MATH444 HW1

Jiasen Zhang

1 Problem 1

1.1 (a)

There are totally 15 figures.

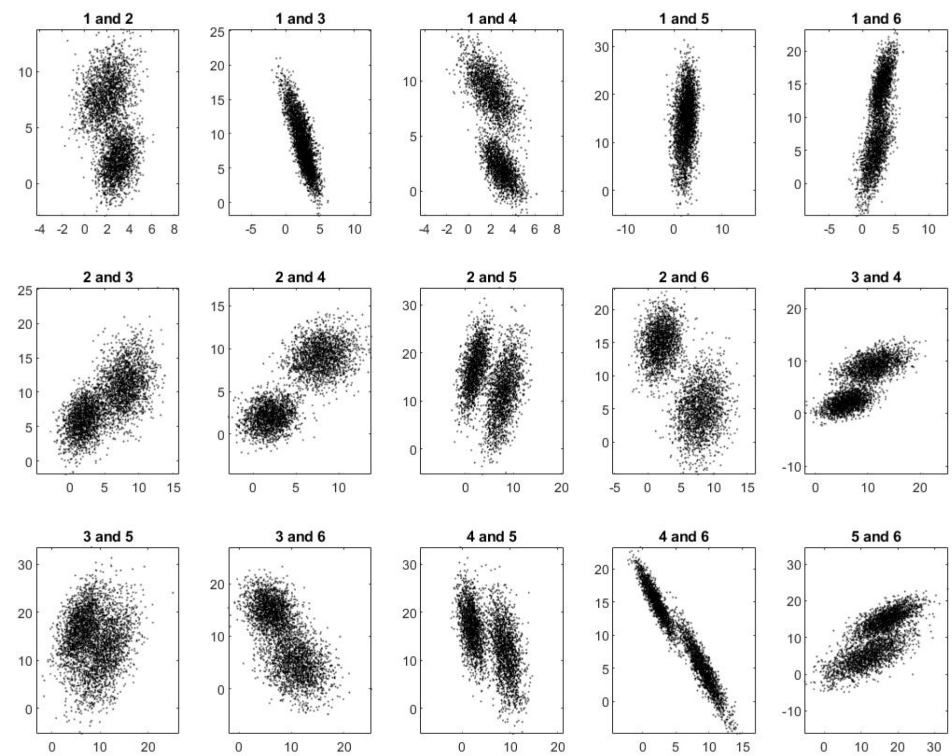

Figure 1: Visualizing the raw data selecting 2 principal components at a time.

1.2 (b)

Computed with Matlab, the singular values of the centered data X_c are:

$$\sigma_1 \approx 557.6798 \quad \sigma_2 \approx 321.0784 \quad \sigma_3 \approx 135.6906$$
$$\sigma_4 \approx 6.4063 \quad \sigma_5 \approx 6.3180 \quad \sigma_6 \approx 6.1953$$

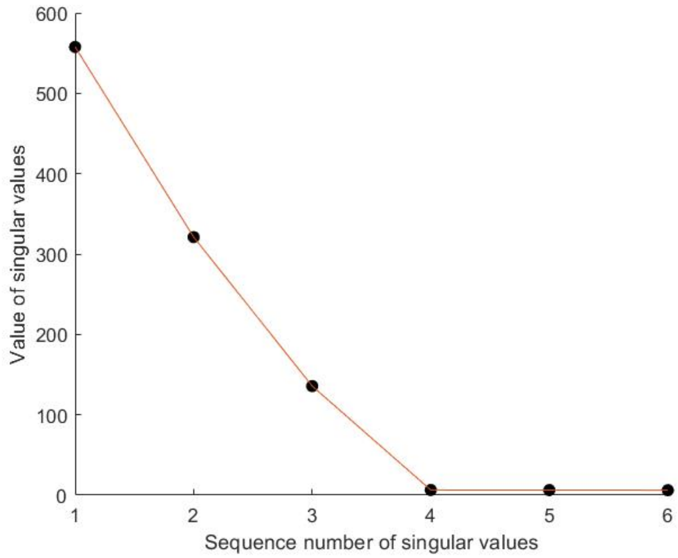

Figure 2: The singular values.

Obviously, σ_4, σ_5 and σ_6 are really close to zero compared with σ_1, σ_2 and σ_3, so we can say the effective dimensionality of the data is 3. I reduce the data dimension to 3 and visualize the reduced data with 3-dimensional scatter plot in Figure 3. We can see the 3D data has 2 clusters.

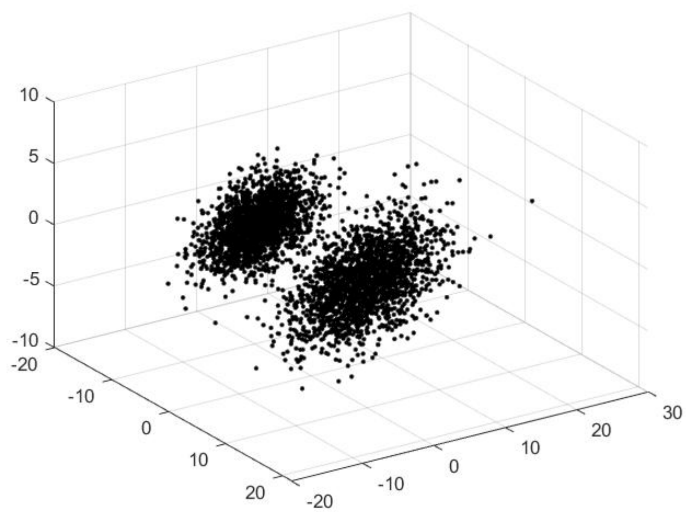

Figure 3: The first three principal components.

2 Problem 2

Centralize the data and calculate the SVD, the singular values are shown in Figure 4.

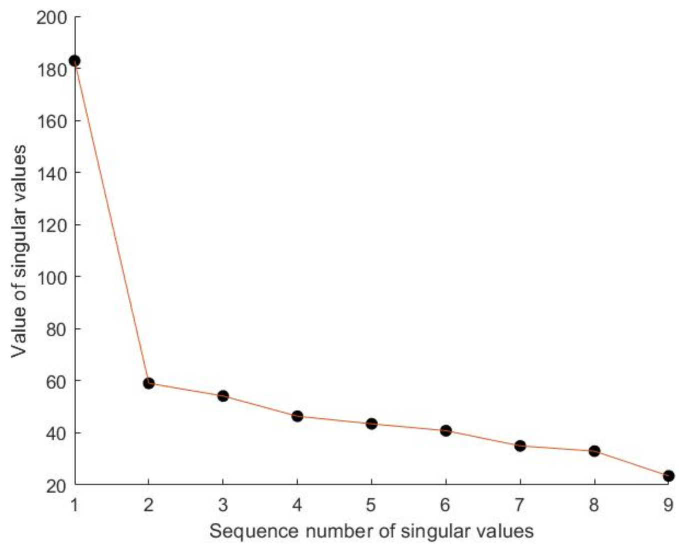

Figure 4: The singular values.

According to Figure 4, the first singular value σ_1 is much larger than other singular values, while the other singular values are close to each other. So we can plot the first two principal components. As shown in Figure 5, the points form a compact cluster on the right. It suggests that one of the two states, benign or malignant tumors, is generally located in the cluster.

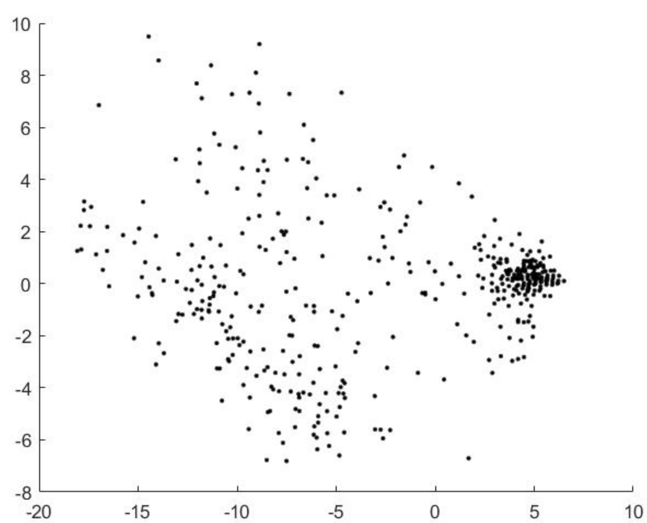

Figure 5: The first two principal components.

3 Problem 3

Centralize the data and calculate the SVD, the singular values are:

$$\sigma_1 \approx 25.0899 \qquad \sigma_2 \approx 6.0079 \qquad \sigma_3 \approx 3.4205 \qquad \sigma_4 \approx 1.8785$$

Choose and plot the first 2 principal components in Figure 6:

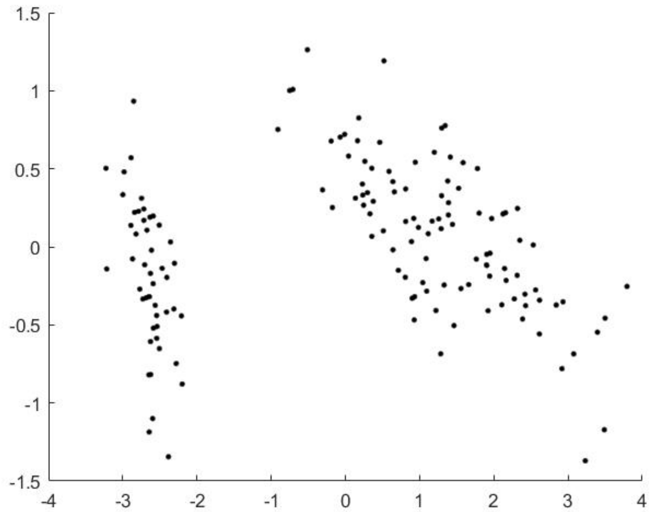

Figure 6: The first two principal components.

According to Figure 6, one cluster is seperated clearly on the left, meaning that it's easy to distinguish this kind of flower. The other two clusters stay close to each other, which means that this two kinds of flower are relatively more similar and can't be distinguished easily.

4 Problem 4

4.1 (a)

Set $r = 10$, choose 5 columns of X and X_r and plot them in Figure 7. The images are approximated with the first 10 feature vectors. As we can see, an approximation of rank 10 of X is enough to preserve these facial attributes.

Columns	22	23	97	125	139
Category	glasses	glasses	wink	leftlight	rightlight

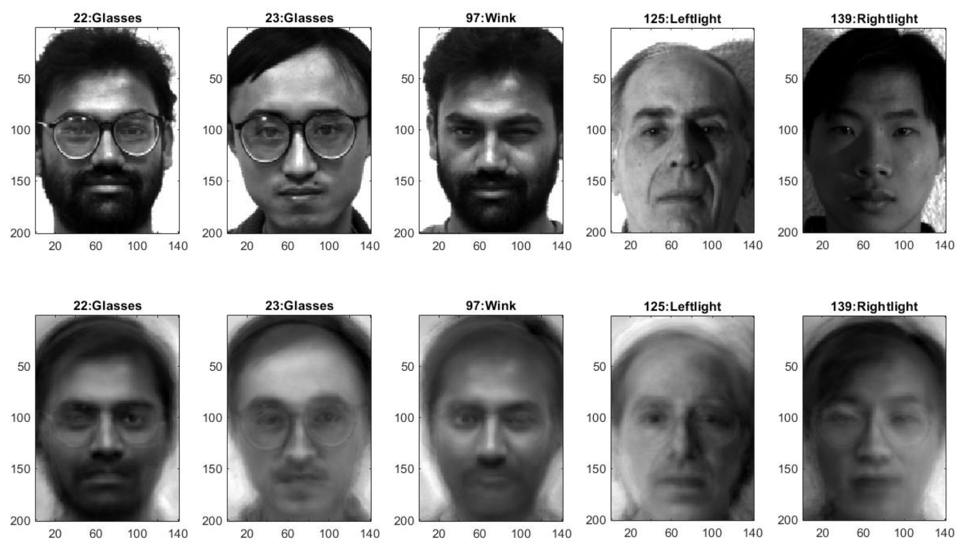

Figure 7: First row: original images. Second row: approximated images with $r = 10$.

4.2 (b)

Increasing r and plot the first r singular values in Figure 8, we can see the singular values decrease quickly at first. Then the speed of decreasing goes down and rise again in the last few singular values.

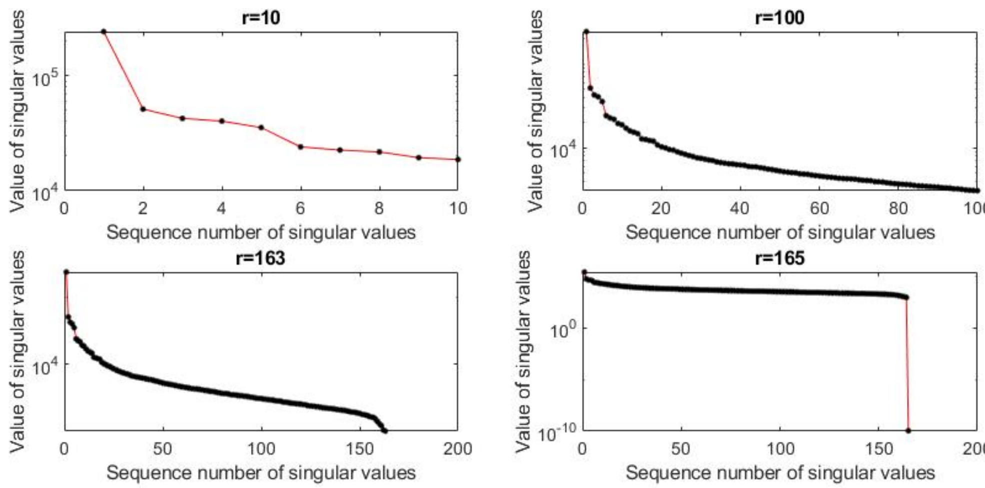

Figure 8: The first r singular values.

4.3 (c)

Plot the first 5 feature vectors in Figure 9(not centered) and Figure 10(centered).

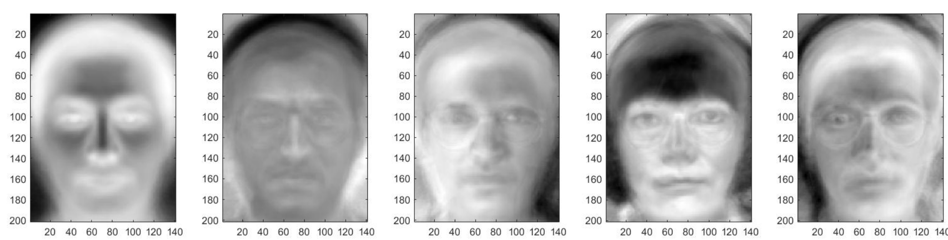

Figure 9: The first 5 feature vectors: not centered.

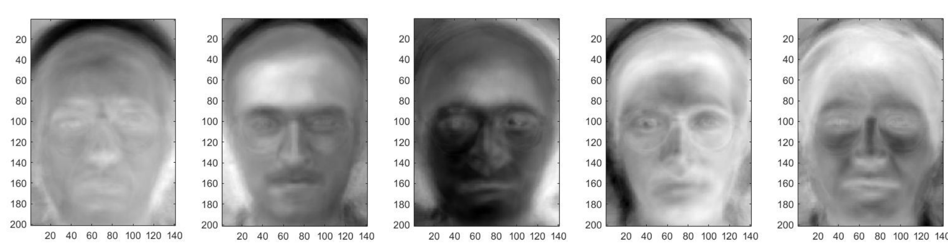

Figure 10: The first 5 feature vectors: centered.

4.4 (d)

The Figures 11-15 show the chosen 5 original images, their approximations and differences. With k increasing, the approximate images are clearer and more similar to the individuals in the corresponding original images. On the other hand, their differences are mroe and more blurred. Besides, when k is larger, the approximate images gradually have the attributes in the original images, such as glasses, lighting and facial expressions.

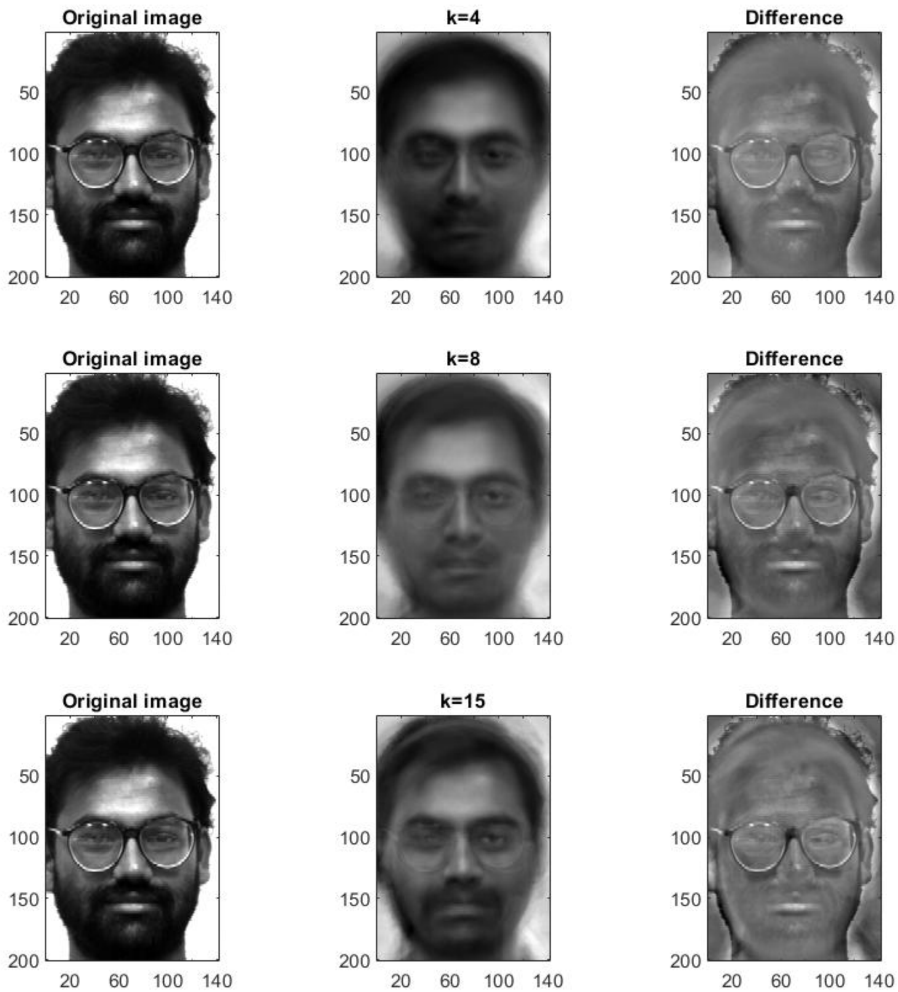

Figure 11: The image 22: glasses.

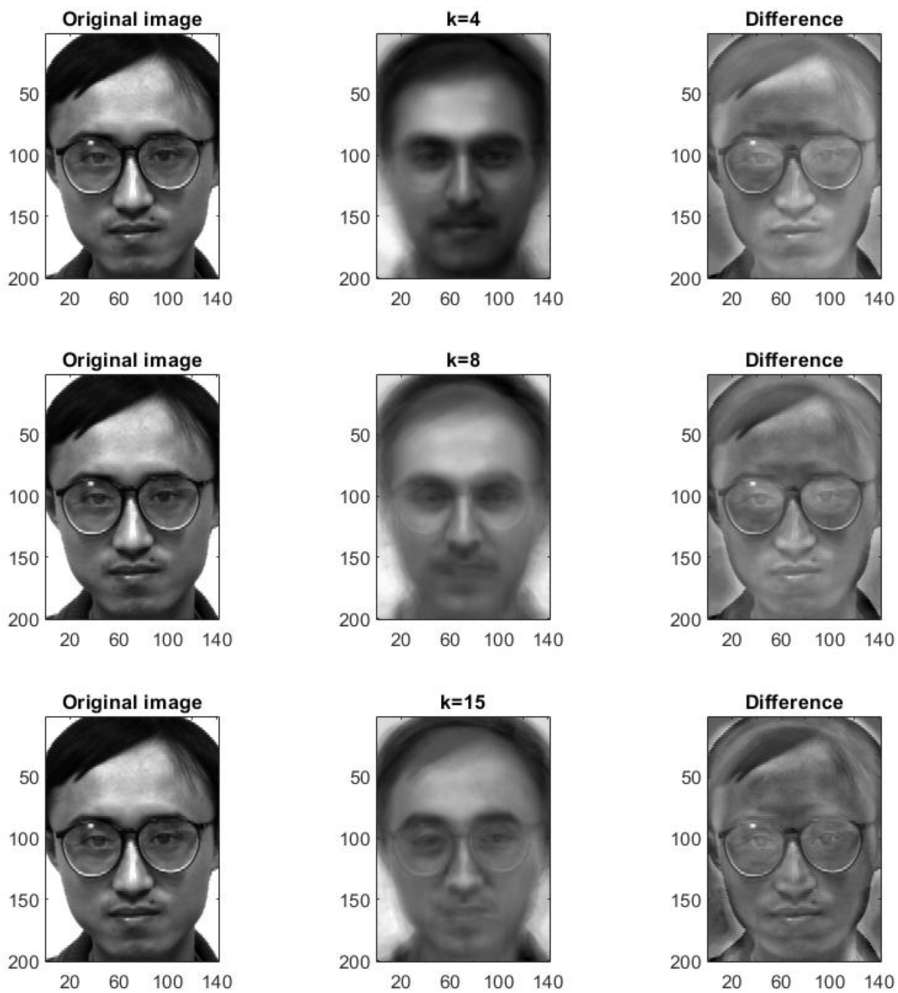

Figure 12: The image 23: glasses.

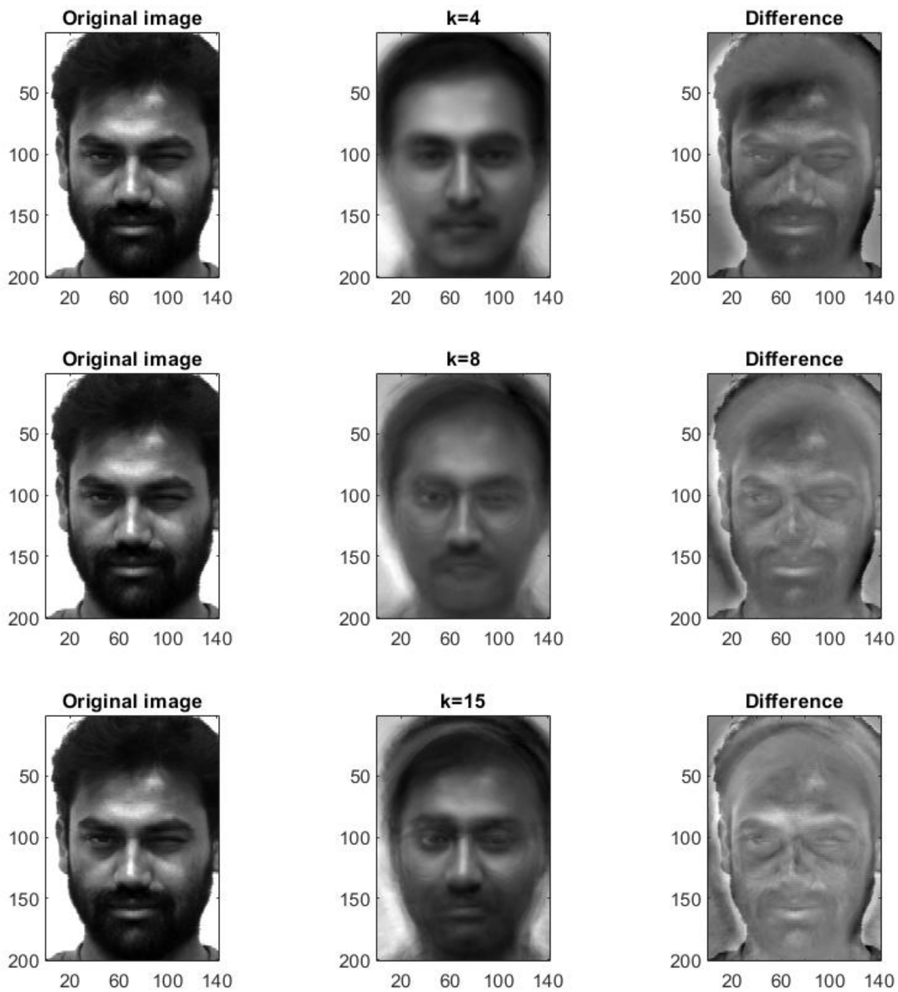

Figure 13: The image 97: wink.

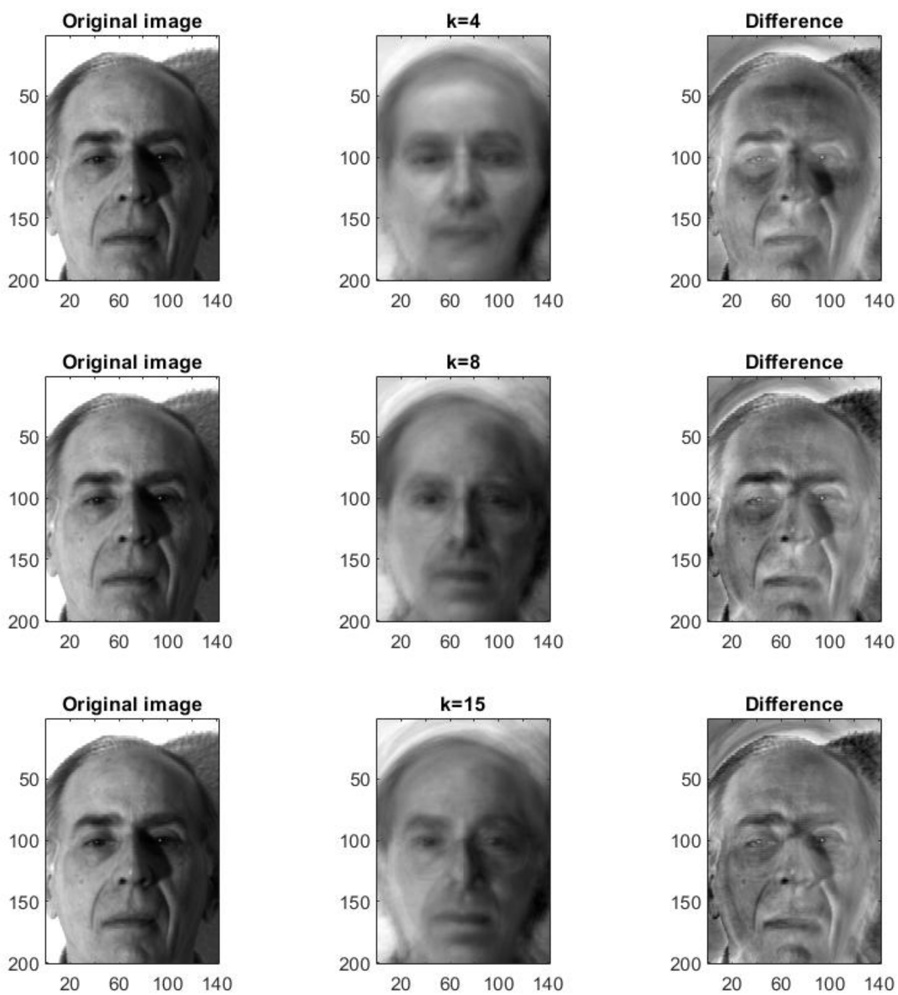

Figure 14: The image 125: leftlight.

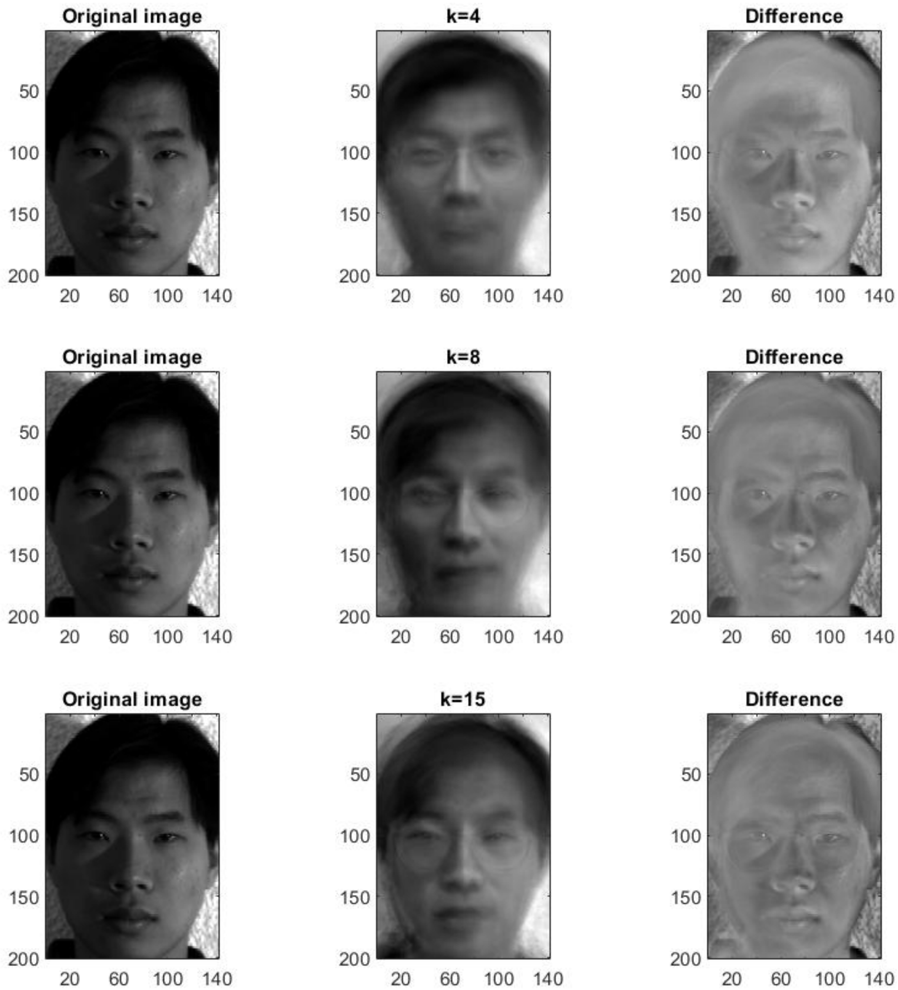

Figure 15: The image 139: rightlight.

5 Matlab Code

5.1 Problem 1

```
clear all;clc;

% Ploting the raw data
load ModelReductionData;
k=1;
figure(1);
for i=1:5
    for j=i+1:6
        subplot(3,5,k);
        plot(X(i,:),X(j,:),'k.','MarkerSize',1);
        axis equal;
        %set(gca,'FontSize',10);
        title(strcat(num2str(i),{32},'and',{32},num2str(j)));
        k=k+1;
    end
end

% Centering the data and compute SVD
Xbar = mean(X,2);
Xc = X - Xbar;
[u,d,v] = svd(Xc);
sigma=diag(d);
% Plot the singular values
figure(2);
scatter([1:6],sigma,'k.','SizeData',500);hold on;
plot(sigma);
ylabel('Value of singular values');
xlabel('Sequence number of singular values');

% Reduce the data dimension
z = u(:,1:3)'*Xc;
figure(3);
scatter3(z(1,:),z(2,:),z(3,:),'k.','SizeData',50);
```

5.2 Problem 2

```
clear all;clc;
load BiopsyData;
% Remove the columns from the data containing missing data
sumColumn=sum(X,1);
aa=find(¬isnan(sumColumn)); % index of nonmissing data
X = X(:,aa); % the data without missing
% Center the data
Xbar = mean(X,2);
Xc = X - Xbar;
[u,d,v] = svd(Xc);
sigma=diag(d)
scatter([1:length(sigma)],sigma,'k.','SizeData',500);hold on;
plot(sigma);
ylabel('Value of singular values');
xlabel('Sequence number of singular values');
% Reduce the data dimension
z = u'*Xc;
figure();
scatter(z(1,:),z(2,:),'k.','SizeData',50);
```

5.3 Problem 3

```
clear all;clc;
load IrisData;
% Centering the data and compute SVD
Xbar = mean(X,2);
Xc = X - Xbar;
[u,d,v] = svd(Xc);
sigma=diag(d)

z = u(:,1:4)'*Xc;
figure();
scatter(z(1,:),z(2,:),'k.','SizeData',80);
```

5.4 Problem 4(a)-(c)

```
clear all;clc;
load AlignedYaleFaces
Xbar = mean(X,2);
Xc = X - Xbar;

%(a) non-centered data is better here
% get X_r
r = 10;
[U,D,V] = svds(X,r);
sigma=diag(D);
X_r = U*D*V';
% choose 5 columns and plot
chosen=[22;23;97;125;139];
for i=1:5
    j=chosen(i);
    subplot(2,5,i);
    imagesc(reshape(X(:,j),ImageSize));
    axis equal;axis tight;colormap(gray);
    title(strcat(num2str(j),': ',AnnotationCode{I(j)}));
end
for i=1:5
    j=chosen(i);
    subplot(2,5,i+5);
    imagesc(reshape(X_r(:,j),ImageSize));
    axis equal;axis tight;colormap(gray);
    title(strcat(num2str(j),': ',AnnotationCode{I(j)}));
end

%(b)
figure();
n=[10,100,163,165];
for i=1:length(n)
    r = n(i);
    [¬,D,¬] = svds(X,r);
    sigma=diag(D);
    %X_r = U*D*V';
    subplot(2,2,i);
    semilogy(sigma,'red');hold on;
    scatter([1:length(sigma)],sigma,'k.','SizeData',70);
    ylabel('Value of singular values');
    xlabel('Sequence number of singular values');
    title(strcat('r= ',num2str(r)));
end

%(c)
r = 165;
[U,D,V] = svds(Xc,r); % using centered data
```

```
figure();
for i=1:5
    subplot(1,5,i);imagesc(reshape(U(:,i),ImageSize));
    axis equal;axis tight;colormap(gray);
end
```

5.5 Problem 4(d)

```
clear all;clc;
load AlignedYaleFaces
r = 165;
[U,D,V] = svds(X,r);

n=22; % columns: 22, 23, 97, 125, 139

k0=[4,8,15]; % number of feature vectors
for i=1:3
    k=k0(i);
    X0=U(:,1:k)*D(1:k,:)*V(:,:)';
    dX=X-X0;
    subplot(3,3,1+3*(i-1));
    imagesc(reshape(X(:,n),ImageSize));
    axis equal;axis tight;colormap(gray);
    title('Original image')
    subplot(3,3,2+3*(i-1));
    imagesc(reshape(X0(:,n),ImageSize));
    axis equal;axis tight;colormap(gray);
    title(strcat('k= ',num2str(k)));
    subplot(3,3,3+3*(i-1));
    imagesc(reshape(dX(:,n),ImageSize));
    title('Difference')
    axis equal;axis tight;colormap(gray);
end
```