Programming Assignment 3

CS 479

`https://github.com/alexander-novo/CS479-PA3`

Nikhil Khedekar	Mehar Mangat	Alexander Novotny
33%	33%	33%
??	??	??

Due: April 26, 2021
Submitted: April 26, 2021

Contents

1 Theory

This assignment considers the usage of Principal Component Analysis for dimensionality reduction for a dataset of images for the purpose of face recognition.

1.1 Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a method for reducing the dimensionality of high dimensional data while preserving as much information as possible. The way this works is by the projection of the sample data along its principal distribution directions. Intuitively, this can be understood by finding the directions that capture the highest variance of the data and projecting the data along these directions.

For some given sample data $\mathbf{X} = \{\vec{x}_i \in \mathbb{R}^N, 1 \leq i \leq M\}$, where M is the number of samples in the data set, the sample covariance matrix Σ_x can be found using eq. (1) and eq. (3), where $\bar{x}$ is the sample mean and Φ_i are the normalized data points:

$$\bar{\mathbf{x}} = \frac{1}{M}\sum_{i=1}^{M}\mathbf{x}_i \tag{1}$$

$$\mathbf{\Phi_i} = \mathbf{x}_i - \bar{\mathbf{x}} \tag{2}$$

$$\mathbf{\Sigma_x} = \frac{1}{M}\sum_{i=1}^{M}(\mathbf{x}-\bar{\mathbf{x}})(\mathbf{x}-\bar{\mathbf{x}})^T = \frac{1}{M}\sum_{i=1}^{M}\mathbf{\Phi}_i\mathbf{\Phi}_i^T \tag{3}$$

Equation (3) can be further simplified to eq. (4) using eq. (5).

$$\mathbf{\Sigma_x} = \frac{1}{M}\mathbf{A}\mathbf{A}^T \tag{4}$$

$$\mathbf{A} = [\Phi_1\Phi_2...\Phi_M] \tag{5}$$

The eigenvectors and eigenvalues of the covariance matrix have the special property that the eigenvectors form an orthogonal basis in R^N for the sample data with the corresponding eigenvalues representing the variance of the data in these directions. The eigenvectors $\mathbf{u}_i$ and corresponding eigenvalues λ_i are obtained using eq. (6).

$$\mathbf{\Sigma_x}\mathbf{u_i} = \lambda_i\mathbf{u}_i \tag{6}$$

In our notation, we assume that the eigenvalues are indexed in descending order, namely, $\lambda_1 > \lambda_2 > ... > \lambda_N$. Using the eigenvectors as the new basis we can represent our data samples $\mathbf{x} \in \mathbb{R}^N$ by eq. (7)

$$\mathbf{x} - \bar{\mathbf{x}} = \sum_{i=1}^{N} \mathbf{y}_i \mathbf{u}_i = y_1 \mathbf{u}_1 + y_2 \mathbf{u}_2 + ...y_N \mathbf{u}_N \tag{7}$$

where $y_1, y_2...y_N$ are the coefficients in this new basis and are given by eq. (8).

$$y_i = \frac{(\mathbf{x} - \bar{\mathbf{x}})^T \mathbf{u}_i}{\mathbf{u}_i \mathbf{u}_i^T} = (\mathbf{x} - \bar{\mathbf{x}})^T \hat{\mathbf{u}}_i \tag{8}$$

where $\hat{\mathbf{u}}_i$ represents the unit vector in the direction of $\mathbf{u}_i$. The final dimensionality reduction is carried out on the basis of eq. (7) in that only the components corresponding to the first $K (K << N)$ eigenvectors are kept and the remaining components are discarded. This formulation approximates $\mathbf{x}$ by $\hat{\mathbf{x}}$ for which eq. (7) can be rewritten as eq. (9).

$$\hat{\mathbf{x}} - \bar{\mathbf{x}} = \sum_{i=1}^{N} \mathbf{y}_i \mathbf{u}_i = y_1 \mathbf{u}_1 + y_2 \mathbf{u}_2 + ...y_K \mathbf{u}_K \tag{9}$$

eq. (9) can be written in matrix form instead of the summation as eq. (10)

$$\hat{\mathbf{x}} - \bar{\mathbf{x}} = \mathbf{U} \begin{bmatrix} y_1 \\ y_2 \\ ... \\ y_K \end{bmatrix} \tag{10}$$

where $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_K]$ is an $N \times K$ matrix with its columns as the K largest eigenvectors of the covariance matrix $\mathbf{\Sigma}_x$. Equation (10) can also be written as

$$\begin{bmatrix} y_1 \\ y_2 \\ ... \\ y_K \end{bmatrix} = \mathbf{U}^T (\hat{\mathbf{x}} - \bar{\mathbf{x}}) \tag{11}$$

now represents each data sample as x This can be re-written as a linear transformation as in

$$TODO \tag{12}$$

1.2 Eigenfaces and Facial Recognition

The first widely successful method of solving the facial recognition problem is the eigenface method. An eigenface

2 Implementation

3 Results and Discussion

The calculated sample mean images for both data set can be found in fig. 1. Note the appearance of a single pixel-wide border of dark pixels around both means - this seems to come from some sort of artifact in the original data sets consisting of slightly darker pixels on the border.

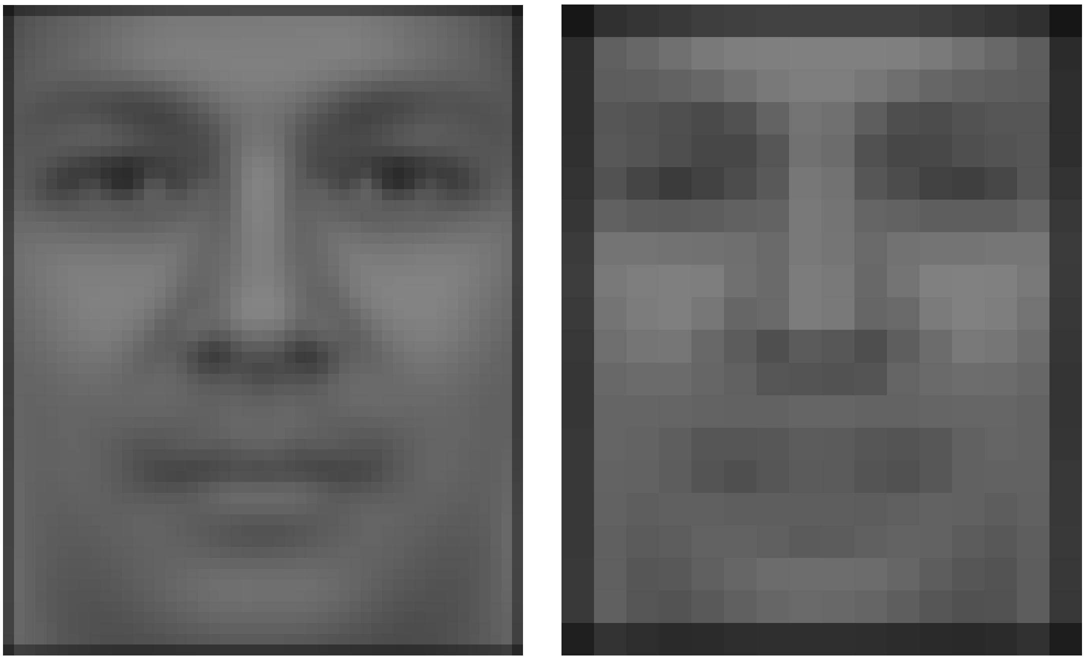

Figure 1: The sample mean face for the high resolution data set (left) and the low resolution data set (right).

The 10 largest and smallest eigenfaces can be found in figs. 2 and 3. Note that the largest eigenfaces all look like faces in some way, while the smallest eigenfaces might as well be random noise. This is a good justification for dimensionality reduction using PCA - these smallest eigenfaces are contributing very little meaningful data. Also note that each eigenface calculated from the low resolution data corresponds to one calculated from the high resolution data almost exactly. This makes sense, since the low resolution images are just sampled version of the high resolution images, so it would make sense that the principal components are just sampled version of the principle components of the high resolution images. Note that eigenfaces 7 and 8 correspond but are inverted between data sets - this is an example of how inverted eigenvectors are still eigenvectors.

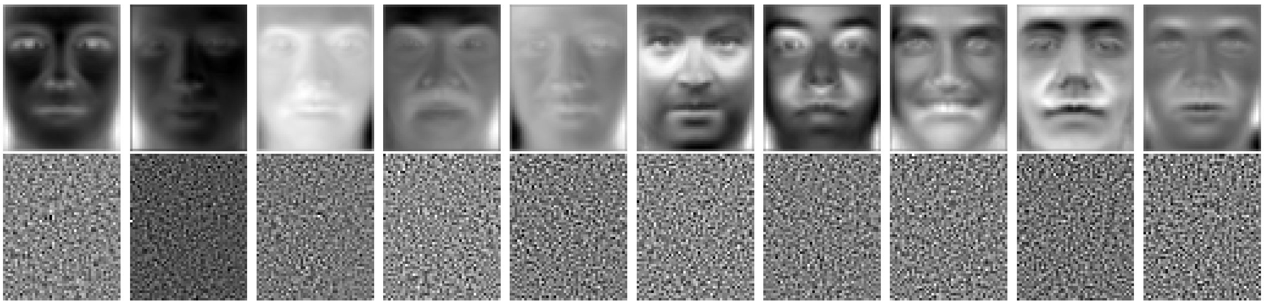

Figure 2: The 10 largest eigenfaces in decreasing order for the high resolution data set, followed by the 10 smallest eigenfaces in increasing order.

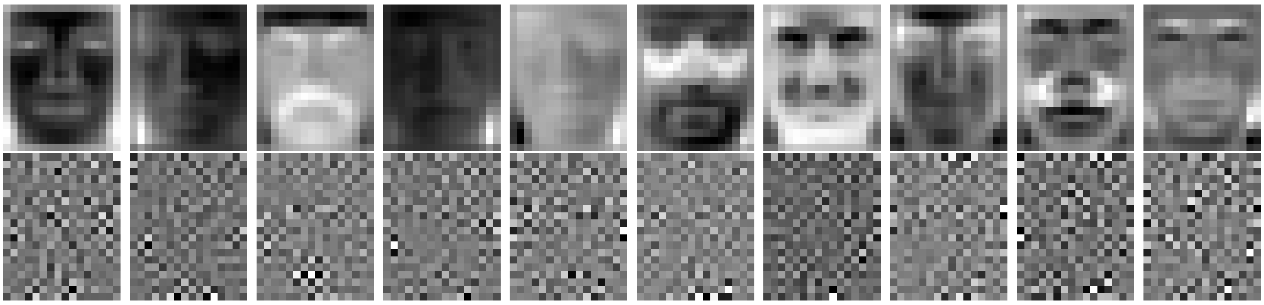

Figure 3: The same figure as fig. 2, but for the low resolution data set.

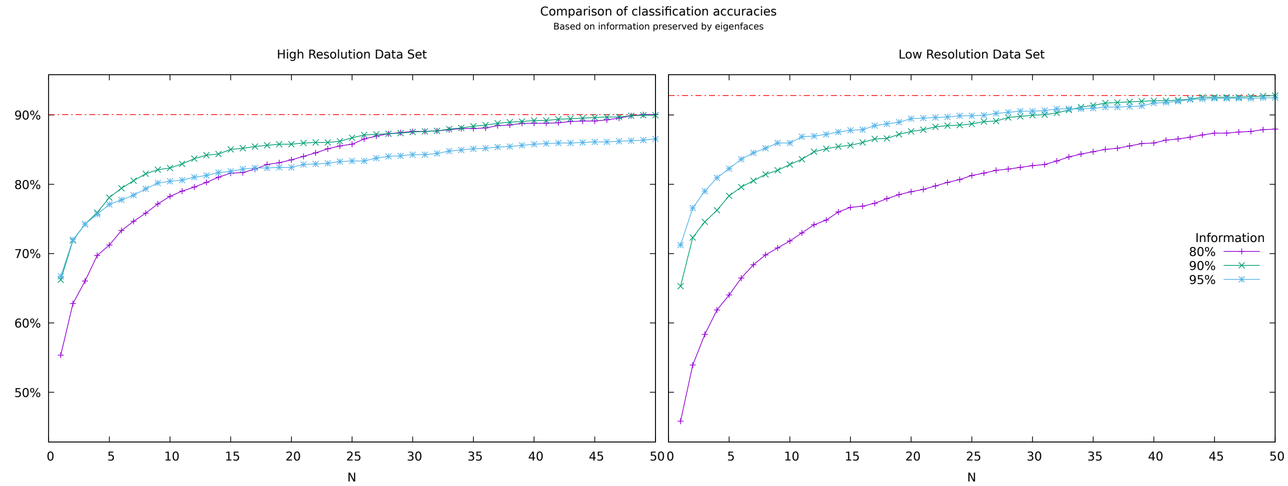

Figure 4: A comparison of classification accuracy against number of images considered (N), the percentage of information preserved by eigenfaces, and the data set used.

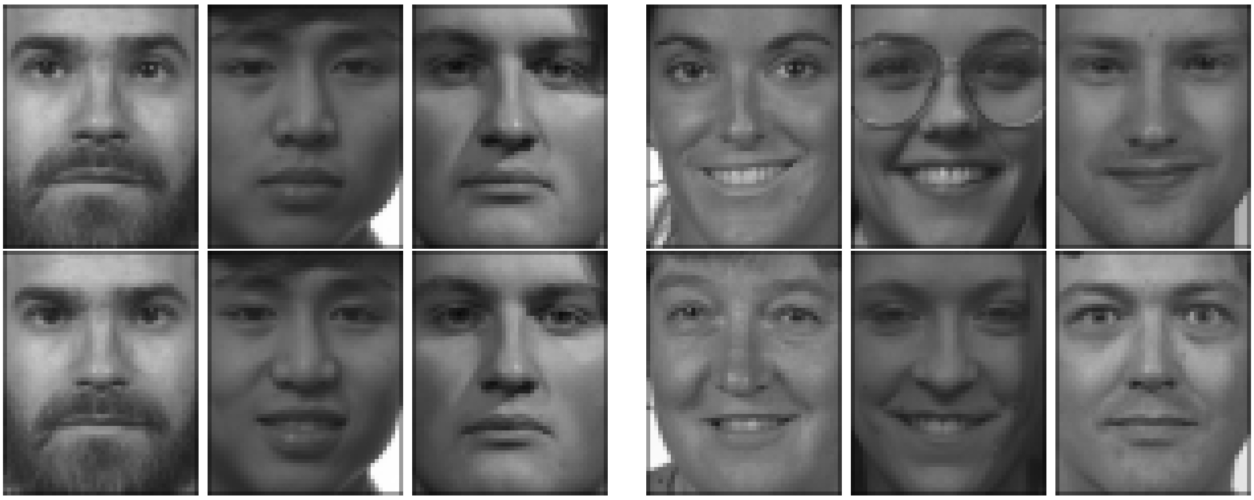

Figure 5: High resolution testing images (top) and the training images they were matched against (bottom). Images which were correctly classified (i.e. both pictures are of the same subject) on the left, and incorrect classifications on the right.

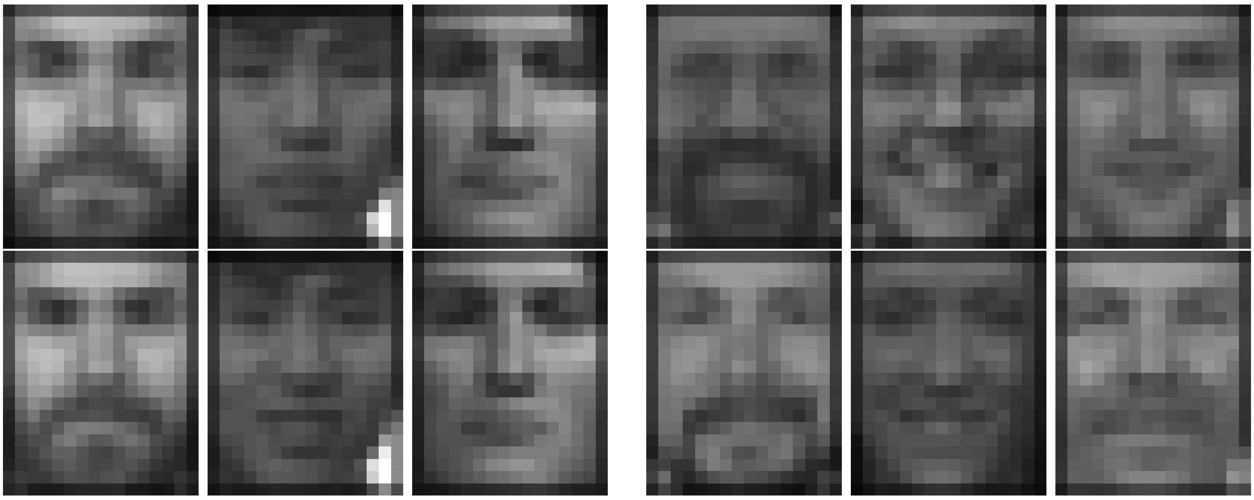

Figure 6: The same figure as fig. 5, but with images from the low resolution data set.

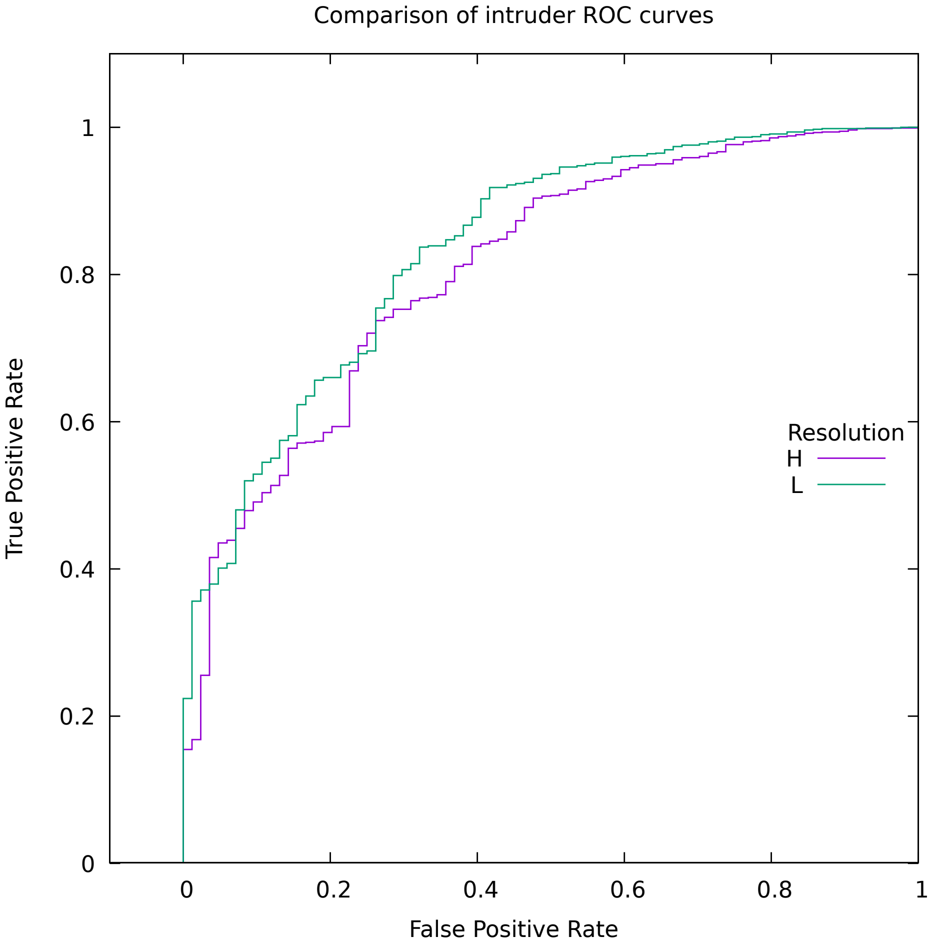

Figure 7: A comparison of intruder-detection ROC curves based on resolution.