

COEN 140 Machine Learning and Data Mining

Lab Assignment #9: Face Recognition

Guideline: Submit a pdf report to Camino. Also submit all the source code needed to generate the results as a separate zip file to Camino.

You are given a face image database of 10 subjects (in the att_faces_10.zip file). Each subject has 10 gray-scale images of 112×92 pixels. You will use the database for a face recognition task. For simplicity, for each subject, use face images 1,3,4,5,7,9 as the training images, and face images 2,6,8,10 as the test images. Convert each image to a vector of length $D=112 \times 92 = 10304$. Stack 6 training images of all 10 subjects to form a matrix of size 10304×60 . Apply singular-value decomposition (SVD) for dimensionality reduction. Find the top- K left singular vectors ($K=1,2,3,6,10,20, 30$, and 50) corresponding to the K largest singular values of the data matrix. Project the face images to the top- K left singular vectors and apply the nearest-neighbor classifier in the reduced dimensional space. Plot the recognition accuracy rate ($\frac{\text{number of correct classification}}{\text{total number of test images}} \%$) versus different K values. Analyze the results you observe.

Note: the plotted figure should also include: xlabel, ylabel, x-coordinates, y-coordinates, legend, grid, and a figure caption.

Demo/Explain to TA (10%):

1. How do you construct the training data matrix?
2. How do you do SVD and find principal components of a different rank?
3. How do you project training and test images onto the principal components?
4. How do you do nearest-neighbor classification?

Grading:

Report (60%)

Source code (30%)

Demo (10%)