STA 5106: Midterm Project

(Thursday, October 10)
Due: Thursday, October 24

1. Problem

Discovering known objects in camera images is an important and challenging area of research with applications in medical diagnosis, military surveillance, underwater search, satellite navigation, etc. The basic problem is as follows: given an image (or multiple images) of a particular scene the goal is to detect and identify objects of interest in that image. We will look at the specific problem of recognizing people from their facial images. We will assume that we have some images of known people that can be used to compare and recognize new images (of the same set of people).

2. Method

Since images are very high dimensional, it is not easy to analyze them directly. A common approach is to reduce their dimensionality using principal component analysis (PCA), linear discriminant analysis (LDA), and other similar methods. There are two sets of images available for such experiments: (i) training set: these images are already recognized and labeled by some expert (maybe humans), and (ii) test images: these are new images that are to be recognized and labeled. The goal is to use the similarities between the test and the training images to label the test images.

In PCA, one computes the principal components of the training images, projects them down to a smaller size. That is, each training image is now represented by a small vector of size k. For any test image, one can also project it down to a k vector, using the same projection, and then find the nearest image in the training set by comparing their k vectors. In LDA, the low dimensional projection is determined using scatter matrices.

We will assume n_2 training images each for n_1 people in our database. The size of each image is $s_1 \times s_2$. These images are taken at different orientations, different facial expressions, etc. Figure 1 shows some examples of the images. The left panel shows an image of original size, while the remaining panels show images of the size provided with the dataset.

There are two parts to the procedure. One is to analyze the training images and compute a projection to their principal k-dimensional subspace. The second part is to recognize test images by projecting them into this k-dimensional subspace and comparing with the training data. We start by outlining the first part: Let the training images be arranged as vectors in a matrix of size $(s_1 \times s_2) \times (n_1 \times n_2)$. Call it Y_{train}. (You will be given Y_{train}). First n_2 columns are images of person 1, next n_2 columns are images of person 2, and so on, with a total of $n_1 \times n_2$ columns.
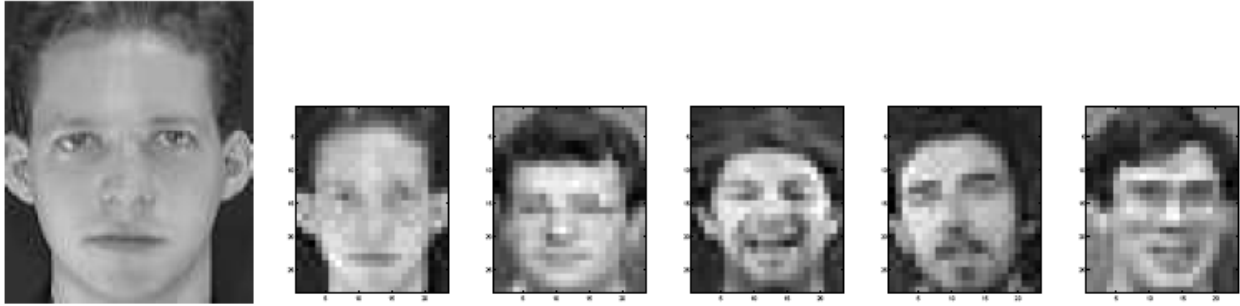
Figure 1: An example of the original Picture and some down-sampled pictures provided for this project. These images are taken from ORL face database.

Feature Extraction:

(a) **PCA:** Perform PCA of Y_{train}. A small difference from the earlier homework problem is that now each column (not each row) is an observation. Let U_1 be the first k columns of the orthogonal matrix U in the SVD decomposition. The size of U_1 is $(s_1 \times s_2) \times k$.

(b) **Simple Projection:** For comparison, we use another projection where U_1 is simply the first k columns of $(s_1 \times s_2)$ identity matrix.

Each image in Y_{train} can now be reduced to a k-dimensional vector using the projection $Y_1 = U_1^T Y_{\text{train}}$. Y_1 is of the size $k \times (n_1 \times n_2)$.

Classification: Now we are ready to perform classification (recognition). You are given n_2 test images per person, and the test set Y_{test} is in the same form as the training set. Take one image I from the test set, i.e. select a column from Y_{test}.

(a) **Form feature vector:** Compute the projection of I using $I_1 = U_1^T I$. I_1 is a $k \times 1$ vector.

(b) **Compute Metric:** Compute the distant between I_1 and each column of Y_1 using the 2-norm.

(c) **Find Nearest Neighbors:** Find the label of the column that has the smallest distance to I_1. If this label matches the true label then recognition is successful otherwise it is a failure.

To compute an average performance, perform the above procedure for each image in the test data and compute the percentage of successful recognition. Call this number $F(k)$. Generate this plot for each of the two projections: PCA and the simple projection. Study the variation of $F(k)$ versus k for $k = 1, 2, \ldots, 40$.

3. Assignment:
Two files, one containing the training set and the other containing the test set, are posted on the blackboard website for this class. For these image the image size is $s_1 = 28$, $s_2 = 23$,

the number of people is $n_1 = 40$, the number of training images per person is $n_2 = 5$. So Y_{train} is a matrix of size 644×200. Y_{test} happens to be of the same size.

Your final report should include descriptions of:
(a) The problem statement,
(b) Methodology or Approach,
(c) Matlab programs,
(d) Results: Demonstrate by a few examples the test image and the closest image in the training set. Tabulate the average performance of correct recognition versus the number k.
(e) Finally, write a short summary analyzing the effectiveness of both projections. Comment of the variation of $F(k)$ versus k in each case.

Points are allocated towards presentation of results and clarity of your report.