gesture Recognition,” but often the methods suggested involved looking into systems and hardware. We were searching for an elegant and simple solution that involved using only a web camera. We wanted to do something with just a simple web camera.

Specific goal of the project (Give a problem definition. Why is this an interesting and difficult problem?)

The goal of our project was to try and emulate the functionalities of the Theremin. This involved sub-problems of being able to control volume with one hand and controlling pitch with the other.

The video version of this “Theremin Problem” is an interesting problem because there is a loss of information (uhh what?). A real Theremin can use oscillators to track the user’s hands in all three dimensions. However, a simple camera does not have any information on depth. This was one of the many complications we had to find a solution for. Additionally, a virtual Theremin has to account for possible inaccuracies in hand position detection, hand movement detection, changes in lighting, and changes in the background. Furthermore, after gaining all visual data, a virtual Theremin needs to use that using that to predictably(I don’t get this adverb…) and reliably reflect a change in audio output.

* Methods (based on material from class, the literature, or on your own ideas)

Before we could output any sound, we first needed to be able to segment the hands. To do this, we decided to use hand segmentation using skin detection which is a method similar to the one used in the hand gesture homework earlier this semester. To make processing of the image more manageable, the image was split in half down the center into two separate images.

Volume:

The first half of the image was used to control the volume of the audio. The segment was then iterated through pixel by pixel. For each pixel, the RGB values were taken and thresholding was performed on their values based for skin detection (based for?). This resulted in a grayscale image, and the contours of the grayscale image were obtained using the opencv findContours function.

A function was then written to iterate through all contours in order to find the one with the largest contour. The largest contour was then taken to be the left hand. We also decided that the centroid of the contour would be the best data to be used in controlling the volume. In order to find the centroid we…

Different amplitude equations

-finally chose to use decibel

Centroid Moments

Optical Flow

-gray scale image

-skin detect

-finally skni detect and contours (best result yet)

Matrix Norm

-l2 norm of flow vectors

* Experiments

A bunch of optical flow experiments

Audio experiments

* Results
  + Volume can pan but the transition could be worked on
  + Can sense hand motion for vibrato, the audio could be worked on
* Discussion of results (Is the method successful? Are your results satisfactory? What are the limitations of the method used? Did you improve on the state-of-the-art? Give a critical evaluation.)
* Conclusions.

github:

Links:

<http://en.wikipedia.org/wiki/Theremin>

<http://www.utpalsolanki.com/project/project3/P1.pdf>

<http://davywybiral.blogspot.com/2010/09/procedural-music-with-pyaudio-and-numpy.html>

<http://stackoverflow.com/questions/9235368/realtime-sound-synthesizer-from-a-varying-input-in-python>

<http://docs.opencv.org/modules/video/doc/motion_analysis_and_object_tracking.html#calcopticalflowfarneback>