* General topic, motivation, background

In 1920’s 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 therapy. In our research we found a few interesting papers related to hand gesture recognition. Solanki and Desai from Gujarat Technological University had attempted to make a remote control for home appliances. 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.

Our motivation was to experiment with way to connect audio sounds with hand gesture recognition. If we could connect hand gesture information to audio output, we that this could add to a new dimension to recognition systems and be used in fields such as occupational therapy. We looked to the Theremin instrument for inspiration. We had found papers such as such as Svilen Dimitrov’s “Analyzing Theremin Sounds for Touch Free Gesture Recognition” but often involved looking into systems and hardware. 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 involves sub problems being able to control volume with one hand and controlling pitch with the other.

The video version of this “Theremin Problem” is interesting problem because there is a loss of information. 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. A virtual Theremin also 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, using that to predictably 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. This method was 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. This resulted in a grayscale image. 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 then decided that the centroid of the contour would be the best data to use in controlling the volume.

Different amplitude equations

Centroid Moments

Optical Flow

Matrix Norm

* Experiments
* Results
* 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.

Links:

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

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