# CSE 471: Mini Project: Face Recognition using Eigen Faces

## Goal

The goal of this mini project is to get you familiarized with the ideas of image representation, PCA and LDA, and face recognition. You will also understand the practical difficulties in developing real-world systems that work with acceptable accuracies.

## Basics of Face Recognition

Biometric recognition, including face recognition, can be carried out in two modes: a) Identification, and b) Verification. In both cases, a few training samples from each class (each person) is collected and stored in a reference database during training. This process is called enrollment. During testing, a test sample is presented, which should be compared against the enrolled samples to decide if they match or not. The matching process differs in the two modes described above. In the identification mode, the test sample is compared against training samples of each of the enrolled persons, and the best match is identified to decide the identity of the person providing the test sample. This is a one-to-many matching process. In the verification mode, the test sample is presented along with a claim of identity. The test sample is matched only against enrollment samples of the claimed identity to decide whether it is a match or not. The process hence verifies whether the identity claim is correct or not. This can be thought of as a one-to-one matching, even though one might compare the test sample against multiple enrollment samples of the claimed person.

## Experiments

### 1. Yale Dataset: Eigen Faces + K-NN/Euclidean Distance

In the first experiment, you will take the Yale face dataset that contain face images that are captured under controlled conditions (illumination, pose and expression) and cropped to contain the face only. The dataset should be divided into four parts and you should perform a 4-fold cross validation while reporting the accuracies. While dividing the training data, make sure that you divide images of each person separately so that a few images of each person is present in each of the partitions. Use the training set to develop the Eigen faces and represent each of the training images as a linear combination of the Eigen face basis. Each of the test samples is then represented in the same Eigen basis, and compared against the training samples using Euclidean distance metric. You should perform both identification and verification experiments using this data, and report the average of results from the four folds.

### 2. CMU-PIE Dataset: Eigen Faces + K-NN/Euclidean Distance

Repeat the above experiment, but this time using the CMU-PIE dataset. Note that this dataset contains face image captured under various poses, illuminations and expressions (hence PIE). How does the accuracy compare against the Yale dataset? What do you infer?

### 3. Using an SVM Classifier for Face Recognition

Compute the Eigen face representation of all the training and test samples and use an SVM classifier (libsvm) to carry out the identification experiment for Yale dataset. Compare the results with the identification experiment using Euclidean distance based k-nearest neighbor classifier used in Exp. 1.

### 4. Representation of Unknown faces and Non-faces

Consider a variety of images (both faces and non-faces, including your own face image) of the same size as that of the Eigen basis computed in the previous experiment. Compute the representation of these images using the Eigen Face basis and reconstruct the image from the representation. How face-like is each image? How faithful is the reproduction? What do you infer from the results? Try to come up with interesting examples for both faces and non-faces.

### 5. Extra Credit Problem: Classroom Attendance

You will be given some face images detected from the photographs taken in a selected set of classes. These faces may be of different sizes, pose and illumination. Carry out both identification and verification experiments using the hold-one-out methods. That is a cross validation experiment where one of the classes is used as test set and the others as training data.

## Reporting the Results

The identification results should be presented as accuracy of recognition (percentage of time when the input is correctly classified). The result reported should be the average of the 4-fold cross validation. For verification experiments, you should form a large number of image pairs with each pair containing one training image and one test image. The pairs are divided as genuine pairs and imposter pairs depending on whether the pair contains sample from the same person or different persons. Compute the distance or similarity (inverse of distance) for each pair to obtain a list of genuine similarity scores and another list of imposter similarity scores. Use the two lists to plot the Receiver Operating Characteristic Curve (ROC Curve) for the matching.

## Datasets

The Yale and CMU-PIE datasets are available on the internet for free download. The classroom dataset will be provided to you by the TAs.

## Report and Submission

The mini project should be done in groups of two. The submission contains the code that you used for running the experiments and computing the ROC curves and accuracies. You may do the project is Matlab/C/C++/Python or any other language with the permission from the TAs. Submit your codes (likely to be very small) along with a detailed report that describes the problem, datasets, how the experiments were conducted, what the results are, and most importantly, a detailed analysis of what the results mean.

Make the report as professional as possible and include representative and interesting results and images from your experiments. Include the images and graphs as figures with numbers and meaningful captions, and refer to them in the body of your report during the analysis. Similarly, present the

identification experiment results as tables again with table numbers and captions. Please refer to any research paper (including the one on Eigen Faces) to see how the figures and tables are used in a technical report. Include a list of references at the end of your report. Once again each reference you include at the end should be referred to from at-least one point in the body of the report. Give appropriate titles for each of the section/sub-section.

Please note that the report should be completely written in your own words and not copy-pasted from any source of any part of the report (including dataset description). The report is evaluated for your ability to describe the experiments, results and analysis, correctly and clearly. So by copying you are claiming the actual authors ability to write as your own, which is clearly cheating. Similarly, the code used for experiments (except for SVM) should be completely your own. You will get a zero in the project in case any copy is detected and may further result in failure in the course. The project will be scored on a scale of 0 to 100, and the extra credit problem will be given a maximum of 20 points.

The deadline for submission will be 15$^{th}$ Nov., 2013, at 10pm. Please note that this mini project is to be done in groups of 2, and the groups should be informed to the TAs by Wednesday, 6$^{th}$ Nov.