Problem 4. (**SVD,** 20 points)

In this problem, we will see an application of SVD: eigenfaces. The dataset (stored in the file "allFaces.mat") contains 2410 images of faces from 38 persons. Each image is 192×168 (vectorized into a 32256-dimensional face vector). Using SVD, we will embed the high-dimensional data into a low-dimensional space, and perform a simple classification with the embedded data.

1) In the file "eigenfaces.mlx", we have loaded the dataset and plotted the average face. Now perform SVD on mean-subtracted data to get $\mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, where $\mathbf{M} \in \mathbb{R}^{32256\times 2410}$ and each column $\mathbf{m}_i$ is a image subtracting the average face, the main diagonal of $\mathbf{\Sigma}$ consists of $\sigma_1, \cdots, \sigma_{2410}$ with $\sigma_1 > \sigma_2 > \cdots > \sigma_{2410}$. The columns of $\mathbf{U}$ are called "eigenfaces". Plot the first 6 eigenfaces. (5 points)
2) In fact, as indicated by the magnitudes of singular values, most of the variance in the dataset is captured by the first few eigenfaces (also called the principle components). For example, $\sum_{i=1}^{150} \sigma_i > 0.5 \sum_{i=1}^{2410} \sigma_i$. Therefore, to compress the dataset, it is reasonable to embed the high-dimensional face data into a rather low-dimensional eigenface space. Let us take the first 50 eigenfaces, and compute the embedded dataset in this $50-$dimensional space. The coordinates of the embedding of image i is given by $\bar{\mathbf{m}}_i = [\langle \mathbf{m}_i, \mathbf{u}_1 \rangle, \cdots, \langle \mathbf{m}_i, \mathbf{u}_{50} \rangle]^T$. Then use the compressed dataset and eigenfaces to recover the average faces of the first 5 persons. Plot these 5 average faces. (5 points)
3) In the eigenface space, images of the same person tend to form a cluster, so we can do a simple classification using the average faces in the last problem. (To see this, you can try to plot the images of the first two persons on a 2-dimensional plane, using the 5th coordinate $\langle \mathbf{m}_i, \mathbf{u}_5 \rangle$ as x-axis and the 6th coordinate $\langle \mathbf{m}_i, \mathbf{u}_6 \rangle$ as y-axis. This plot does not count for points.)

   Specifically, randomly draw 100 samples of the images of the first 5 persons, and we want to classify which person do they belong to. To do this, compute their euclidean distances with the 5 average faces in the r-dimensional eigenface space. The estimated person would be the closest average face. Compute the accuracy rates with different r's and complete the following table. (10 points)

Table I

ACCURACY RATES

r	6	10	30	80	150
accuracy rate					