ENCM 509 Lab 6

March 1, 2019

Group Members: Andrew Schneider, Patrick Settle

Instructor: Dr. Svetlana Yanushkevich

Introduction

The focus of this lab is to investigate the basic eigenfaces approach for facial recognition. First, a training step is used to create eigenfaces for images in the database. Then the feature vector or eigenvector of the probe image is compared against the feature vectors of the database to determine it's difference (or euclidean distance) from each of the database images. In the first exercise, two users are considered to be in the database, and a threshold on euclidean distance is determined that separates users who are in the database and those who are not. A third user who is not in the database is used to help choose a second threshold that separates images of faces and images that are not faces.

In the second exercise, synthetic images of a genuine user are created using FaceGen, and then compared with the existing database of real images to see if they match or are rejected.

Exercise 1

Evaluation of Genuine User Threshold

The first threshold we need for a facial recognition system is a score to define what a genuine or accepted user is. To evaluate this threshold we did k-fold training of our 9 training images (9-fold training, can be seen in figure 1). Each image or each genuine user was matched against all of their other images, and the maximum and minimum scores were recorded. The resulting scores are recorded in table 1.

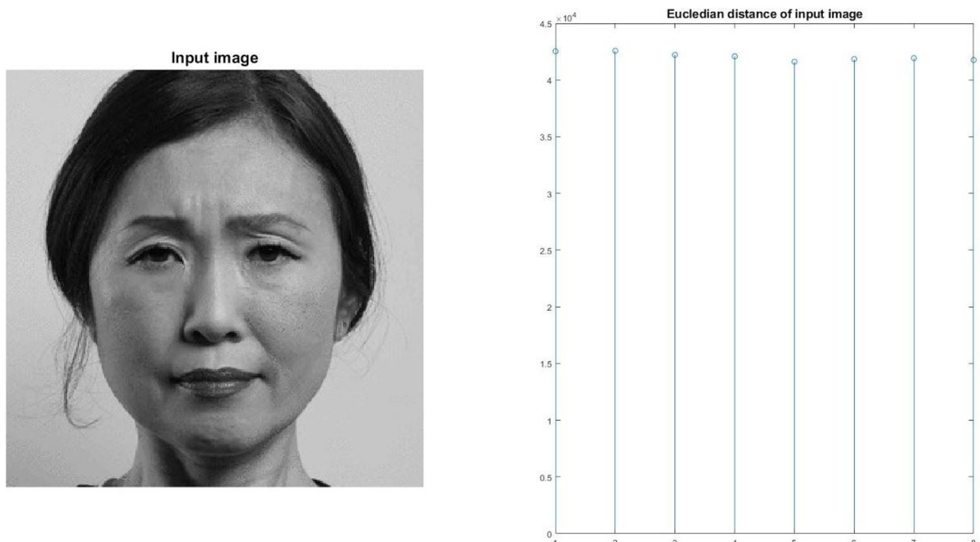


Figure 1: A user matched against 8 of their own images.

Table 1: Results of k-fold training, the minimum and maximum score of each probe.

User Type	Minimum Score	Maximum Score
Genuine, Class A	41614.84682	42572.81695
Genuine, Class A	41442.1433	43982.14314
Genuine, Class A	41626.86362	44259.2016
Genuine, Class A	41433.11021	42968.22834
Genuine, Class A	41703.63173	42948.96211
Genuine, Class A	41413.76256	42324.32261
Genuine, Class A	41541.77709	42755.1408
Genuine, Class A	41893.50305	43170.20109
Genuine, Class A	42142.78278	43727.41279
Genuine, Class B	44099.34829	44659.65729
Genuine, Class B	43678.77711	44231.08892
Genuine, Class B	43756.50389	44428.8852
Genuine, Class B	43825.97739	44467.80394
Genuine, Class B	43992.53078	44383.55979
Genuine, Class B	44123.29767	44677.73688
Genuine, Class B	43730.87721	43883.81529
Genuine, Class B	43802.30781	44389.55547
Genuine, Class B	43893.26222	44449.04421

Using these minimum and maximum scores, we can calculate:

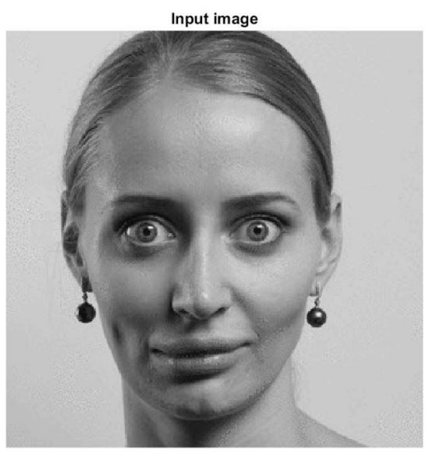
$\mu_{min} = 42761.96$
$\mu_{max} = 43793.31$
$\sigma_{min} = 1165.56$
$\sigma_{max} = 785.52$

When we go to do matching, we will be using either a rank-1 or a rank-3 approach, that both use the minimum scores. Therefore we will select a threshold by selecting an error range on μ_{min} at 97.5% confidence:

$$thresh_1 = \mu_{min} + \varepsilon_{0.025} = 42761.96 + \frac{z_{0.025} \cdot \sigma_{min}}{\sqrt{N}} = 43300.41$$

Evaluation of Personhood Threshold

The second threshold we need for a facial recognition system is a score to evaluate if a user is an image of a person at all. To evaluate this threshold, we create a database of all our genuine users, then analyze the scores that imposters receive when compared against the database. An example of this comparison can be seen in figure 2. The scores of these comparisons can be seen in table 2.



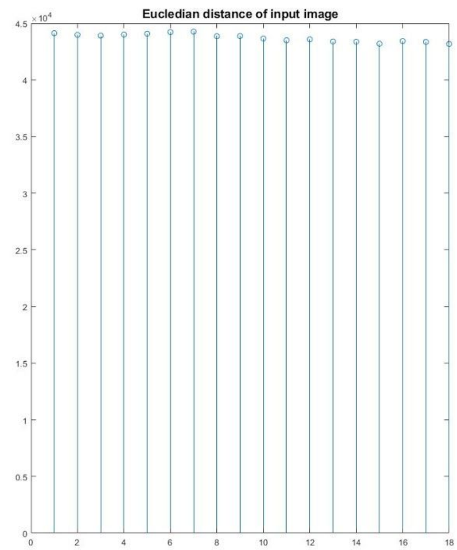


Figure 2: Imposter user compared against all genuine users

Table 2: Results of imposter training, the minimum and maximum score of each probe.

User Type	Minimum Score	Maximum Score
Imposter, Class C	43125.88735	43890.49512
Imposter, Class C	43259.07402	44142.64536
Imposter, Class C	43155.01656	43880.58068
Imposter, Class C	43313.59036	44128.20945
Imposter, Class C	43278.24585	44044.30234
Imposter, Class C	43494.80112	44417.62189
Imposter, Class C	43359.99998	44274.1526
Imposter, Class C	43763.40873	45004.34991
Imposter, Class C	43179.37745	44258.88745

Using these minimum and maximum scores, we can calculate:

$\mu_{min} = 43325.49$
$\mu_{max} = 44226.80$
$\sigma_{min} = 199.58$
$\sigma_{max} = 340.47$

When we go to do matching, we will be using either a rank-1 or a rank-3 approach, that both use the minimum scores. Therefore we will select a threshold by selecting an error range on μ_{min} at 97.5% confidence:

$$thresh_2 = \mu_{min} + \varepsilon_{0.025} = 43325.49 + \frac{z_{0.025} \cdot \sigma_{min}}{\sqrt{N}} = 43455.88$$

Facial Recognition with Rank-1 Matching

To do rank-1 matching, a probe image is matched against every image in the database. Each match gives a distance score, and the minimum distance score is compared against the designed thresholds to produce a recognition result. Figure 3 shows one of these matches with a genuine user. The results of this matching approach are described in table 3.

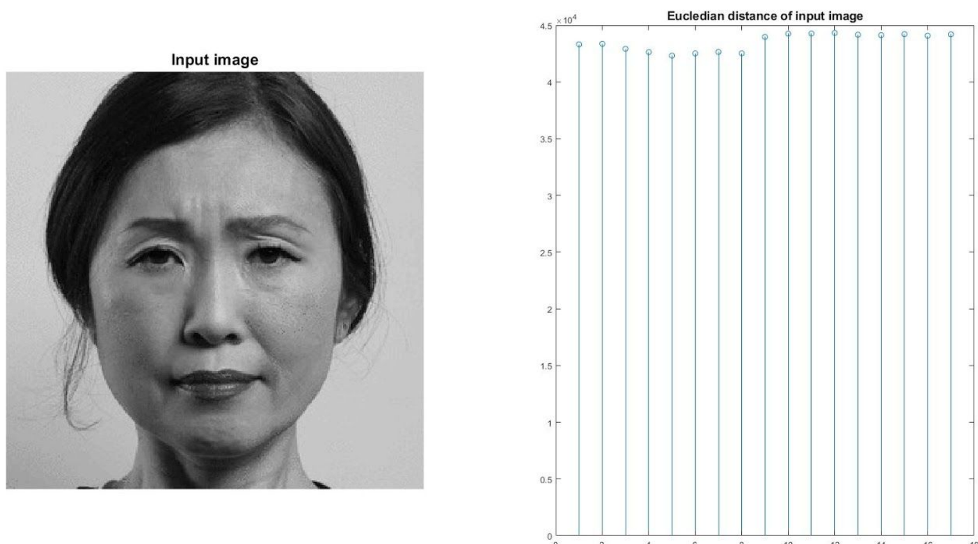


Figure 3: Genuine user compared against the database of genuine users

Table 3: Results of facial recognition, rank-1 approach.

User Type	Minimum Matching Score	Decision Result
Genuine, Class A	42323.85899	Known Face (Class A)
Genuine, Class A	42267.01325	Known Face (Class A)
Genuine, Class A	42324.02428	Known Face (Class A)
Genuine, Class A	42379.87384	Known Face (Class A)
Genuine, Class A	42518.72973	Known Face (Class A)
Genuine, Class A	42331.61676	Known Face (Class A)
Genuine, Class A	42323.82842	Known Face (Class A)
Genuine, Class A	42605.87809	Known Face (Class A)
Genuine, Class A	42816.27755	Known Face (Class A)
Genuine, Class B	42384.99657	Known Face (Class B)
Genuine, Class B	42156.69688	Known Face (Class B)
Genuine, Class B	42241.99508	Known Face (Class B)
Genuine, Class B	42329.49298	Known Face (Class B)
Genuine, Class B	42381.64709	Known Face (Class B)
Genuine, Class B	42517.0049	Known Face (Class B)
Genuine, Class B	42195.45098	Known Face (Class B)
Genuine, Class B	42317.69715	Known Face (Class B)
Genuine, Class B	42487.44294	Known Face (Class B)
Imposter, Class C	43125.88735	Known Face (Class B)
Imposter, Class C	43259.07402	Known Face (Class B)
Imposter, Class C	43155.01656	Known Face (Class B)
Imposter, Class C	43313.59036	Unknown Face
Imposter, Class C	43278.24585	Known Face (Class B)
Imposter, Class C	43494.80112	Not a Face
Imposter, Class C	43359.99998	Unknown Face
Imposter, Class C	43763.40873	Not a Face
Imposter, Class C	43179.37745	Known Face (Class B)

Assuming that “Not a Face” results can be considered true negatives:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{18+2}{18+2+5+2} = 74.1\%$$
$$Sensitivity = \frac{TP}{TP+FN} = \frac{18}{18+2} = 90\%$$
$$FAR = \frac{FP}{TN+FP} = \frac{5}{2+5} = 71.4\%$$
$$FRR = \frac{FN}{TP+FN} = \frac{2}{18+2} = 10\%$$

The results of this matching technique were 74.1% accuracy with a FAR of 71.4% and a FRR of 10%. Looking at the results of the matching scores, it appears that the thresholds could have been selected to improve the accuracy of the matching with our dataset. Since all the genuine users have minimum matching scores below 43000 and all imposter users have matching scores above 43000, setting $thresh_1 = 43000$ and $thresh_2 = 100000$ would have resulted in a perfect matching result

Facial Recognition with Rank-3 Matching

To do rank-3 matching, a probe image is matched against every image in the database. Each match gives a distance score, and the minimum distance score is compared against the designed thresholds to produce a recognition result. A result is generated for the second and third smallest distance scores. The recognition category with the most results is taken as the final result. In the event of a tie, the overall smallest distance score is used as a tiebreaker. The results of this matching procedure can be seen in table 4.

Table 4: Results of facial recognition, rank-3 approach

User Type	Result 1	Result 2	Result 3	Final Result
Genuine, Class A	Known Face	Known Face	Known Face	Known Face (Class A)
Genuine, Class A	Known Face	Known Face	Not a Face	Known Face (Class A)
Genuine, Class A	Known Face	Known Face	Not a Face	Known Face (Class A)
Genuine, Class A	Known Face	Known Face	Known Face	Known Face (Class A)
Genuine, Class A	Known Face	Known Face	Known Face	Known Face (Class A)
Genuine, Class A	Known Face	Known Face	Known Face	Known Face (Class A)
Genuine, Class A	Known Face	Known Face	Known Face	Known Face (Class A)
Genuine, Class A	Known Face	Known Face	Known Face	Known Face (Class A)
Genuine, Class A	Known Face	Known Face	Known Face	Known Face (Class A)
Genuine, Class B	Known Face	Known Face	Known Face	Known Face (Class B)
Genuine, Class B	Known Face	Known Face	Known Face	Known Face (Class B)
Genuine, Class B	Known Face	Known Face	Known Face	Known Face (Class B)
Genuine, Class B	Known Face	Known Face	Known Face	Known Face (Class B)
Genuine, Class B	Known Face	Known Face	Known Face	Known Face (Class B)
Genuine, Class B	Known Face	Known Face	Known Face	Known Face (Class B)
Genuine, Class B	Known Face	Known Face	Known Face	Known Face (Class B)
Genuine, Class B	Known Face	Known Face	Known Face	Known Face (Class B)
Genuine, Class B	Known Face	Known Face	Known Face	Known Face (Class B)
Imposter, Class C	Known Face	Unknown Face	Unknown Face	Unknown Face
Imposter, Class C	Known Face	Unknown Face	Unknown Face	Unknown Face
Imposter, Class C	Known Face	Known Face	Known Face	Known Face (Class B)
Imposter, Class C	Unknown Face	Unknown Face	Not a Face	Unknown Face
Imposter, Class C	Known Face	Known Face	Unknown Face	Known Face (Class B)
Imposter, Class C	Not a Face	Not a Face	Not a Face	Not a Face
Imposter, Class C	Unknown Face	Not a Face	Not a Face	Not a Face
Imposter, Class C	Not a Face	Not a Face	Not a Face	Not a Face
Imposter, Class C	Known Face	Known Face	Unknown Face	Known Face (Class B)

Assuming that “Not a Face” results can be considered false negatives:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{18+3}{18+3+3+3} = 77.7\%$$
$$Sensitivity = \frac{TP}{TP+FN} = \frac{18}{18+3} = 85.7\%$$
$$FAR = \frac{FP}{TN+FP} = \frac{3}{3+3} = 50.0\%$$
$$FRR = \frac{FN}{TP+FN} = \frac{3}{18+3} = 14.3\%$$

The results of this matching technique were 77.7% accuracy with a FAR of 50.0% and a FRR of 14.3%. The rank-3 approach appears to increase the accuracy of the test slightly and also reduce the FAR somewhat. However, the rank-3 approach does also increase the number of imposter users that are incorrectly identified as “Not a Face”, which is not accounted for in the test statistics.

Effectiveness of Procedure

In general, the facial recognition system used for this lab is not very effective. The main issue with the procedure is that the scores that result from K-fold training within a class appear to be on a different scale than the scores obtained when comparing a probe against a collection of classes. For example, these are the threshold scores obtained by following the same process as above, except by doing 18-fold training across 2 classes. (In a similar manner to the recognition procedure used to test the thresholds).

$$thresh_1 = 42456.54$$
$$thresh_2 = 43455.88$$

The first threshold is noticeably different. When comparing the first threshold against table 3, it is easy to see that these new thresholds will be much more effective at reducing the FAR.

Another possible modification to this procedure would be to treat each class independently. In this case we would have two threshold values for each class of genuine user. This would accommodate variations between genuine classes quite easily.

Exercise 2

Face Modeling and Recognition using Synthetic Faces.

For this exercise, we used an image of a neutral expression from one of our classes in the database from exercise 1, and created a 3D model with FaceGen. Then, the model was modified to create 21 different synthetic pictures:

- Original synthetic image
- 15, 30 and 45 degree rotation
- 4 facial expressions (angry, smiling, disgust, surprise)
- Aged to 30,40,50,60 years
- Neutral face with 5 different backgrounds (black, red, green, blue, and the ubiquitous image of Lena Söderberg)
- Neutral face with 4 different lighting conditions (Ambient brightness 0, ambient brightness 1.0, lighting preset 2, lighting preset 3)

Next, as an intermediate step to try and make the facial recognition algorithm as successful as possible, the images are resized and cropped to be sized and aligned well to the real images. For comparison, the neutral expression for the real and synthetic version of a face is shown below:

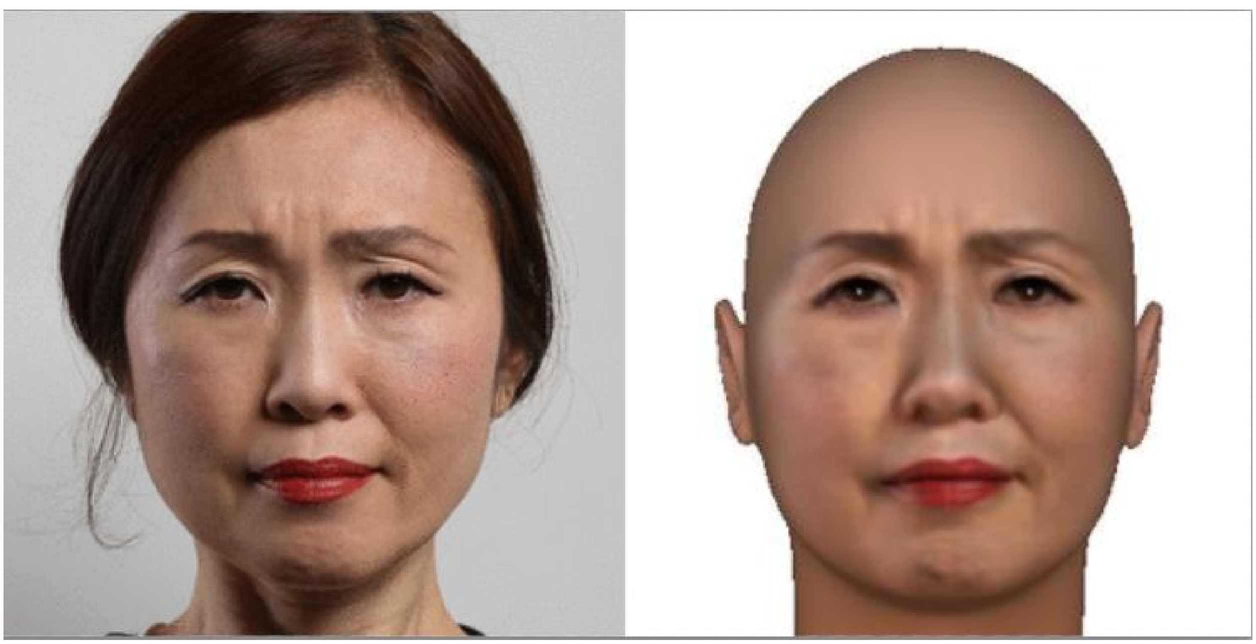


Figure 4: Real and synthetic versions of a genuine user

For each synthetic picture, we perform facial recognition using the database of genuine users from exercise 1, under the assumption that the synthetic face should be considered as a genuine face. I.e., If the synthetic face is considered part of the database, that is a true positive. Both the rank-1 and rank-3 approaches were used.

Table 5: Results of synthetic facial recognition, Rank-1 approach

Synthetic image description	Minimum Matching Score	Decision Result
Normal	44805.44371	Not a Face
15° rotation	44827.57178	Not a Face
30° rotation	44916.01677	Not a Face
45° rotation	44848.47976	Not a Face
Angry	44808.29987	Not a Face
Smile	44785.44278	Not a Face
Disgust	44800.21756	Not a Face
Surprise	44821.1653	Not a Face
Age 30	44813.80698	Not a Face
Age 40	44818.25436	Not a Face
Age 50	44809.0754	Not a Face
Age 60	44781.32306	Not a Face
Black background	48932.11334	Not a Face
Red background	47908.90875	Not a Face
Green background	45023.17278	Not a Face
Blue background	48753.19906	Not a Face
Lena background	46404.55824	Not a Face
Ambient Brightness 0	44946.63317	Not a Face
Ambient brightness 1.0	44848.17717	Not a Face
Lighting preset 2	44493.37413	Not a Face
Lighting preset 3	44721.15122	Not a Face

$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{0+0}{0+0+0+21} = 0\%$
$Sensitivity = \frac{TP}{TP+FN} = \frac{0}{0+21} = 0\%$
$FRR = \frac{FN}{TP+FN} = \frac{21}{0+21} = 100\%$

Table 6: Results of imposter training, the minimum and maximum score of each probe.

Synthetic image description	Result 1	Result 2	Result 3	Final Result
Normal	Not a Face	Not a Face	Not a Face	Not a Face
15° rotation	Not a Face	Not a Face	Not a Face	Not a Face
30° rotation	Not a Face	Not a Face	Not a Face	Not a Face
45° rotation	Not a Face	Not a Face	Not a Face	Not a Face
Angry	Not a Face	Not a Face	Not a Face	Not a Face
Smile	Not a Face	Not a Face	Not a Face	Not a Face
Disgust	Not a Face	Not a Face	Not a Face	Not a Face
Surprise	Not a Face	Not a Face	Not a Face	Not a Face
Age 30	Not a Face	Not a Face	Not a Face	Not a Face
Age 40	Not a Face	Not a Face	Not a Face	Not a Face
Age 50	Not a Face	Not a Face	Not a Face	Not a Face
Age 60	Not a Face	Not a Face	Not a Face	Not a Face
Black background	Not a Face	Not a Face	Not a Face	Not a Face
Red background	Not a Face	Not a Face	Not a Face	Not a Face
Green background	Not a Face	Not a Face	Not a Face	Not a Face
Blue background	Not a Face	Not a Face	Not a Face	Not a Face
Lena background	Not a Face	Not a Face	Not a Face	Not a Face
Ambient Brightness 0	Not a Face	Not a Face	Not a Face	Not a Face
Ambient brightness 1.0	Not a Face	Not a Face	Not a Face	Not a Face
Lighting preset 2	Not a Face	Not a Face	Not a Face	Not a Face
Lighting preset 3	Not a Face	Not a Face	Not a Face	Not a Face

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{0+0}{0+0+0+21} = 0\%$$
$$Sensitivity = \frac{TP}{TP+FN} = \frac{0}{0+21} = 0\%$$
$$FRR = \frac{FN}{TP+FN} = \frac{21}{0+21} = 100\%$$

Effectiveness of Procedure

The performance on the synthetic images has a 100% failure rate, using the rank-3 approach or the rank-1 approach. The euclidean distance between the real faces and synthetic faces were very high, to the point where the synthetic faces were not even considered faces. Using different backgrounds, expressions, lighting, age, or rotation of the synthetic images did not have an impact. In conclusion, the algorithm performs very poorly with synthetic images compared to real ones.

As an experiment to try and find the cause of this poor performance, we used photoshop to modify a synthetic face to have hair similar to the real face:

Figure 5: Real image of genuine user (left), synthetic image of genuine user (middle) and synthetic image of genuine user with hair (right)

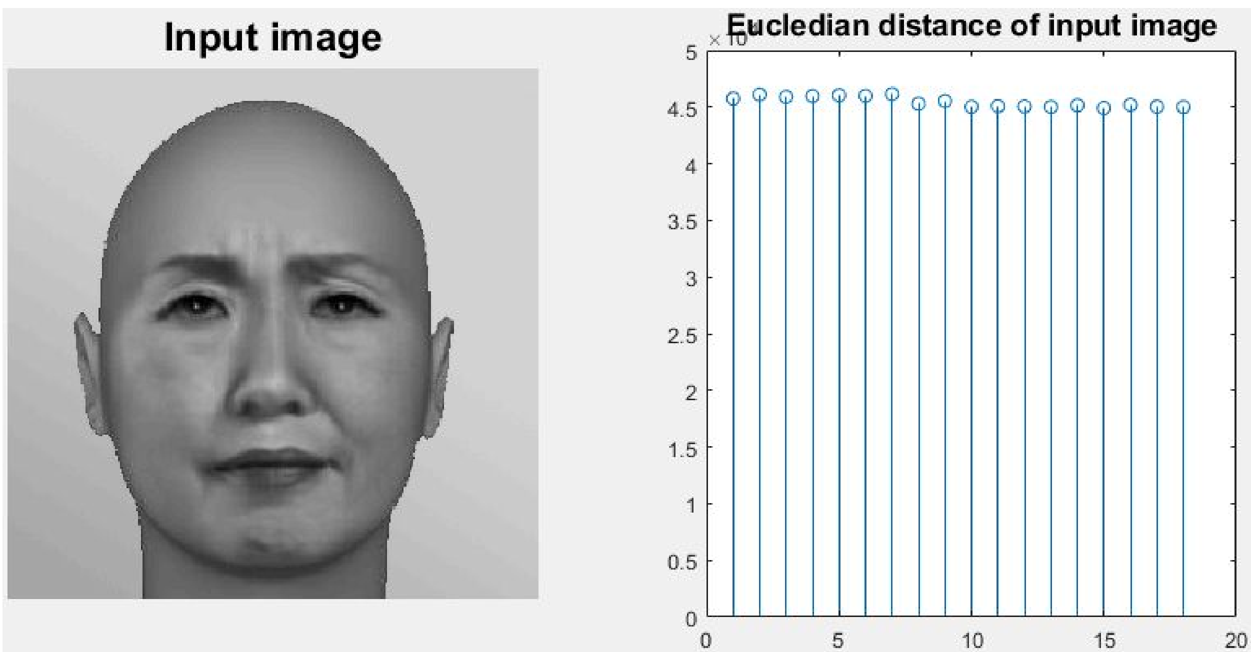


Figure 6: Synthetic image compared against the database of genuine users

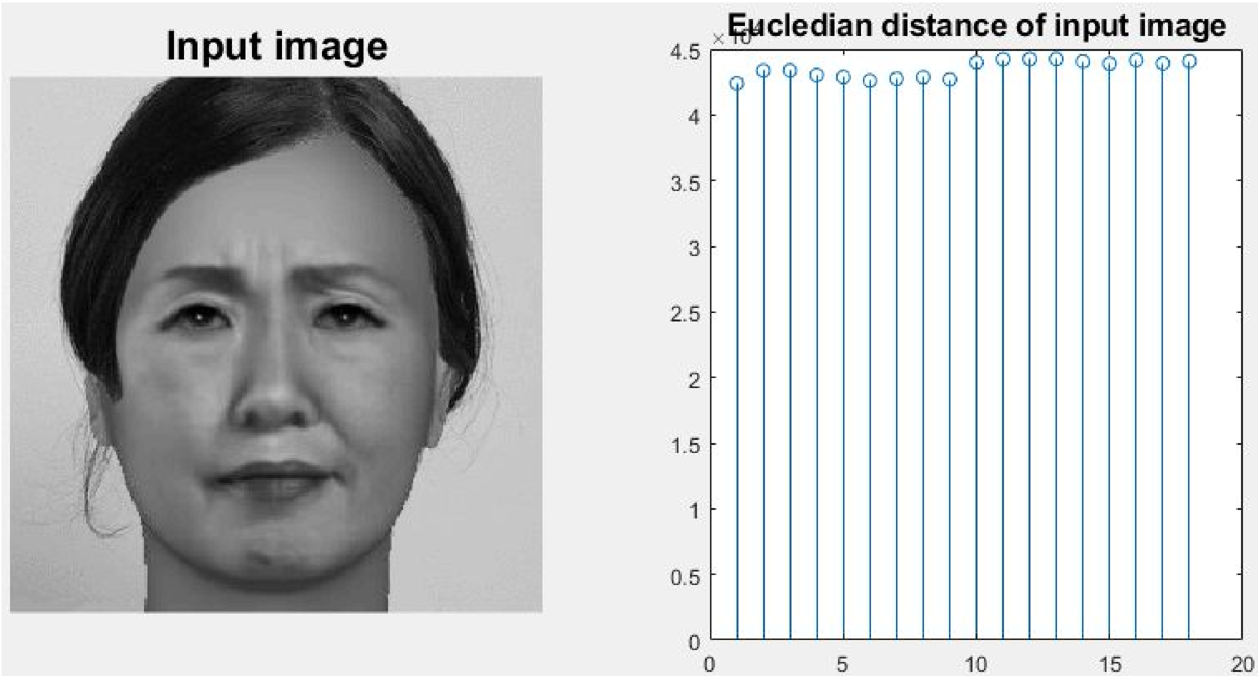


Figure 7: Synthetic image with hair compared against the database of genuine users

The basic synthetic face performed poorly and was considered not a face ("Normal") in the tables above). However, the synthetic face with hair had a minimum distance of 42403, which is under the first threshold and matches the corresponding real user in the database (true positive).

Considering that the lack of hair in the synthetic images compared to real images seems to be causing false rejection, the procedure for this lab could be improved by providing sets of images of real users who are bald. That way, the synthetic images created with FaceGen have a chance of being considered genuine. Another alternative, which would require somewhat more effort, is choosing a different facial recognition algorithm which is robust enough to ignore or properly accommodate the presence of hair. A final alternative of improving the procedure would be to update FaceGen to create synthetic hair. However, this is probably the highest effort and least feasible alternative.

Conclusion

The eigenvector-based 'eigenfaces' facial recognition algorithm used in this lab did not work very well. It achieved an overall accuracy of 74.1 when using the rank-1 approach, and an accuracy of 77.7 when using the rank-3 approach. It had particular difficulty knowing what to do with faces that were not originally in the database and used to train the algorithm. Real faces of users not in the database were often falsely accepted as in the database, or rejected as not being a face. Synthetic images of a genuine user were rejected as not being faces 100% of the time, which seems to be caused by the lack of hair in the synthetic 3D models.

Appendices

The matlab code used for this lab can be found in the lab6code folder.

All real face images used were taking from the following website, used under the non-commercial use policy:

- https://www.bodiesinmotion.photo/browse?page=1&order=popular&type=faces
- https://www.bodiesinmotion.photo/terms-conditions-website