PA2 Report

By: Chris-Emio Raymond
Teammate: Zhan Hao Xu, Ivan Mata
Professor Betke

Problem:

Our program can detect 3 different static hand gestures:palm, thumb, and rock. One dynamic hand gesture: wave.

Process:

When we first start up the program it immediately runs background removal on our camera feed and it displays that side by side with our colored camera feed. We now have two windows open side by side. One that's in color and is just a regular camera feed, and another in black and white where the white aspects represent our skin (Refer to the colored feed as feed 1 and the B&W feed as feed 2 from now on).We then upload our hand gesture templates into the program. The templates are in black and white as well, not in color. We then compare the frames captured by feed 2 to our templates and see if any hand gestures are detected. If a gesture is detected then a green box will surround the gesture indicating that it has identified it as one of the gestures presented in the templates. It then displays the name of the gesture, but if no gesture is recognized then the program will display, "None". The way the comparison is done is that we run the function matchTemplate on our feed 2 frame and on each hand template. So the program is constantly executing matchTemplate on all 3 hand templates to check to see if there's a match.

Now for frame differencing, the way we went about this is by having a variable named frame equal the current frame at the beginning of our while(True) loop. We then have a variable named old_frame equal the current frame at the end of the loop. What this does is that when we go back through the loop, frame is now the current frame that the camera is capturing and old_frame is the frame that the camera had captured previously. Now to find the frame differencing all we do is find the absolute difference between frame and old_frame, convert it to grayscale, and run the function threshold on it. What this does is that if the difference between one frame to the other was big enough we set the pixel value to 255 (white) and if it wasn't we set the value to 0 (black). We then display this new frame on a separate window next to our other two windows.

For the dynamic hand gesture detection we used the matrix provided by our frame differencing frame. We counted up all the white pixels in the matrix and if that number exceeded 5000 then we stated that a wave was detected.

Statistics:

Confusion Matrix:

N = 25

Fist		Predictions		
	Fist	1	0	sum
Actual	1	8 (0.8)	2 (0.2)	10
	0	6 (0.4)	9 (0.6)	15

N = 25

Thumb		Predictions		
	Fist	1	0	sum
Actual	1	8 (0.571)	6 (0.429)	14
	0	3 (0.273)	8 (0.727)	11

N = 25

Palm		Predictions		
	Fist	1	0	sum
Actual	1	6 (0.429)	8 (0.571)	14
	0	0	1	11

N = 25

Wave		Predictions		
	Fist	1	0	sum
	1	8 (0.471)	9 (0.529)	17

Actual	0	1 (0.125)	7 (0.875)	8

Fist:

Precision = TP / (TP +FP) = 8 / (8+ 6) = 8/14 = 0.571

recall = TPR = 8/10 = 0.8

F1-Score = 2PR/(P+R) = (2*0.571*0.8)/(0.571+0.8) = 0.9136 / 1.371 = 0.67

Thumb:

Precision = TP / (TP +FP) = 8 / (8+ 3) = 8/11 = 0.727

recall = TPR = 8/ 14 = 0.571

F1-Score = 2PR/(P+R) = (2*0.727*0.571)/(0.727+0.571) = 0.8302 / 1.298 = 0.64

Palm:

Precision = TP / (TP + FP) = 6 / (6 + 0) = 6 / 6 = 1

Recall = TPR = 6 / 14 = 0.429

F1-Score = 2PR/(P+R) = (2*1*0.429)/(1+0.429) = 0.6

Wave:

Precision = TP / (TP + FP) = 8 / (8 + 1) = 8 / 9 = 0.889

Recall = TPR = 8 / 17 = 0.471

F1-Score = 2PR/(P+R) = (2*0.889*0.471)/(0.889+0.471) = 0.616

Using the templates, we can pick up our gestures that show up in the camera. The reason why fist pops up very often is because the template that we use for the fist is basically a white blob it is very ambiguous to the machine, The camera will pick up any white blob that is a similar shape to the template. The thumbs up is a white blob and a finger pointing up and sometimes the machine will confuse the 2 as seen in the picture below.