Assignment 2

Baochen Wang

◆ Questions

Question-1: After you apply PCA, what is the dimension of the transformed (1D) image vectors?

I used princomp(X) performs principal components analysis (PCA) on the n-by-p data matrix X. The returned COEFF is a p-by-p matrix, each column containing coefficients for one principal component. The columns are in order of decreasing component variance. The input matrix X is 10304x400. The covariance matrix (COEFF) is 400x400.

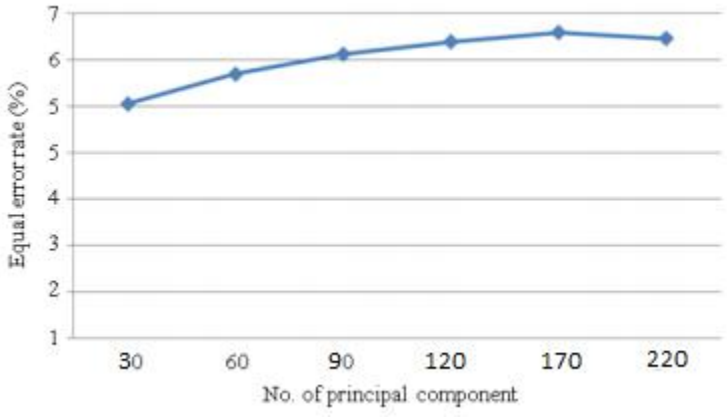
Question-2: The ORL database has 40 users and there are 10 samples for each user. Using 6 of them, create templates for each user. Use the remaining 4 for testing (i.e., creating genuine and impostor distributions). How you create the template?

In feature extraction, a mathematical representation of original image called a biometric template. For each user, computed as the mean of the projected faces. To use the 6 images transformed using PCA and calculate the mean image.

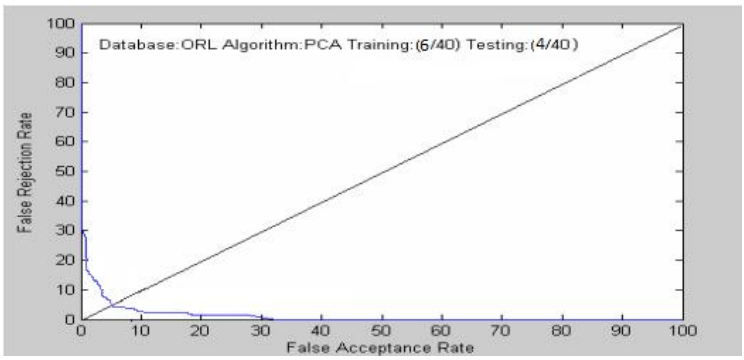
Question-3: Using first k components of the transformed feature vectors, determine the EER value, EER(k). For different values of k, draw the plot EER vs. k for k=1, 5, 10, 15, … max possible k. Which value of k is the best? For the best value of k, plot the genuine and impostor distributions and corresponding ROC curve.

In this case, the ORL database has 40 users and there are 10 samples for each user. Using 6 of them, create templates for each user(training set). Then using the remaining 4 samples to test. So, 4 test data for every user is used to generate 40x4=160 genuine authentication attempts and 39x40x4=6240 impostor authentication attempts (4 attempts by 39 remaining users for every user in the system).

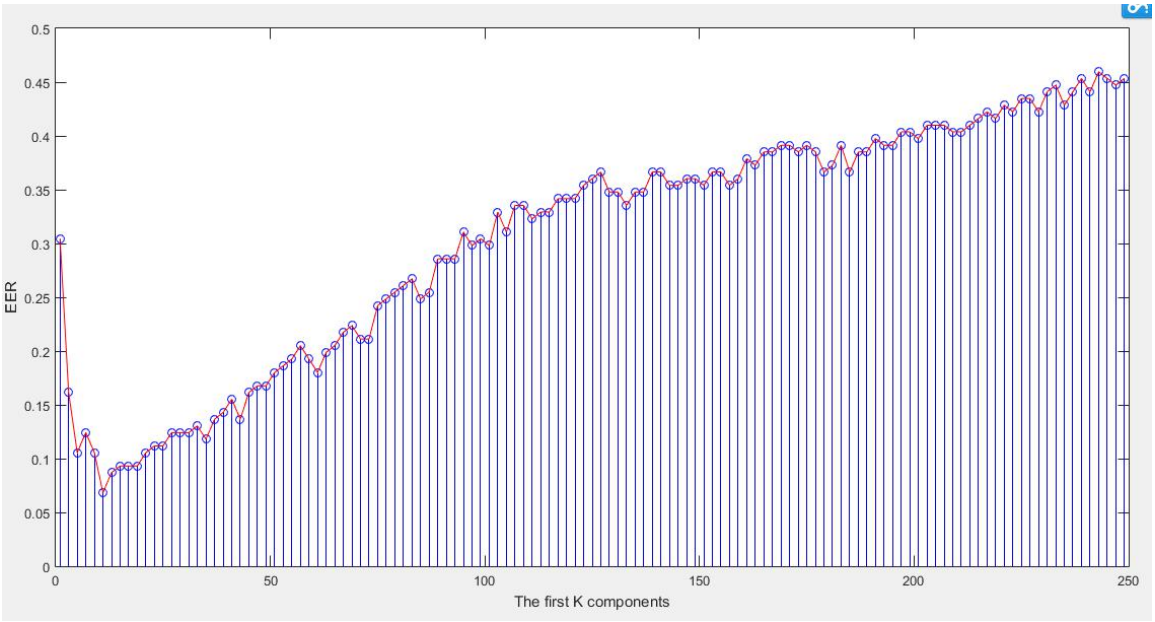
Through the research:



Best EER is 0.05 when k =30.

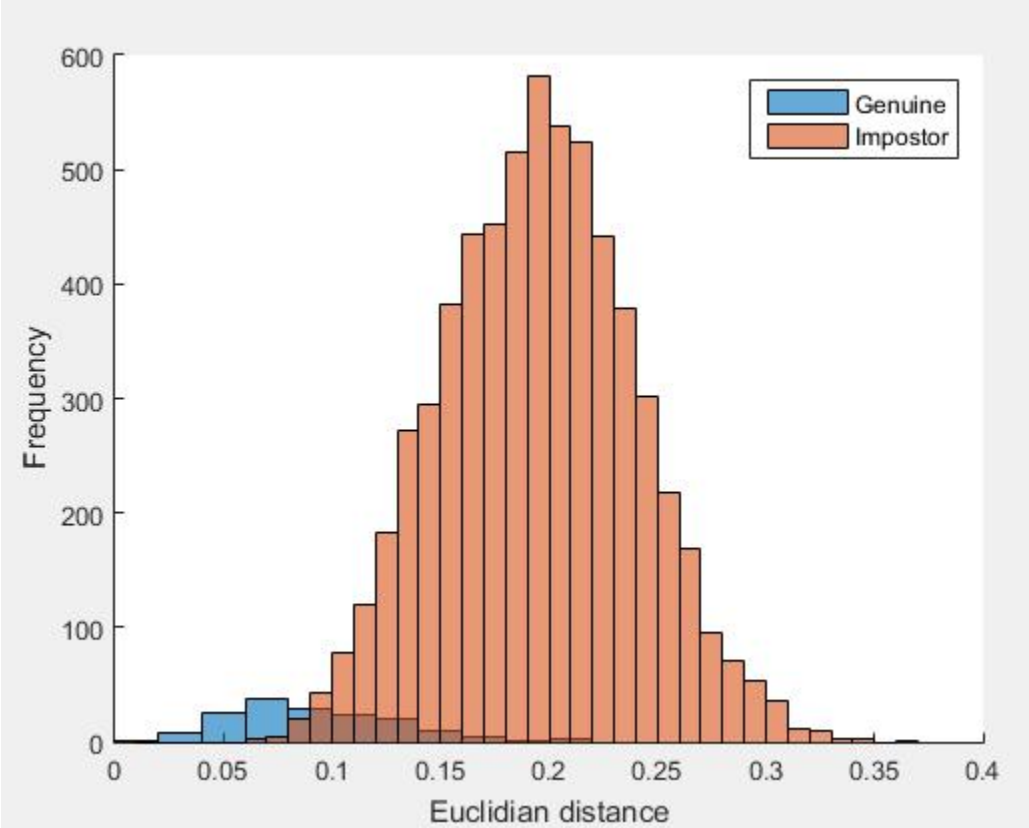


Through my program:

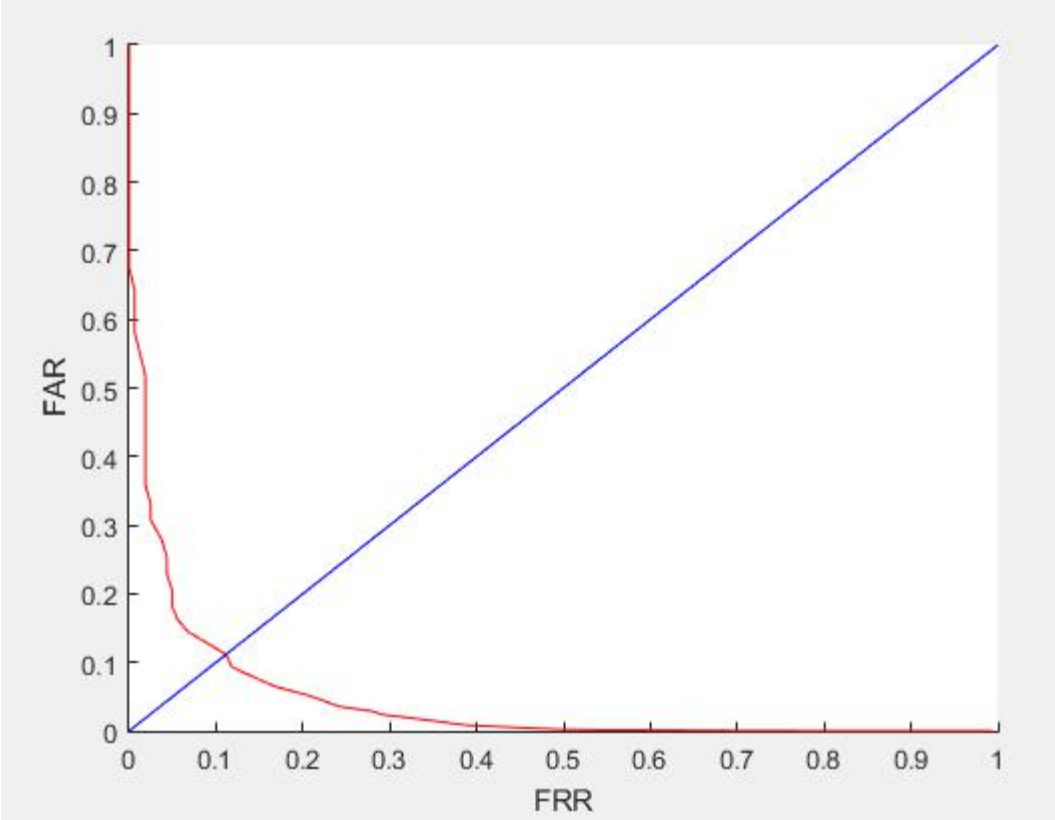


Best EER is 0.068 when k = 10.

The genuine and impostor distributions:



The corresponding ROC curve:

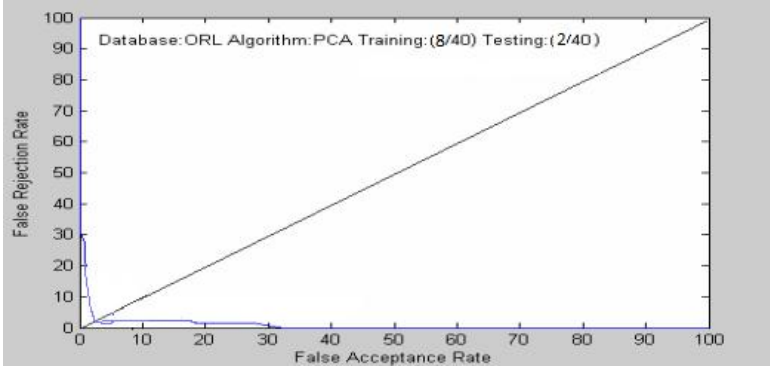


Question-4: For the best value of k that you found in Q3, plot the genuine and impostor distributions and corresponding ROC curve again. But this time, use 8 feature vectors (for each user) to create the template and use the remaining 2 for testing. What is the value of EER this time? Is it better/worse? Why?

In this case, the ORL database has 40 users and there are 10 samples for each user. Using 8 of them, create templates for each user(training set). Then using the remaining 2 samples to test. So, 2 test data for every user is used to generate 40x2=80 genuine authentication attempts and 39x40x2=3120 impostor authentication attempts (2 attempts by 39 remaining users for every user in the system).

Through the research:

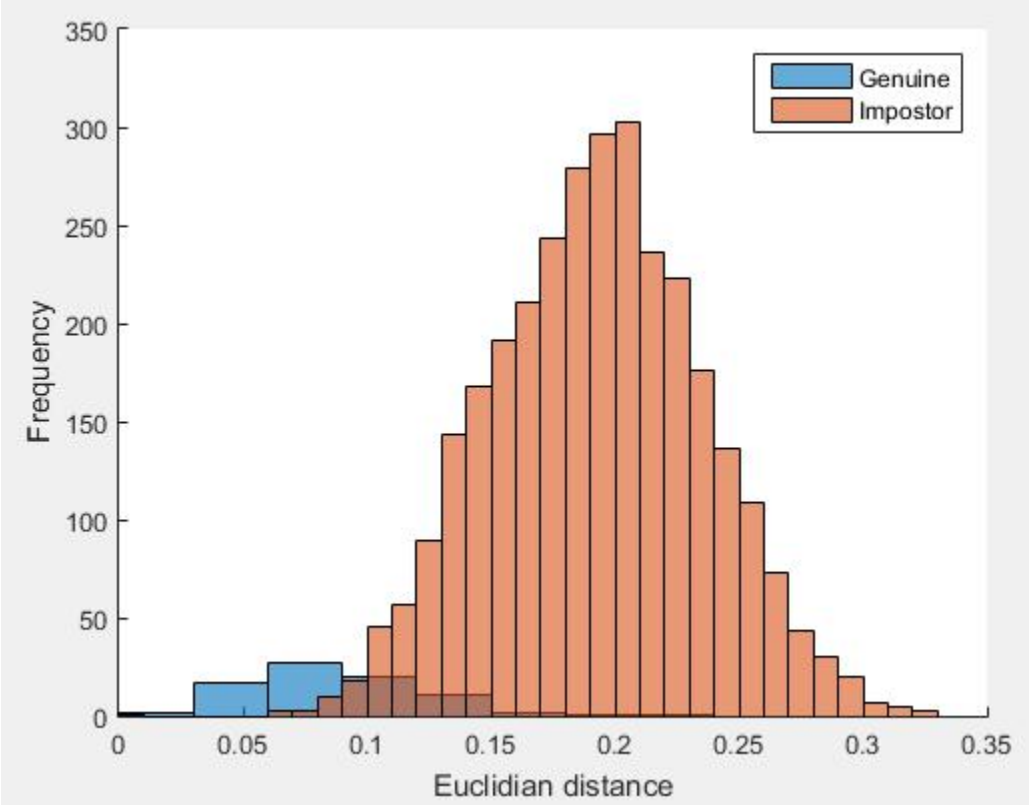
The EER value becomes smaller(around 1%), so it is better. More training images produce better EER.



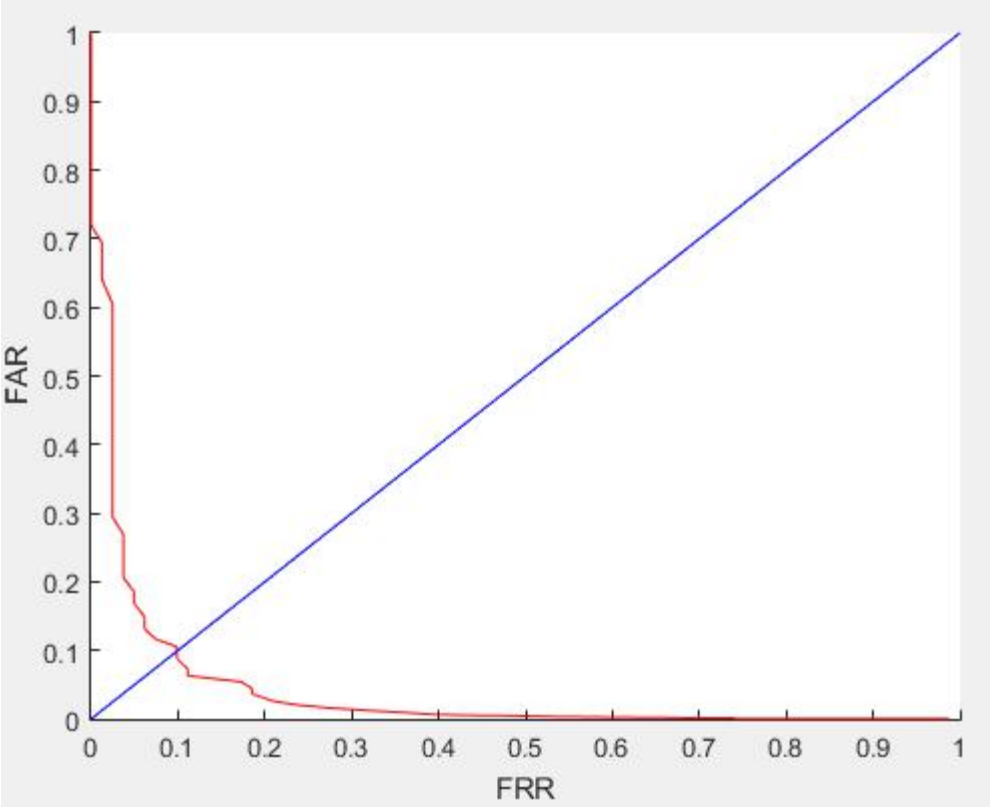
Through my program:

The EER value becomes less than 0.068(there is error in ROC plot), so it is better. More training images produce better EER.

The genuine and impostor distributions:



The corresponding ROC:



Question-5: What other (than Euclidian distance) possible measure(s) you may use to compare two feature vectors?

City Block distance. The sum of absolute differences between two vectors is called the L1 distance, or city-block distance. This is a true distance function since it obeys the triangle inequality. reason why it is called the city-block distance, and also as the Manhattan distance or taxicab distance is that going from a point A to a point B is achieved by walking 'around the block', compared to the Euclidean 'straight line' distance.

Mahalanobis distance. The Mahalanobis distance is a measure of the distance between a point P and a distribution D. It is a multi-dimensional generalization of the idea of measuring how many standard deviations away P is from the mean of D. This distance is zero if P is at the mean of D, and grows as P moves away from the mean: along each principal component axis, it measures the number of standard deviations from P to the mean of D.

Range-based measure. It is more sensitive to feature variation since that requires each feature to be in pre-estimated ranges.

Reference

1. https://en.wikipedia.org/wiki/Mahalanobis_distance

2. http://ijaiem.org/volume2issue7/IJAIEM-2013-07-24-087.pdf

3. http://www.biometrics.org/bc2013/presentations/iris_sutcu_tuesday_1540.pdf

4. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.177.2547&rep=rep1&type=pdf