Evaluation of Face Photos

Michigan State University

DATA

The data used in this evaluation is a subset of the previous evaluation with 1 additional image (Subject A – Exhibit 13). The previous evaluation contained many low-quality images and was filtered to retain only high-quality images. This helps ensure that the matching results are mainly a function of the biometric content of the images rather than the biometric content *and* the image quality. Tables 1, 2, 3, and 4 show the images for subject A, B, C, and D. The tables show the original images and the "cropped" image. The cropped image is the face area of the image as detected by an automated commercial face recognition SDK (RankOne Computing ROC SDK v1.18.1). The face area shows the region of image where the SDK will extract features from. The face area is resized to 96x96 in the tables.

Table 1. Images of Subject A.

Original					
Cropped					
Subject	A	A	A	A	A
Exhibit	1	4	11	12	13

Table 2. Images of Subject B. Note: the original Exhibit 10 image was cropped to reduce the width of the image and ensure the table fits on the page.

Original						

Cropped						
Subject	B	B	B	B	B	B
Exhibit	1	2	6	9	10	11B

Table 3. Images of Subject C.

Original				
Cropped				
Subject	C	C	C	C
Exhibit	1	2	3	4

Table 4. Images of Subject D.

Original				
Cropped				
Subject	D	D	D	D
Exhibit	0	2	14	16

Face Comparison

Goal

The goal of this evaluation is to assess 4 claims:

Cliam I. Subject A and Subject B are the same individual.
Cliam II. Subject C and Subject D are the same individual.
Cliam III. Subject B is a different individual than Subject C.
Cliam IV. Subject B is a different individual than Subject D.

Evaluation Setup

For our evaluation, we use the RankOne Computing ROC SDK v1.18.1 to match face images. We also make use of a reference set of images from the Labeled Faces in the Wild (LFW) dataset.[1] This assists us in classifying the scores as genuine or imposter. A genuine score is a match score that results from comparing two face images of the same subject. An imposter score is a match score that results from comparing two face images from different subjects. As the ROC SDK produces similarity scores in the range [0, 1], we expect the majority of imposter scores to occur in the lower part of the [0, 1] range and the majority of genuine scores to occur in the higher part of the [0, 1] range. The images in the reference dataset are compared to obtain a set of genuine scores and a set imposter scores. We can then visualize the distribution of both the genuine score set and the imposter score set and find specific ranges where genuine and imposter matches are most likely to occur.

Results Interpretation

We visualize the results using the boxplot and histogram techniques. An example of a boxplot is shown in Figure 1. A boxplot marks the value of certain percentiles given a set of scores. A percentile indicates the value where a certain percentage of the scores in a score set are located. For example, the 20^{th} percentile may be a value of 0.3 where 20% of the scores in the score set are below 0.3. The boxplot marks certain locations:

- The red bar shows the median (50^{th} percentile).
- The blue box shows first quartile (25^{th} percentile) and the third quartile (75^{th} percentile). Thus, the box contains the interquartile range (IQR).
- The black lines that extend from the blue box are known as whiskers. The whiskers extend to 1.5 times the IQR in either direction.
- A red plus indicates a scores which falls outside the whiskers and are called “outliers.”

An example of a histogram is shown in Figure 1. A histogram shows the distribution of a score set over a range. Each bar represents the scores in a small range and the height of the bar indicates the frequency of scores in that range. For example, bar 1 in Figure 1 represents the number of scores with value between x_1 and x_2. Since the height of the bar 1 is less than that of bar 4, the scores in the score set occur more frequently in the range $[x_4, x_5]$ than $[x_1, x_2]$.

[1] http://vis-www.cs.umass.edu/lfw/

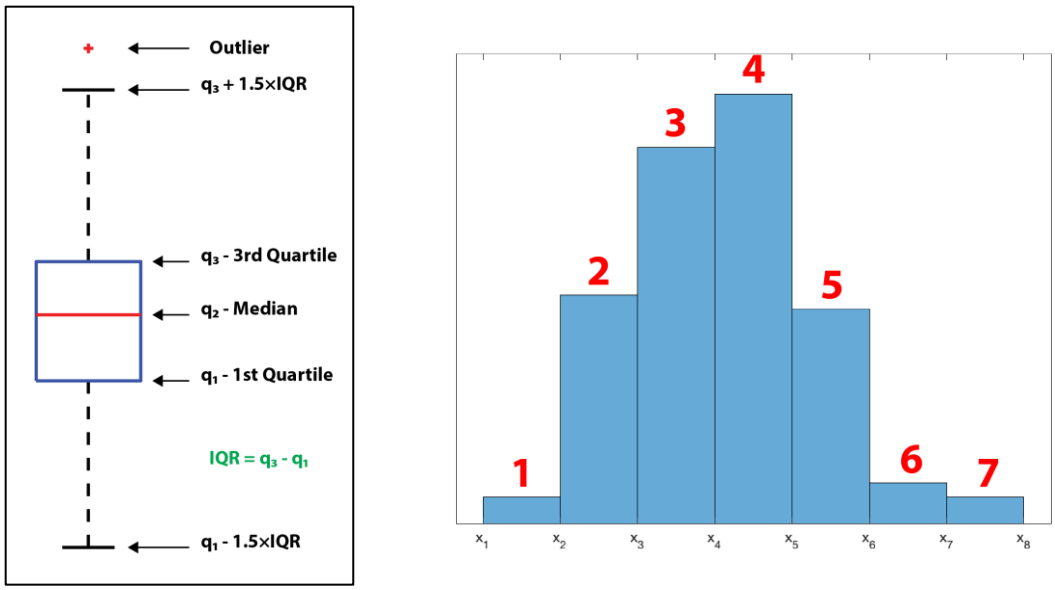

Figure 1. Example of a boxplot (left) and a histogram (right).

See https://towardsdatascience.com/understanding-boxplots-5e2df7bcbd51 for further description of box plots and https://datavizcatalogue.com/methods/histogram.html for further description of histograms.

Claim I: Are Subject A and Subject B The Same Individual?

Figure 2 shows the match scores between the Subject A images and the Subject B images. We see that the majority scores are above 0.5 and exist in a range where the probability that a score is genuine is greater than the probability that a score is imposter for the reference distributions.

Conclusion: Claim is true
Confidence: Moderately Likely

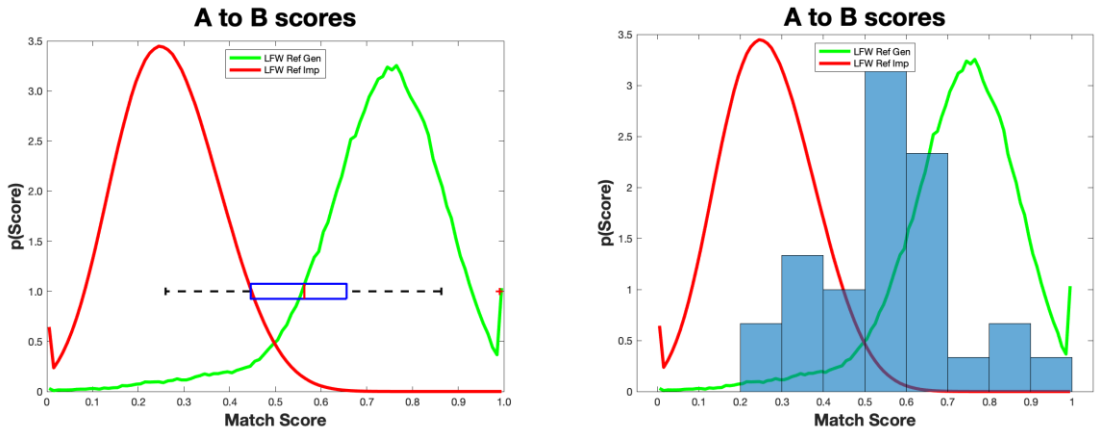

Figure 2. Boxplot (left) and histogram (right) showing the match scores between the images of Subject A and Subject B.

Claim II: Are Subject C and Subject D The Same Individual?

Figure 3 shows the match scores between the Subject C images and the Subject D images. We see that all of the match scores are above 0.6 and exist in a range where the probability that a score is genuine is greater than the probability that a score is imposter for the reference distributions.

Conclusion: Claim is true
Confidence: Very Likely

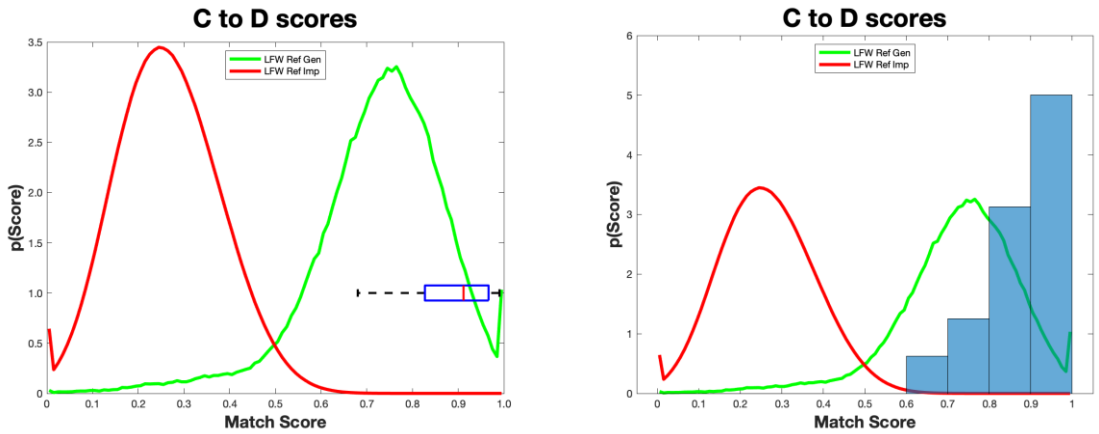

Figure 3. Boxplot (left) and histogram (right) showing the match scores between the images of Subject C and Subject D.

Claim III: Is Subject B a Different Individual Than Subject C?

Figure 4 shows the match scores between the Subject B images and the Subject C images. We see that the majority scores are above 0.5 and exist in a range where the probability that a score is genuine is greater than the probability that a score is imposter for the reference distributions.

Conclusion: Claim is false
Confidence: Likely

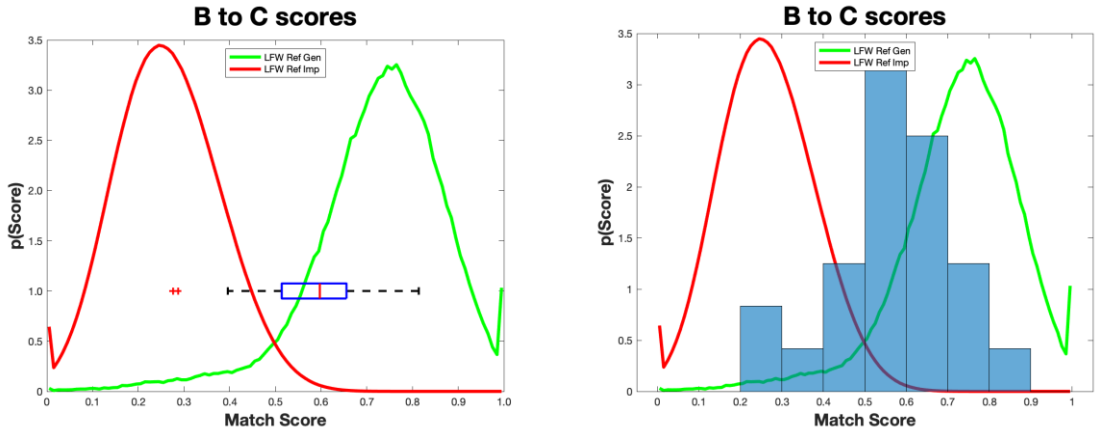

Figure 4. Boxplot (left) and histogram (right) showing the match scores between the images of Subject B and Subject C.

Claim IV: Is Subject B a Different Individual Than Subject D?

Figure 5 shows the match scores between the Subject B images and the Subject D images. We see that a large majority scores are below 0.5 and exist in a range where the probability that a score is imposter is greater than the probability that a score is genuine for the reference distributions.

Conclusion: Claim is true
Confidence: Likely

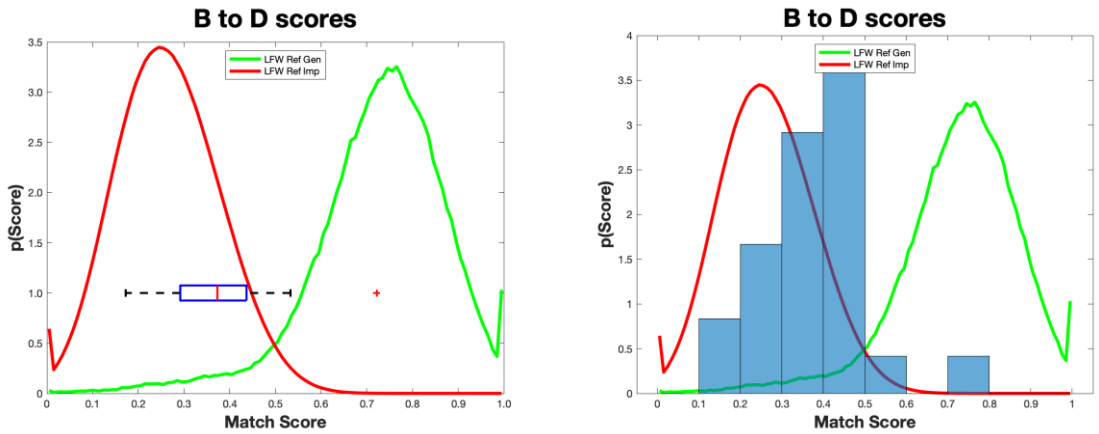

Figure 5. Boxplot (left) and histogram (right) showing the match scores between the images of Subject B and Subject D.