Jen Johnson CSCI 454 HW3

In HW2, PCA was implemented for face matching on non-aligned images. Data was normalized by subtracting the mean of each image. The first 10 eigenvectors were used to calculate the weights for each image. The distribution of genuine and imposter scores overlapped greatly, so there was not a very good threshold to set to distinguish distance scores. The equal error rate was 0.41, which means that even at the best threshold, the false accept and false reject rates were wrong 41% of the time.

In Part 1, the first 30 eigenvectors were used instead of the first 10. A larger number of eigenvectors means that more information is retained, so the comparisons between images had improved accuracy. However, the plots were still very similar in that the imposter and genuine scores overlapped. However, the imposter curve had shifted to the right a little bit. Both curves shared the same shape, as the same two points on the left hand side of the genuine curve are seen in both plots. The equal error rate decreased to 0.38, but this was not very much improvement.

In Part 2, we performed another step to normalize the data. This extra normalization reduced range of lighting in the entire dataset. This in turn allowed the variation detected by PCA to address more directly differences in the faces rather than differences in the image due to lighting. The shape of the score distribution curve changed, since this extra step was more significant than just using a larger number of eigenvectors. The overlap of the imposter and genuine curves decreased. The equal error rate decreased to 0.36. This was not a very large numerical change, but the placement of the curves suggests that this step is essential.

In Part 3, another preprocessing step was used to normalize the data. An external program rotated the images by detecting the eyes in the image. This decreased the noise caused by variations in pose or when a face was not centered in the frame. The equal error rate was 0.37. This was higher than the previous result, even though min-max scale was also used. However, it was still considered an improvement because the overlap of the imposter and genuine scores decreased. In the third plot, the most frequent scores for imposter and genuine occur in different positions along the x axis. The most frequent genuine score was around 48, while the most frequent imposter score was around 60.

In Part 4, local region PCA was used to increase the quality of comparisons between images. This decreased the noise in the data even further, as it reduced the variation caused by pose even more than in Part 3. The weighting of different sectors addressed the fact that the middle of the face is more useful for distinguishing differences than the edges of the face/image. The equal error rate decreased to 0.29. This magnitude of change in equal error rate is greater than in previous iterations, which suggests that this was the best method to improve the accuracy of PCA. However, even small improvements from each variation contributed to the final accuracy. Compared to HW2, the variations implemented in this homework were effective, since an error rate of 29% is much better than the initial 41%.

I have neither given nor received unauthorized aid on this assignment. Jen Johnson