Alexandria University
Faculty of Engineering
Computer and Systems Engineering Dept.

Prof Dr. Nagwa M. El-Makky
Prof. Dr. Marwan Torki
Special Topics in Information Systems

Assignment 2
Face Recognition

Problem Statement

You are required to perform face recognition. Face recognition means that for a given image you can tell the subject id. The given database of subjects is very simple. It has 40 subjects. The required steps to achieve the goal of the assignment are shown below.

1. Download the dataset and understand the format
    a. You will use the ORL dataset for this assignment. http://cam-orl.co.uk/facedatabase.html
    b. The data is available at the following link. http://www.mediafire.com/file/83akhrn9n7ynj0f/orl_dataset.zip/file
    c. The dataset has 10 images per 40 subjects. Every image is a grayscale image of size 92x112.

2. Generate the Data Matrix and the Label vector
    a. Convert every image into a vector of 10,304 values corresponding to the image size.
    b. Stack the 400 vector into a single Data Matrix D and generate the label vector y. The labels are integers from 1:40 corresponding to the subject ids.

3. Split the Dataset into Training and Test sets
    a. From the Data Matrix $D_{400x10304}$ keep the odd rows for training and the even rows for testing. This will give you 5 instances per person for training and 5 instances per person for testing.
    b. Split the labels vector accordingly.

4. Dimensionality Reduction using PCA and Classification using 1-NN
    a. Use the pseudo code of PCA given in lecture to find the projection matrix U. try the following values of alpha, {0.8,0.85,0.9,0.95}
    b. Project the training set and test sets **separately** using the same projection matrix.
    c. Use a simple classifier (first Nearest Neighbor to determine the class labels).
    d. Report Accuracy for every value of alpha separately.
    e. Can you find a relation between alpha and the classification accuracy

5. Hyper-parameter Tuning
    a. Set the number of neighbors in the K-NN classifier to 3,5,7.
    b. Break ties at your preferred strategy.
    c. Plot (or tabulate) the performance measure (accuracy) against the K value.

6. BONUS

a. Use different Training and Test splits. Change the number of instances per subject to be 7 and keep 3 instances per subject for testing. compare the results you have with the ones you got earlier with 50% split.

Notes

1. You should deliver well documented code as well as a report illustrating every step in the assignment.
2. Copied assignments will be penalized. So, not delivering the assignment would be much better.
3- You should work in groups of 2

GOOD LUCK