Biometrics (CSE 40537 / 60537)
University of Notre Dame, Spring 2022
Assignment 4: Face Identification
Due date: April 18, 11:59 PM ET
Total: 10 points

1. Introduction

The purpose of this assignment is to train and evaluate the students' capabilities to use, set the operating point, and perform biometric identification with a third-party library of face recognition, over a partition of the Yale Faces Database (https://bit.ly/3uUhBOg). To do so, two zip files are being provided, one containing the third-party library (which was presented in class), and the other containing the dataset partition.

1.1. Face recognition library

The face recognition library is available at https://bit.ly/3qZuTYK. To use it and complete this assignment, you will need a computer with *Python 3* (https://bit.ly/3j0ymBZ), plus the *numpy* (https://bit.ly/3j1mEae), *OpenCV* (https://bit.ly/3j1ElXi), Astropy (https://bit.ly/3K80PBH), and ArcFace (https://bit.ly/3J8Tgtc) libraries and implementations. A quick install of these libraries can be done through *PyPI* (https://pypi.org/), see Figure 1:

```
user@host:~$ pip3 install numpy opencv-python astropy arcface
```

Figure 1. Command line leveraging *PyPI* to install the needed libraries.

A good starting point to solve this assignment is the *Python* program implemented as *main.py*, within the face recognition library. It currently works by providing pairs of file paths that store images depicting faces and the extension to solve this assignment should be straightforward to the students (and it is indeed expected from them).
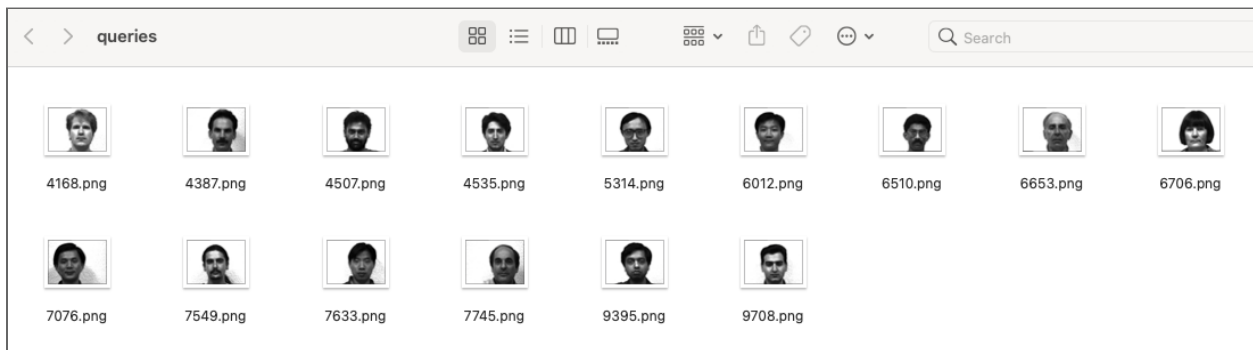
1.2. Face dataset

The face dataset is available at https://bit.ly/3uSFCVT and is a modified subset of the Yale Faces Database (https://bit.ly/3uUhBOg). The data is organized into two folders, namely "dataset" and "queries". The "dataset" folder contains 99 images, each one depicting a single face, whose names obey the following format: "`subject<ID>.<CAPTURE_CONDITION>.png`". As a consequence, two files with the same `<ID>` depict the same individual, hence comprising a genuine pair. Figure 2 provides a summary of the content of "dataset" with 14 images and two individuals.
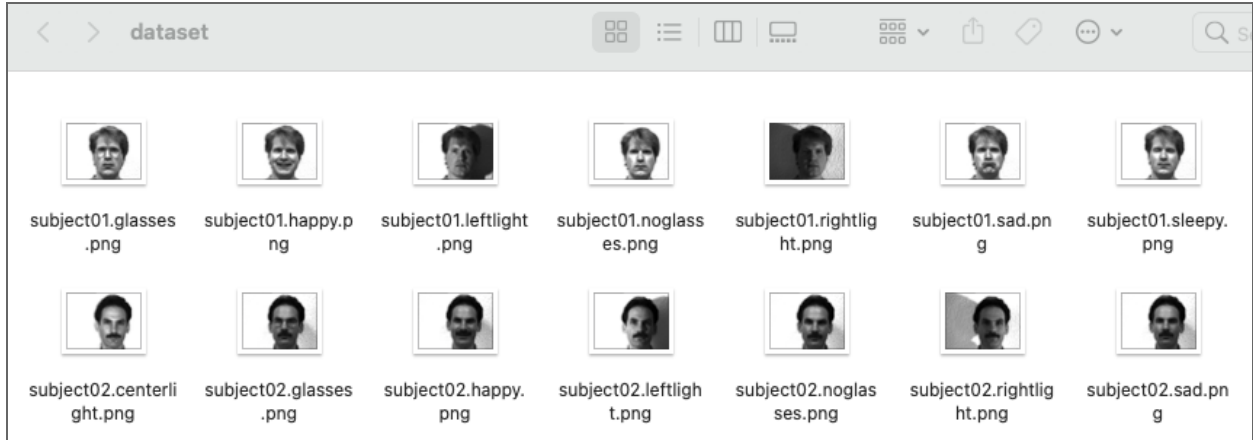
Figure 2. Example of the content of the folder "dataset".

The "queries" folder, in turn, contains 15 image files, whose names are random numbers and have no particular meaning. These images do not have exact copies within the "dataset" folder and the depicted individuals may or may not have images with their faces within the "dataset" folder. Figure 3 depicts the full content of "queries".

Figure 3. Example of the content of the folder "queries".

1.3. Assignment directions
After downloading and unzipping the contents of the face recognition library and face dataset, follow the instructions and answer the questions presented in Section 2.

There is no formal template for providing your answers. You may use the editor you like. The following submission is expected:
- A single PDF file or Word document containing your answers.

Please share your answers through Slack (https://bit.ly/3H8kajL, *@dmoreira*) or send them to daniel.moreira@nd.edu by April 18, 2022, 11:59 PM ET.

2. Questions

2.1. As explained in class, the provided third-party face recognition library is able to extract a 512-dimensional float feature vector from a given face image, as well as calculate the angular distance between two feature vectors, using ArcFace (https://bit.ly/3J8Tgtc). The expected behavior for the software is to generate small distances for two face images that depict the same individual (genuine pair), and large distances for two images that depict different individuals (impostor pair).

Leveraging the content of **only** the "dataset" folder within the provided data, the third-party face recognition library, and metrics learned in class, please determine **what is a good angular distance threshold to separate genuine from impostor pairs**. While providing your answer for the distance threshold, please explain in detail how you computed it. (1 point)

2.2. **What is the AUC** of the face recognition system you are using? In addition, please provide a graph with the system's **ROC curve**. (1 point)

2.3. By leveraging the face recognition system and the distance threshold previously computed, and by either capturing your face with your webcam or providing an image with your face, find within the "dataset" folder what is the individual that is the most similar to you. **Please provide the subject ID and the angular distance between your face and theirs**. In your opinion, do you have anything in common with this subject (e.g., gender, ethnicity, age, etc.)? If yes, what is it? (2 points)

2.4. By leveraging the face recognition system and the distance threshold previously computed, **please provide the subject ID** (or "**UNKNOWN**", if the individual does not have a face within the "datasets" folder), as well as **the respective angular distances** that supported your decision, for each one of the 15 images provided within the "queries" folder. (6 points)