Our program can detect six static gestures and one dynamic gesture. The six static gestures are a thumbs up, one finger, two fingers, three fingers, four fingers, and five fingers. The dynamic gesture is a wave.

Pre-processing consisted of making the gesture templates. First hand images were taken using the Photo Booth application on a Mac laptop. The image was cropped to only consist of the hand. Using cv2, the image was turned into gray scale then into a black and white image. To pre-process the camera image, the camera input was first transformed into HSV instead of RGB values. This allowed better skin detection through thresholding. The camera input was thresholded for skin color and only skin color objects were used. In this way, we only focused on the hand from the camera input. This requires using a dark background and ensuring that one's face is not in the image. Finally, the received skin image was turned into a black and white image for template matching.

The determine what hand gesture was presented, we used template matching through opencv. All templates were stored in a list of template and matched with the camera image. The template with the largest match was then singled out and thresholded to determine if the match was strong enough. If it was, a green box was drawn around the gesture and the label was indicated on the top left of the screen. If there was no match, the program asked whoever was using it to move their hand for better capture angle. For the waving gesture, we overlapped and rotated the five finger template which resembled the movement of motion in the eyes of the camera. We essentially used the non-perfect frame rate to our advantage as a measure of motion.

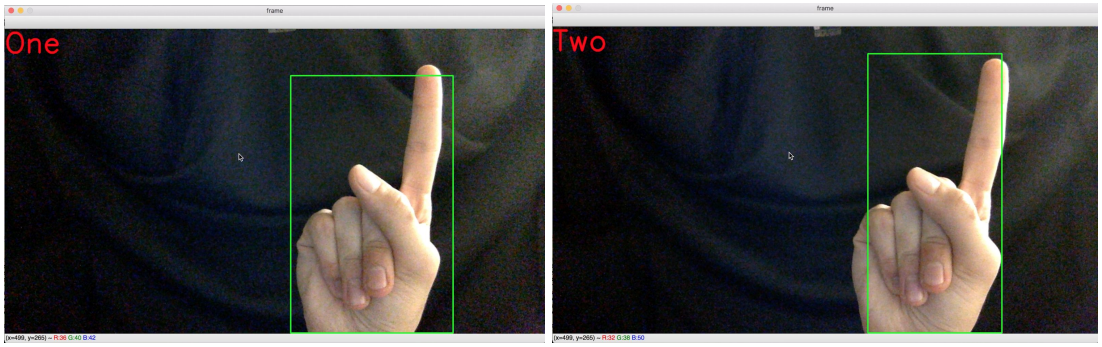
Qualitative results: For our program to work, we had to take into account hand size, skin dcoloretection, and background for success and failure cases. The hand size had to match the template and was not a major issue for our hands; however, problems could arise if the program had to analyze smaller or larger hands. Skin detection was done using HSV values, which is not perfect and could result in error when not done in the proper setting. For this reason, we used a black background in our analysis in order

to specifically separate the hand from the background. We also kept any other skin (arms, face, etc.) out of the frame when attempting to identify gestures.

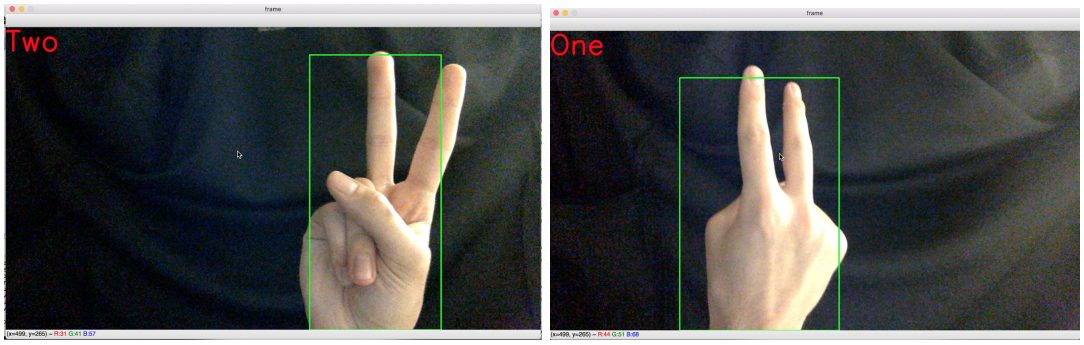
Observations we thought were interesting:

- The program's accuracy decreased when trying to identify different hand sizes.
- Motion can be identified through template matching at times due to the image blur during movement.

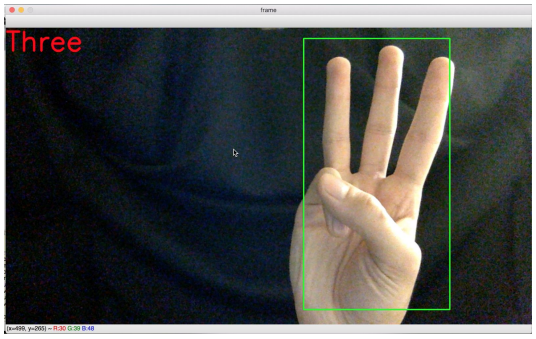
Left column is correct gesture recognition, right column is incorrect gesture recognition

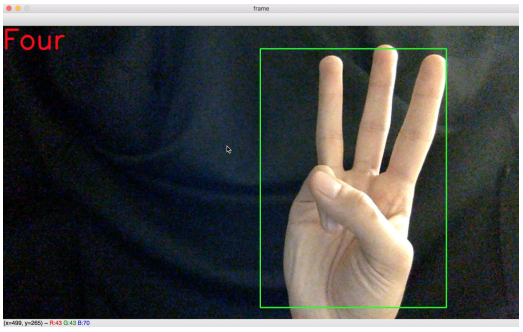


The program most likely failed in the right image because it occasionally read the thumb as a second finger. This only occured when the hand was first presented, then the program was able to correctly indicate one finger was raised.

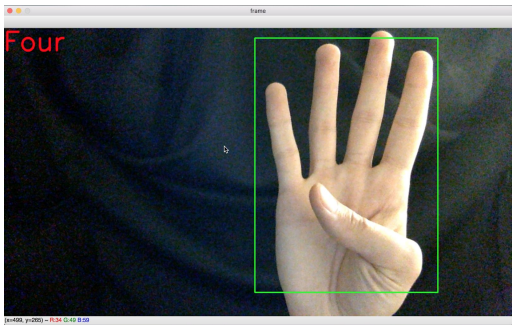


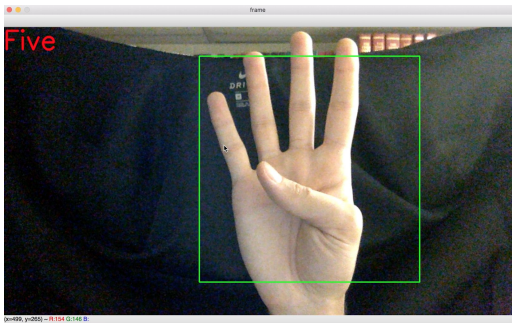
The program was not able to correctly label the image on the right because it does not have a template for hand raising two fingers with the palm facing away from the camera.



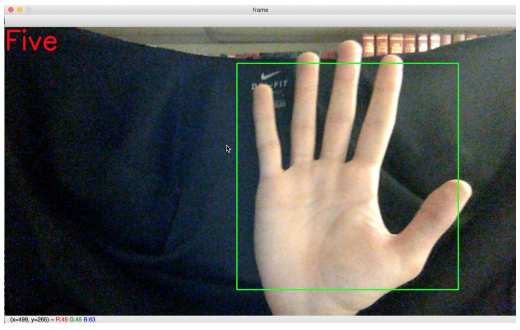


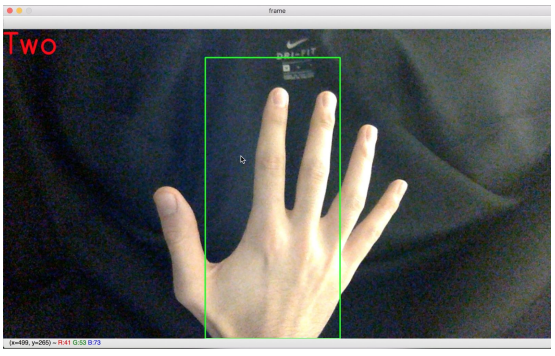
As with the fail case for one finger, the program added a finger here possibly due to the thumb protruding from the hand. This was only transiently observed as the program correctly indicated three fingers when given enough time to see the hand.



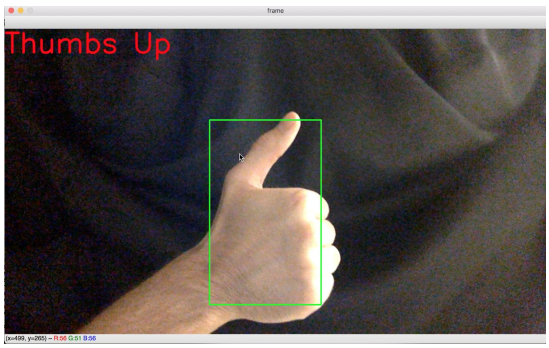


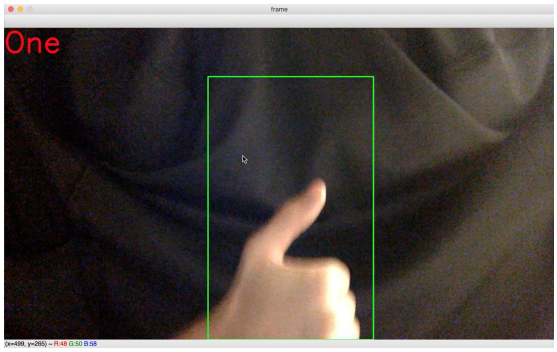
The image on the right was not recognized because the background was not completely black as there was a slight drape in the cloth. The increased noise made the program believe that five fingers were presented.





The image on the right was not recognized because the program only focussed on the two fingers indicated by the green box. Also, the template given was of a hand with the palm facing the camera as in the image on the left, so the image on the right was not accurately recognized by the program.





The program did not correctly indicate a thumbs up on the right because part of the hand was out of the frame of the camera. The program was tricked into labeling it as one because that was the closest template match.

The quantitative results were generally good for the limited amount of trials that were done for each static and motion gesture. There were 10 trials for each gesture with which a confusion matrix was created. Each gesture was presented for about 3 seconds before classifying it.

Classification Matrix

	One	Two	Three	Four	Five	Thumb	Wave
One	10						
Two	2	7			1		
Three			7	1	1	1	
Four				9	1		
Five					10		
Thumb						10	
Wave				1	3		6

Confusion Matrices for each gesture

Gesture: One

	P	N	Recall: 0.8333
P	10		Precision: 1.0
N	2	0	F1: 0.9091

Gesture: Two

	P	N	Recall: 1.0
P	7	3	Precision: 0.7
N	0	0	F1: 0.8235

Gesture: Three

	P	N	Recall: 1.0
P	7	3	Precision: 0.7
N	0	0	F1: 0.8235

Gesture: Four

	P	N	Recall: 0.8182
P	9	1	Precision: 0.9
N	2	0	F1: 0.8571

Gesture: Five

	P	N	Recall: 0.6250
P	10	0	Precision: 1.0
N	6	0	F1: 0.7692

Gesture: Thumb

	P	N	Recall: 0.9091
P	10	0	Precision: 1.0
N	1	0	F1: 0.9524

Gesture: Wave

	P	N	Recall:1.0
P	6	4	Precision: 0.6
N	0	0	F1: 0.75

In general the F1 score was relatively good, but with more trials it will most likely average out to a lower amount. Consequently, the system works well as long as there is about a 3-5 second delay for classification for the system to calibrate. Combined with our qualitative results, it can be seen that our system does consistently work with a solid black background. However, due to the HSV values themselves not properly accounting for the background, the system is not capable of differentiating in high luminance environments. Therefore, for the future we would like to add additional capabilities to account for these issues.