Project 4 - Dimensionality Reduction & Classification

Instructor: Linwei Wang

1 Overview

In this project, you will experiment with the application of dimensionality reduction and classification methods in face recognition. You will be given a well-known dataset of face images, which consists of 10 different images for each of the 40 different subjects. On this dataset, you will experience the use of PCA to extract the so-called *eigenfaces* and use them in combination with linear classification methods for image reconstruction, face recognition, and face identification. As the last project of the course, this project will require more creativity and independent learning ability from you with only a minimal introductory information given for the need of the projects.

2 Problems

The application of PCA in face recognition, first developed in 1980s, has been proved to be a great success. The term *eigenface* represents a set of principal components generated from PCA on a large set of images depicting different human faces. They can be considered a set of *standardized face ingredients*, a linear combination of which can then renders the face of an individual human subject. In another word, the facial image of a person is no longer represented on the image domain, but on the domain of the eigenfaces. In this project, you will experience the application of the simple technique of PCA in a set of classic problems.

3 Task

A dataset of face images are provided on MyCourses. This is also known as the AT&T face database.

This database consists of 400 facial images from 40 subject $\times$ 10 image per subject. After figuring out how to load these images into your matlab, you will go through the following tasks:

- Dimensionality reduction
  1. Use PCA to extract eigenfaces on all 400 images, examine
     (a) How do the (leading) eigenfaces look like as an image (show some examples in your report)?
     (b) How does the importance of the eigenfaces decrease?
  2. Face reconstruction with PCA: On a subset of images from different subjects, reconstruct the face image using different numbers of principal components.
     (a) Observe the difference between reconstructed and original images, as the number of eigenfaces used in reconstruction increases.
     (b) How many eigenfaces are required to recover an original face with reasonable errors?
- Bonus: Classification: use one of the classification methods outlined in lecture notes or external reading materials for the following tasks. You will divide the dataset into a training set and a testset. Make sure you training set has images from 35 subjects and 8 image per subject. Use the remaining 2 image for each of the 35 subjects as well as all images for the other 5 subjects as test images.
  1. Face recognition: Given a new image, recognize whether it is a face image or not. Test it on the left-out 2 $\times$ 35 subjects, the left-out 10 $\times$ 5 subjects, and some "non-face" images.
  2. Face identification: Given a new face image, identify and label it to the correct person in the training set
- 10 points for each clas-sification method applied as described above.

4 Deliverables

1. Code. Please submit all the code, along with a brief documentation. I should be able to run your code based on these.
2. Report. Three page maximum including figures, single column. Please be clear and concise. You are expected to include (but not restricted to) to include the answers to the above questions asked.

3. Presentation. You will give a 10 minute presentation as a group.

5 Grading

- Content: 50%
- Report: 20%
- Presentation: 30%
- 10 points bonus for an extra method attempted