Internal	Technical Assessment	Biometix BIOMETRIC DATA SOLUTIONS

Purpose

The purpose of this task is to better understand your technical abilities. There are a number of elements being tested, you may not excel in all areas.

Description of task

We will be performing a performance assessment of a facial recognition (FR) algorithm. This will take place in the following steps:

1. Establish an appropriate environment to run the FR algorithm and output results
2. Run the FR algorithmic over the dataset, outputting match scores for each pair
3. Do an initial analysis of the produced in stage 2
4. Write a short brief summarising your findings

Further instructions for each stage are provided below.

Note on the data: The data being used is the FERET dataset (pruned). This data was collected by NIST through the 90s for the US Department of Defence in order support Biometric system testing. For the sake of this exercise you may:

- Assume there is only one face in each image (sample)
- Assume that the first five digits in the file name uniquely identifies the subject

You may need to sanitize the directory of files that are not images.
For further information on the dataset refer to the NIST FERET website[1].

Instructions

1. **Setup:** You will need to run be running Python 2 or 3. Following this you must install the face_recognition module (`pip install face_recognition`).
   You can refer to the installation instructions on https://github.com/ageitgey/face_recognition for more information.
2. **Run:** You will have to run the FR engine to output a match score for each combination of images in the data set. Refer to the face_recognition documentation for functions you may wish to call (https://face-recognition.readthedocs.io/en/latest/face_recognition.html).
   You will have to process the file names from the OS and run basic to text matching (first five digits) to determine if the pair of images belong to the same subject (mated).
   This will output a csv with four columns: prode id (image filename), subject id (image filename), match score and mated/non-mated (1/0).
   NB: It may be worth reading into the face_recognition documentation to understand how the engine outputs match scores.
   You may find that you need to use
   ```
   score = 1 - np.linalg.norm(face_encodings - face_to_compare, axis=1)
   ```
   to return a matching score).
3. **Analysis:** You may perform simple descriptive analysis in excel (or python if you prefer). In order perform:

[1] http://www.itl.nist.gov/iad/humanid/feret/feret_master.html

   a. Histograms for each genuine and non-genuine scores (count against match score)
   b. Find the threshold where would there be an equal rate of False Matches (results above the threshold[2] who are not mated) and False Non-Matchs (results below the threshold who are mated)? What is the rate of False Matches and False Non-Matches in this condition (the equal error rate)?
4. **Brief:** Write a short brief (approximately half a page) on the exercise completed, including the following:
   a. Answers to the questions posed in section 3 point a) and b).
   b. A subjective commentary on you experience: issues encountered, challenging tasks, simple tasks and anything else interesting.

Once this is completed please provide your python script, csv/excel notebook and the briefing document.

[2] The threshold is a setting where any score higher than this the system considers to be a match, any score below this the system deems to be a non-match.