Simon: An Interactive Sign Language Recognition Game

Using color segmentation and morphological characteristics to identify hand gestures

Jonecia Keels

Columbia University, Computer Engineering

New York, NY USA

jtk2128@columbia.edu

James Guevera

Columbia University, Electrical Engineering

New York, NY

[enter your email address]

*Abstract*—Using the Open Source Computer Vision Library (OpenCV) and included algorithms, we develop an interactive visual based game. Through heuristics, image segmentation, and morphological analysis, we are able to create a system that successfully recognizes select hand gestures from the America Sign Language chart.

Keywords-component; formatting; style; styling; insert (key words)

# Introduction

The primary objective of this research is to process static images taken with a camera of different hand gestures performing sign language. Our system, named Simon, will successfully be able to define and identify a certain grammar set using OpenCV.

# Problem

The basis of the Simon system is for it to visually identify the hand of a person. This poses some problems for color-identification because of different variables such as lighting conditions, different skin tones, hand sizes etc. Particularly color and skin identification is tricky, especially in the Red-Green-Blue (RGB) color scope. Generally in vision code, there is a range of pixel values set aside for identification of a desired object. For example, we retrieved RGB values of people of different skin tones, which produced the RGB values in *Figure 2.1*.

| African-Americans | Caucasians | Asians |
| --- | --- | --- |
| R:98-140  G:50-97  B:40-56 | R: 235-249  G: 180-202  B: 173-194 | R: 240-255  G: 212-230  B: 170-184 |

*Figure 2.1 RGB Values of different ethnicities*

Though the data generated was taken from dozens of images of people in of different skin tones, the RGB values are still not precise because lighting greatly effects how a color is shown. For example, a person taking a picture with a lot of lighting will have completely different RGB values if taken in a dim setting. There has been research done with finding the optimal method of skin detection that includes converting the color space, thresholding, and performing binarization.

Additionally another posed problem is identifying hand gestures. OpenCV includes vision algorithms such as edge and contour detection to find characteristics in shape. Now the question is: Which combined methods of skin detection and vision algorithms is best to identify sign language?

# Implementation

There are different steps needed in order to find the optimal way to implement accurate sign language recognition in our System. First, there needs to be a way to isolate the hand and remove background noise, then perform some type of combined vision algorithms to find the morphological characteristics of the hand shape, and finally performing tests on multiple gestures and test for accuracy.

## Converting RGB to HSV

An image generally includes a huge combination of different color values, however the desired pixels needed for the purpose of Simon is the ones that define the skin color of a person’s hand. As found earlier, there are many different values that can define a person’s skin color making it hard to threshold these values without noise being identified as well.

According to research, HSV is considered one of the best color space for skin detection [1]. HSV stands for Hue, Saturation, and Value. Hue particularly is valuable in that it represents the color of an image and can almost solely be used for accurate skin detection across all color tones. In OpenCV, converting a RGB image to a HSV color space is simple by using the cvCvtColor()

## Identifying Hand

Through pixel-by-pixel analysis of dozens of images, it was found that the best threshold value for skin identification is:

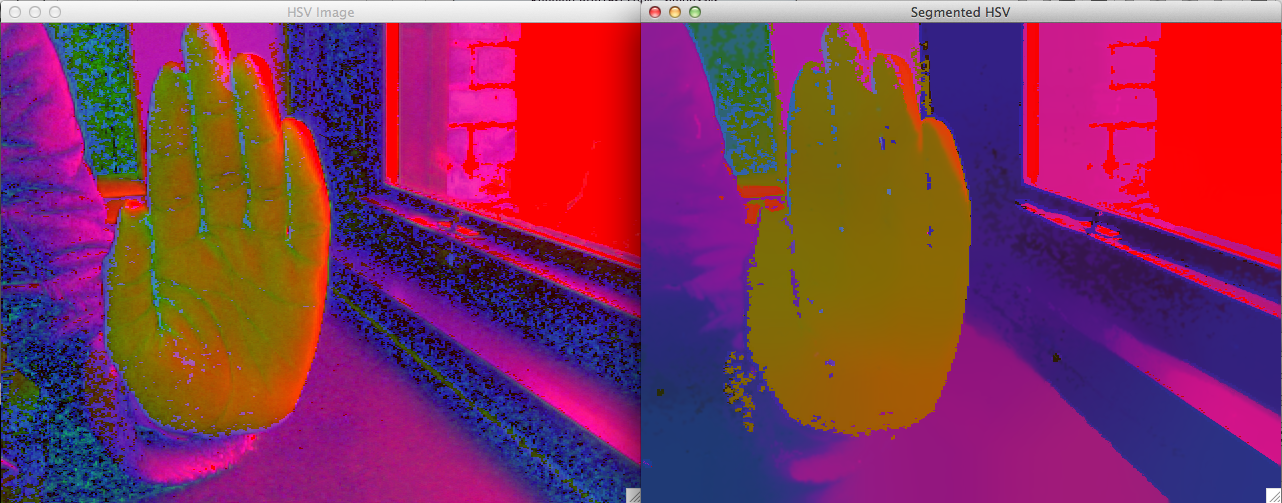
***H: 0-50***

***S: 20-150***

***V: 40-255***

### Color Segmentation

In an attempt to make the process of identifying skin pixels easier, the system performs color segmentation on the HSV image in order to combine similar valued pixels into one. By doing so, we can reduce some stray noise from the image for a cleaner binary image. *Figure 3.2* shows what the hand looks like after using the mean-shift segmentation algorithm on the HSV image



*Figure 3.2 Color Segmentation on Hand Image*

You can see how some of the stray pixels are all blobbed together in a similar colored object. Some of the detail of the image is lost as well, such as the creases on the sleeve. By performing segmentation, it is easier to identify all of the hand since they are around the same HSV value.

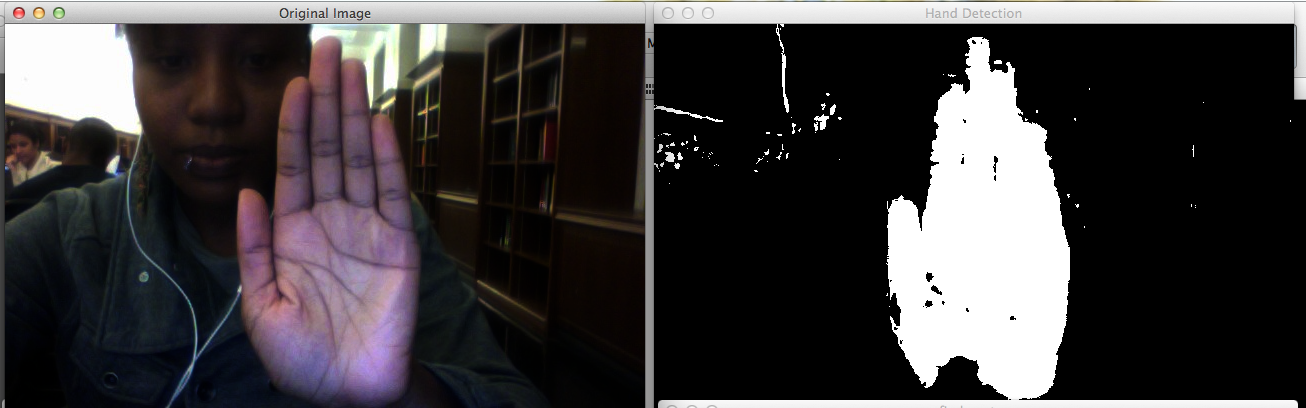
### Image Binarization

Using the threshold HSV values found earlier, we are now able to develop a system for deciding whether a pixel should be white or black, with white (value 255) being our object of interest, the user’s hands. *Figure 3.3* shows the conditions in which the system determines if a pixel should be binary or not.

*Figure 3.3 Threshold function used for image binarization*

*p(x,y) stands for a pixel at coordinates(x,y)*

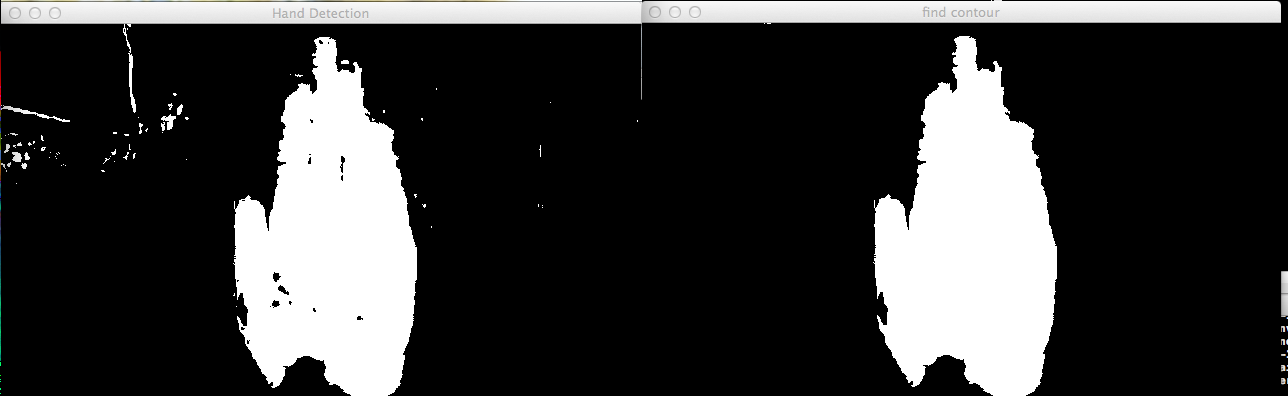
After performing the threshold function on the segmented HSV image we are left with a binary image that isolates the hand shown in *Figure 3.4*.



*Figure 3.4 Performing image binarization*

### 3) Noise Removal

To improve the code even further a method for noise removal was added. After performing image binarization, often times there were stray pixels in the frame that was identified. For the best accuracy in detecting hand gestures, we need to remove as many pixels as possible that is not on the hand. To do this we use cvFindContours()to identify all the bob. we then find the biggest blob as did earlier, afterwards we erase and replace the other smaller blobs with black pixels, erasing them. we are then left with a cleaner binary image as shown in *Figure 3.5*.



*Figure 3.5 Binary Image after noise removal*

## Recognizing Hand Gestures

Now that we have isolated the hand, we need to identify the sign language gestures by analyzing the characteristics of the shape of the hand.

### Morphological Characteristics

Now that we have isolated the hand in the image, we can perform some analysis. Simon needs a way to recognize and differentiate multiple hand gestures. Simon will recognize the sign language gestures of letters; B,C,F,I,L,R,V,Y.

We analyzed multiple static images of hands performing the different hand gestures. We find that the problem of recognizing the hand gestures can be scaled down to a shape identification problem. Reason being is that the different hand gestures have different shapes, in which there are essentially a unique combination of contour points, area, width, length, and skin pixel values. We generated a table that shows the characteristics of these hand gestures in *Figure 3.6*

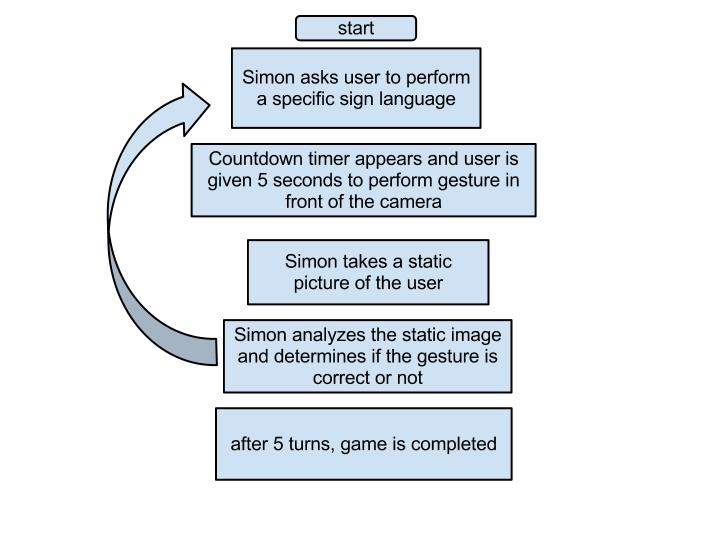
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Letter* | *Pixel Count* | *Width* | *Length* | *Area* | *Sides* |
| *B* | *56000-63000* | *190-230* | *440-470* | *60000-72000* | *4-8* |
| *C* | *48000-80000* | *280-470* | *370-440* | *51000-79000* | *5-10* |
| *D* | *37000-50000* | *220-370* | *440-470* | *50000-60000* | *6-11* |
| *F* | *54000-67000* | *220-260* | *440-460* | *63000-75000* | *6-9* |
| *I* | *49000-58000* | *240-310* | *410-445* | *56000-69000* | *6-9* |
| *L* | *39000-70000* | *380-550* | *430-475* | *40000-95000* | *5-10* |
| *R* | *40000-54000* | *20-100* | *20-70* | *200-3400* | *8-13* |
| *V* | *30000-57000* | *20-60* | *20-45* | *300-1200* | *6-15* |
| *W* | *54000-67000* | *20-70* | *20-40* | *200-9000* | *8-14* |

*Figure 3.6 Chart with hand gesture morphological characteristics*

In order to view the morphological characteristics, we use the included cvFindContours() function that finds all of the blobs in the image. After binarization, the hand is the largest blob, so we include a method to retrieve information on the largest contour. These characteristics are the amount of pixels in the blob, the width, the length, area, and number of sharp points in the blob.

## Implementing Game Logic

To make our system interactive and entertaining, we wrap this sign language recognition system into a game, called Simon. It is based on the popular game in which there is a person designated as Simon, that instructs players to perform different actions, but only if Simon precedes their sentence with “Simon says”. In this system, Simon is the computer and the user is the player. The logic of the game is displayed in *Figure 3.7*. Simon will take the picture of the user performing the gesture action if they are ready, they will have 5 seconds to do so, then Simon will analyze the image and determine if the user was correct or not. Scoring will be based on the users correct input and whether they performed the action only when Simon tells them to. The user will have 5 turns of the game.



*Figure 3.7 Simon System Logic Flow*

# Testing

We tested against 5 different people, of different skin tones in a white background and bright white lights. The Accuracy results are graphed in *Figure 4.1* [haven’t tested yet to fill in this portion]

[insert figure of graphed accuracy results]

# Conclusion

[Need to test before writing conclusion]

1. Rasiwasia, Nikhil, Color Space for Skin Detection – A Review, Fondazione Graphitech, University of Trento, Italy
2. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
3. I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
4. K. Elissa, “Title of paper if known,” unpublished.
5. R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.
6. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
7. M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.