COMPUTER VISION

IV. Eigensystem and Principal Component Analysis

This project allows to implement and understand the merits of the PCA on multispectral images. This document shows the results of applying PCA algorithm over multispectral satellite Images. In particular, I selected an 8-band image (i.e. coastal, blue, green, yellow, red, red edge, NIR1 and NIR2); see below their respective wavelength intervals, in which we can see different geographic properties like a river, forest (trees), building structures, green space, etc. This image is 16-bit depth and its location is Adelaide, Australia. Please see below a representation of the RGB channels of the 8-band image.

Spectral bands considered include:

Coastal band [400–450 nm],
Blue [450–510 nm],
Green (510–580 nm)
Yellow [585–625 nm],
Red [630–690 nm]
Red-edge [705–745 nm]
Near-infrared 1, NIR1 [770–895 nm]
Near-infrared 2, NIR2; [860–1040 nm]

Given these 8-band image, I have constructed a matrix of $R X 8$, where $R = M \cdot N$ and M and N are the corresponding width and height of the input image I, with 8 representing the 8 dimensions or wavebands. Hence the rows of this matrix correspond to observations and columns correspond to variables.

All the below tasks were implemented using Matlab :

- Computed the covariance matrix
- Determined the eigenvalues
- And determined the eigenvectors (principal components) of the covariance matrix.
- Through the inverse PCA,

(a) Reconstructing the original image by keeping each of the Principal Component (PC) starting from the first to the eight PC.
(b) Reconstructing the input image from the eigenvectors of the highest 4 eigenvalues
(c) Reconstructing the input image from the eigenvectors of the lowest 4 eigenvalues
(d) Subtracting the original image from the reconstructed image you obtained in (b)
(e) Subtracting the original image from the reconstructed image you obtained in (c)

INTRODUCITON:

An Eigensystem is a set of Eigenvectors and their corresponding Eigenvalues for a square matrix. In linear algebra, Eigenvectors are the non-zero vectors that do not change their direction when a linear transformation is applied to them. Eigenvalues are scalar values that represent how much an Eigenvector is scaled during the linear transformation.

PCA is a statistical method used to transform a dataset into a new coordinate system. It is based on the Eigenvectors and Eigenvalues of the covariance matrix of the dataset. The covariance matrix describes the relationship between variables in the dataset. In PCA, the Eigenvectors of the covariance matrix are computed, and they represent the directions in the data that contain the most variation. The corresponding Eigenvalues indicate the amount of variation in each direction.

PCA for multispectral satellite image analysis to reduce the dimensionality of the image data while retaining the most important information. In multispectral satellite images, each pixel is represented by a vector of values corresponding to the intensity of the pixel in each spectral band. PCA can be applied to these pixel vectors to extract the most significant spectral features or components that capture the majority of the variance in the data.

PCA can be used to perform data compression on multispectral satellite images by reducing the number of bands required to represent the image while retaining the important information. The compressed image can be reconstructed using only a subset of the principal components.

PROCEDURE:

The covariance matrix is a central concept in PCA because it is used to compute the eigenvectors and eigenvalues of the dataset. The eigenvectors of the covariance matrix correspond to the principal components of the dataset, which are the directions of maximum variance in the data. The eigenvalues represent the variance of the data along each of these principal components. The larger the eigenvalue, the more important the corresponding eigenvector is in explaining the variance in the data.

By computing the eigenvectors and eigenvalues of the covariance matrix, PCA identifies the principal components of the dataset and orders them according to their importance in explaining the variance in the data. The first principal component corresponds to the direction of maximum variance in the data, and each subsequent principal component corresponds to the direction of maximum variance that is orthogonal to the previous components.

In the case of multispectral satellite images, the covariance matrix is computed over the entire image to capture the relationships between the different spectral bands. PCA can then be applied to the covariance matrix to identify the principal components that explain the majority of the variance in the image. These principal components can then be used to reduce the dimensionality of the image, enhance features of interest, and compress the image data for more efficient storage and transmission.

FORMULAS:

The eigensystem and covariance matrix are both important concepts in Principal Component Analysis (PCA).

The formula for the eigensystem of a matrix A is:

$A v = \lambda v$

where A is a square matrix, v is a non-zero vector, λ is a scalar (the eigenvalue), and the equation is satisfied when A is multiplied by v. In other words, the eigenvector v is a direction in which A only stretches or compresses the vector, and the eigenvalue λ is the factor by which the vector is stretched or compressed in that direction.

In PCA, the covariance matrix C is calculated for a dataset X with n observations and p variables (features) as:

$C = (X - m) (X - m)T / (n - 1)$

where m is the mean of the dataset. The covariance matrix describes how the variables in the dataset are related to each other, and its diagonal elements represent the variances of each variable. The eigenvectors and eigenvalues of the covariance matrix are used to identify the principal components (PCs) of the data, which are the directions in which the data varies the most. The eigenvectors represent the directions of the PCs, and the eigenvalues represent the amount of variation explained by each PC. The PCs are sorted in decreasing order of their eigenvalues, so that the first PC captures the most variation in the data, the second PC captures the second most variation, and so on.

Once the eigenvectors and eigenvalues of the covariance matrix are calculated, the dataset can be projected onto the PCs by multiplying the data matrix X by the eigenvectors:

$Y = X V$

where V is a matrix of the eigenvectors, sorted by decreasing eigenvalue. The resulting matrix Y contains the scores of each observation (row of X) along each PC (column of V), and represents a new coordinate system in which the PCs are the axes.

INVERSE PCA:

Inverse PCA is the process of reconstructing the original dataset from a reduced set of principal components (PCs) obtained through PCA. This is done by multiplying the reduced PCs with the corresponding eigenvectors and adding the mean of the original dataset. The resulting reconstructed dataset will have a lower dimensionality than the original dataset, but should still be representative of the original dataset.

FORMULA FOR INVERSE PCA:

the inverse PCA formula can be written as:

$X = \mu + \Phi * Y$

where,

- X is the reconstructed data matrix of size n x d, where n is the number of observations and d is the number of variables
- μ is the mean of the original data matrix of size 1 x d
- Φ is the matrix of size d x k, where k is the number of principal components considered, containing the eigenvectors of the covariance matrix of the original data matrix sorted in descending order of their corresponding eigenvalues
- Y is the matrix of size n x k, containing the reduced representation of the original data matrix obtained by projecting it onto the first k principal components

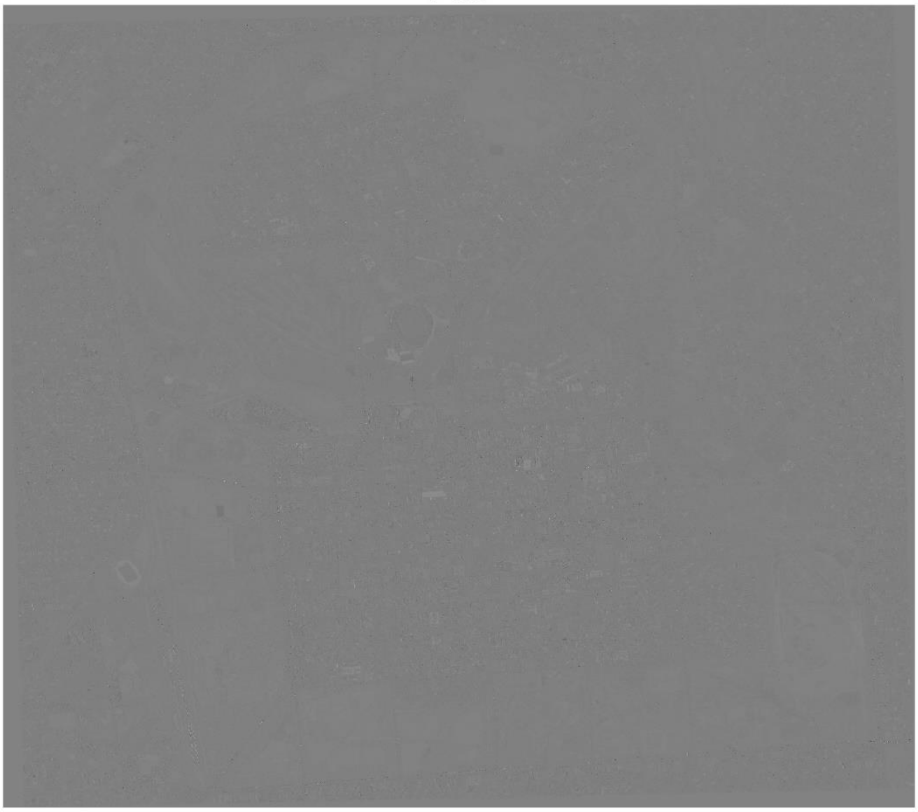
To reconstruct the original data matrix, we first project the reduced representation Y onto the eigenvectors Φ, and then add back the mean vector μ to obtain X. The reconstructed data matrix X will have the same dimension as the original data matrix.

RESULTS:

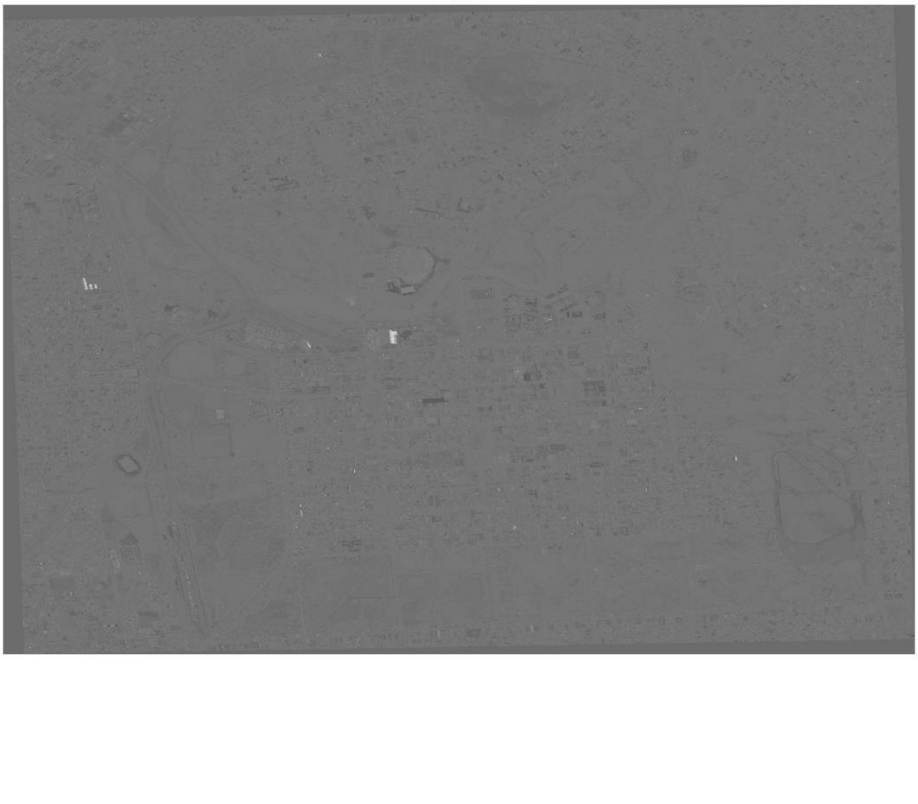
PC1

PC2

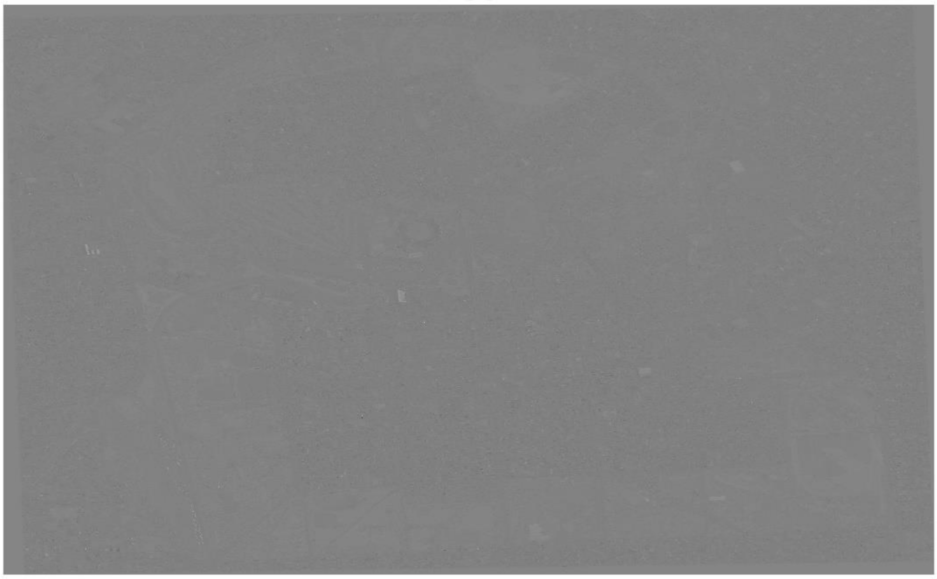
PC3



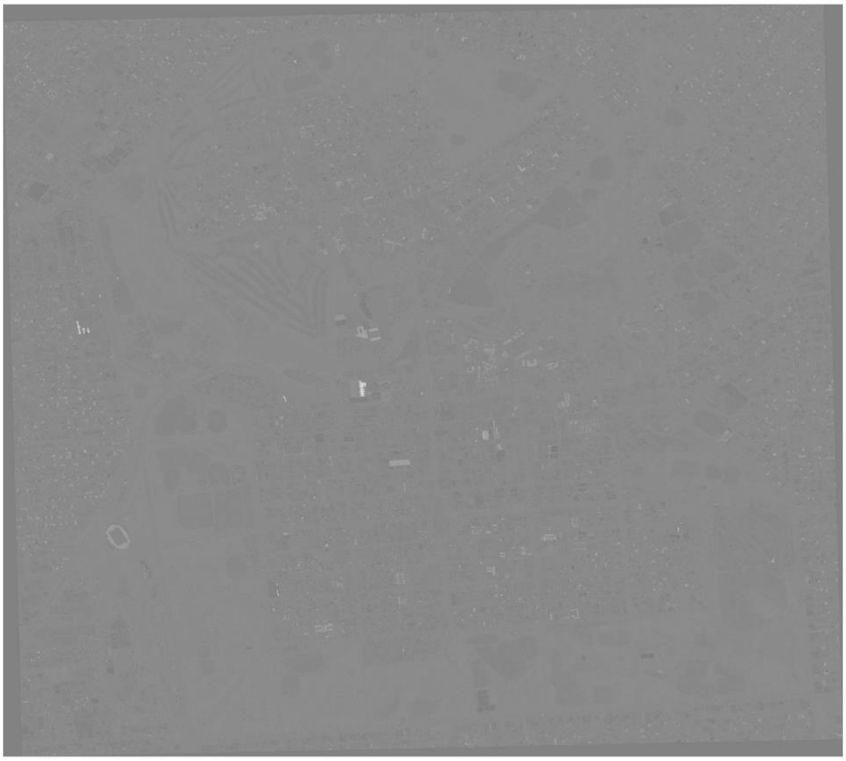
PC4



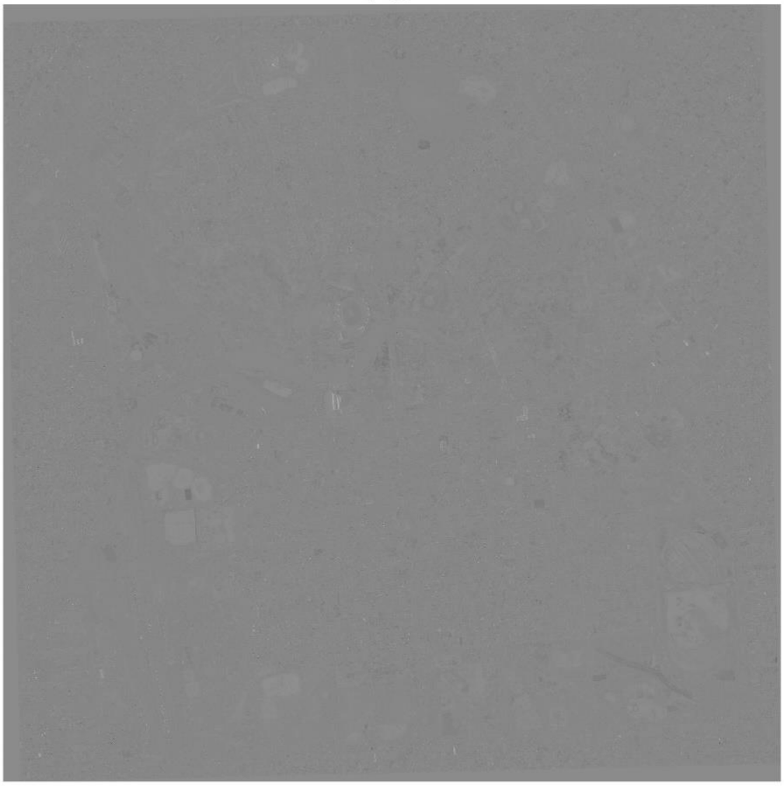
PC5



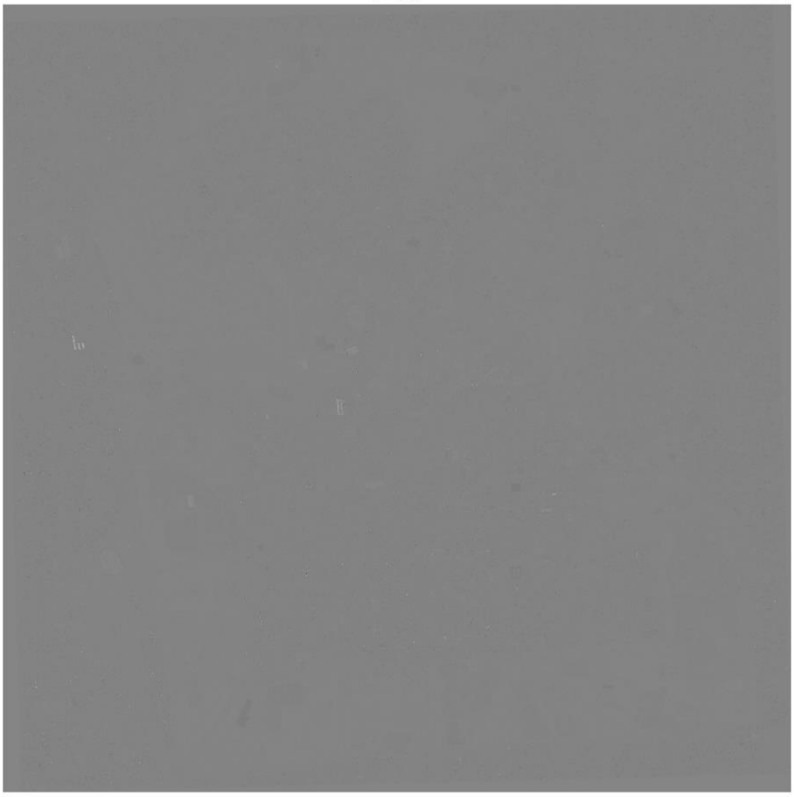
PC6



PC7



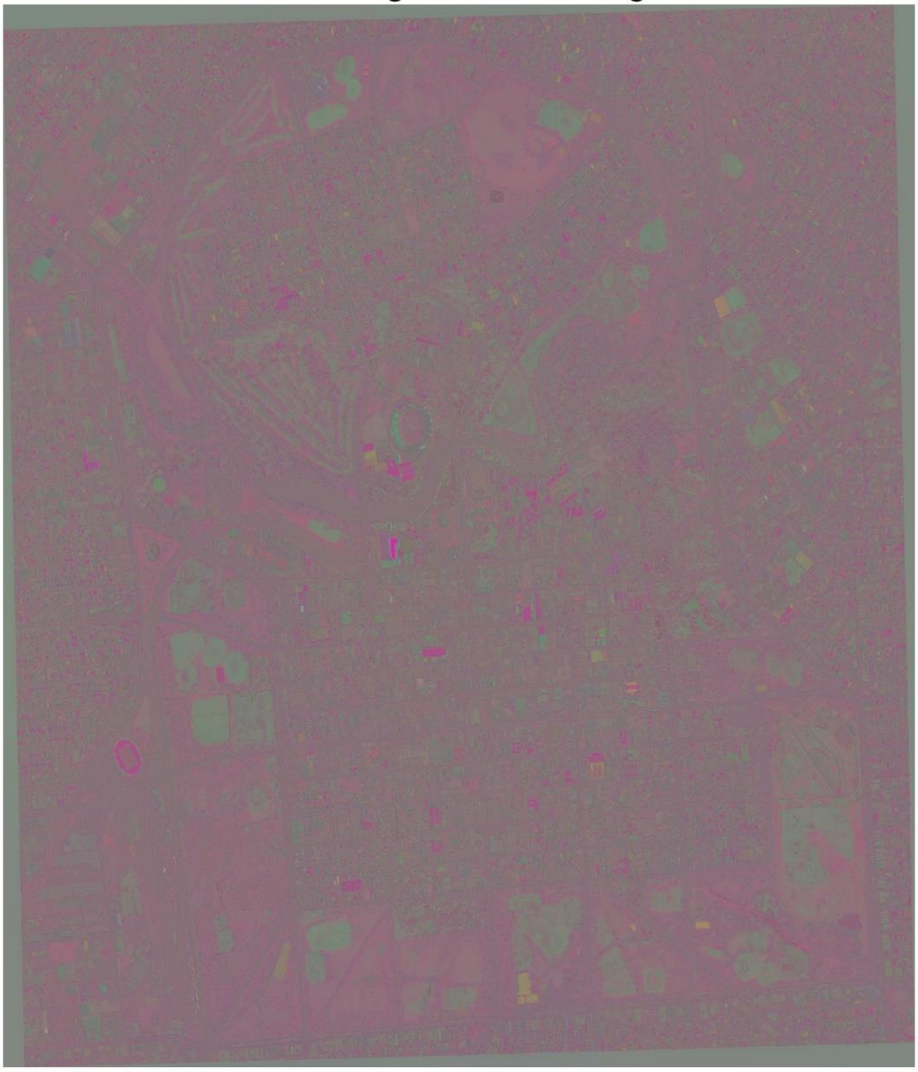
PC8



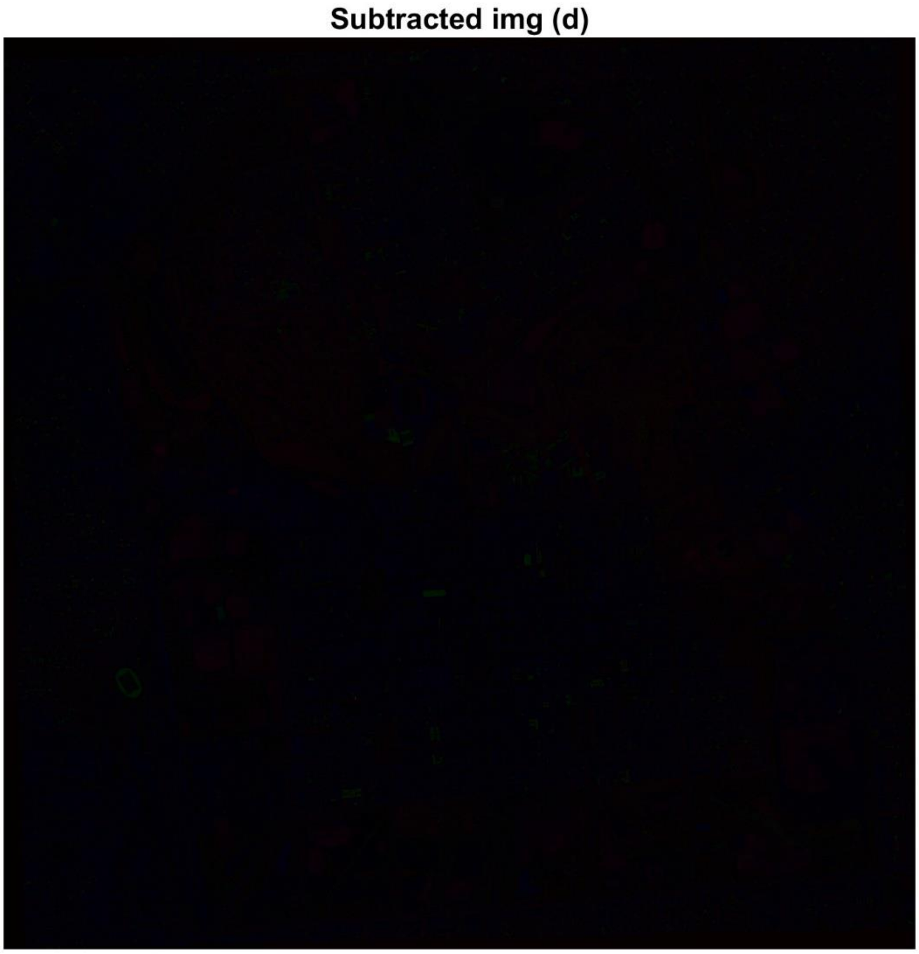
Fully reconstruced image

Reconstructed img with highest 4 eigen values

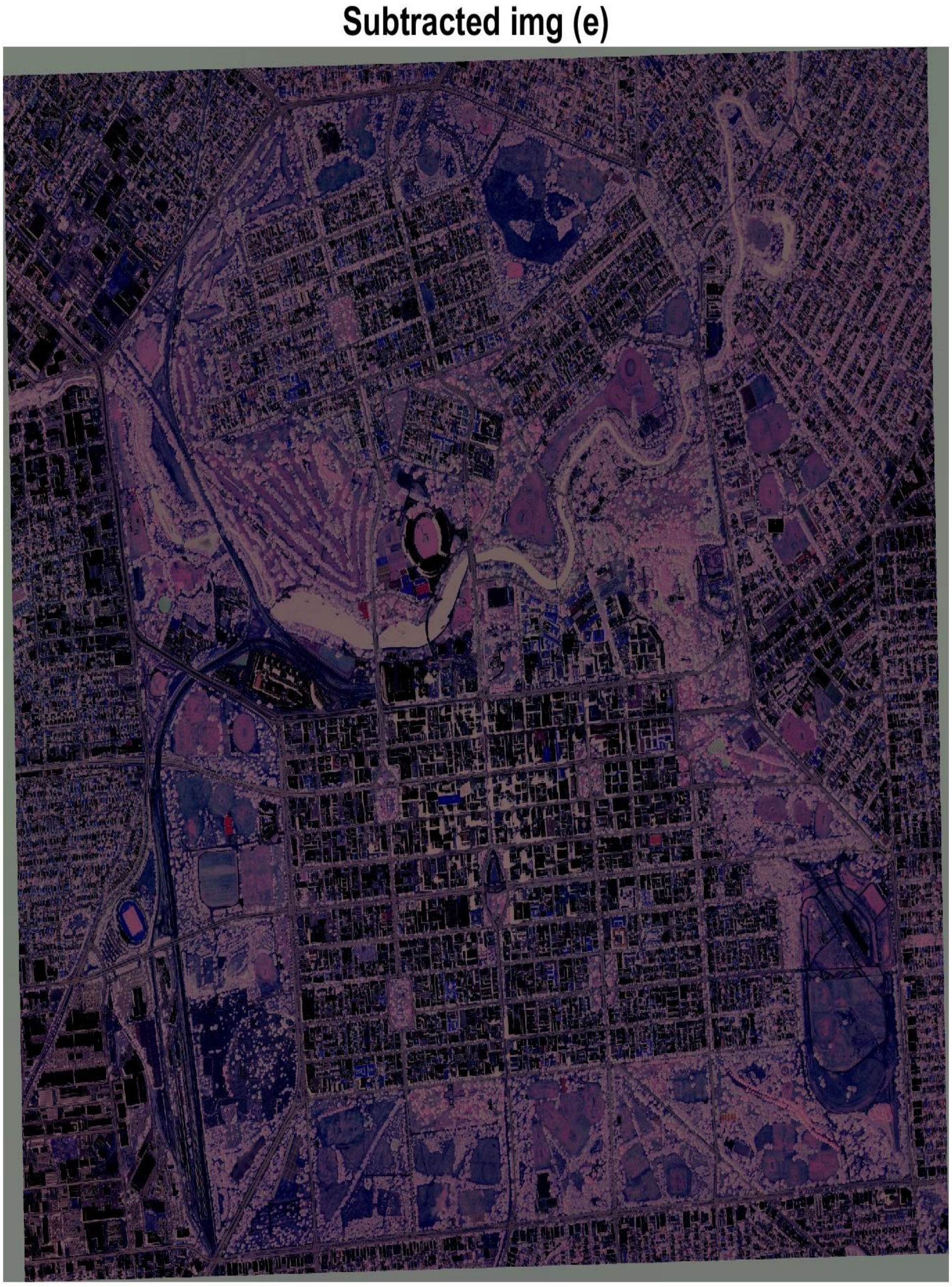
Reconstructed img with lowest 4 eigen values



(d) Subtracting the original image from the reconstructed image obtained from the highest 4 eigen values



(e) Subtracting the original image from the reconstructed image obtained from the lowest 4 eigen values



Covariance Matrix Obtained:

```
>> untitled
   1.0e+05 *

    0.1399    0.1811    0.2599    0.3025    0.2898    0.2167    0.1883    0.1282
    0.1811    0.2560    0.3721    0.4195    0.4197    0.2967    0.2614    0.1688
    0.2599    0.3721    0.5621    0.6377    0.6384    0.4770    0.4646    0.2967
    0.3025    0.4195    0.6377    0.7749    0.7509    0.5681    0.4972    0.3310
    0.2898    0.4197    0.6384    0.7509    0.7700    0.5405    0.4721    0.2973
    0.2167    0.2967    0.4770    0.5681    0.5405    0.5861    0.7781    0.5240
    0.1883    0.2614    0.4646    0.4972    0.4721    0.7781    1.4455    0.9426
    0.1282    0.1688    0.2967    0.3310    0.2973    0.5240    0.9426    0.6597
```

DISCUSSIONS & CONCLUSIONS:

We observe clearly that the first two eigenvectors make up about 97% of the entire information. This demonstrates how closely similar the original image I is and how just two dimensions may account for 97% of the data. Additionally, by using these two orthogonal, uncorrelated eigenvectors, we should be able to segment the image more accurately.

Depending on these values of the eigen vectors the subtracted images clearly says that the highest eigen values image get subtracted by the highest eigen values then the image is full black.