HERIOT WATT UNIVERSITY

B31SE Image Processing: Assignment 4

Hugo Millet
04/04/2023
B31SE: Image Processing
MSc. Robotics
Electronical Electronic and Computer Engineering

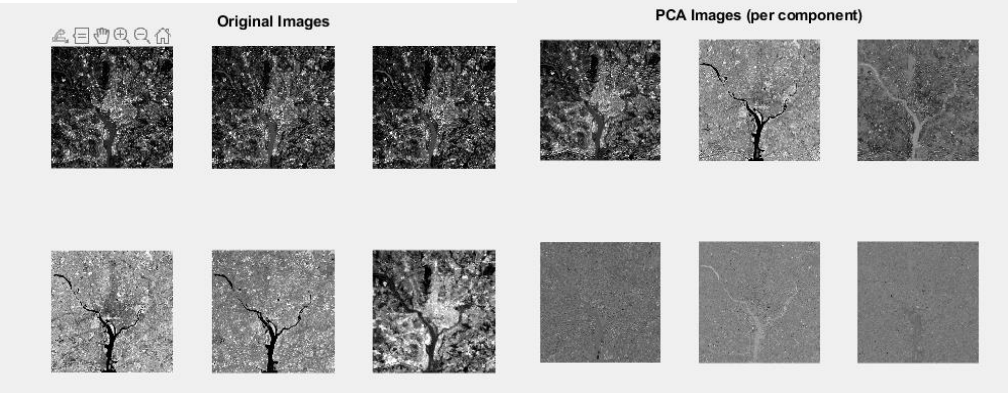
1)

In data science, PCA is a statistical technique used to reduce dataset dimensionality while preserving as much information as possible. PCA identifies the most important features or variables in a dataset and transforms them into principal components, which capture the most variability in the data.

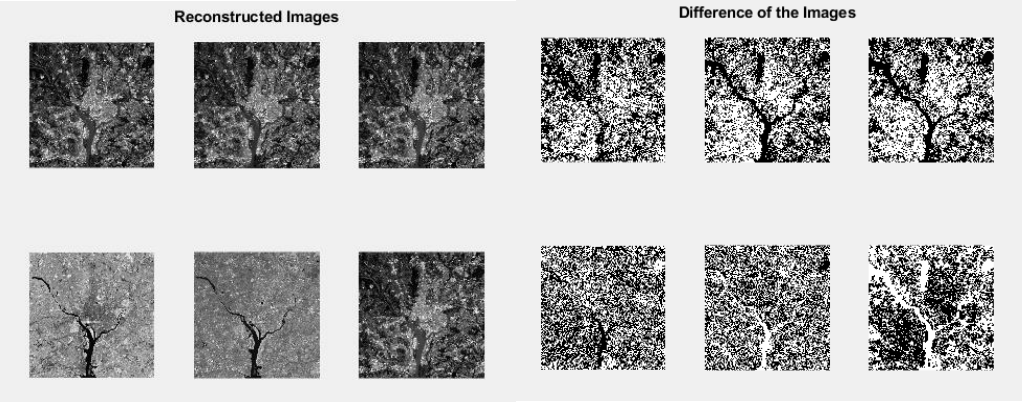
Therefore, PCA is useful for several purposes:

- Identify which features of the dataset/image contribute the most to the total information.
- Obtain the most similar image with fewer components, being able to reduce the size of the image.

2) By executing the main_pca.m MATLAB program, we obtain the following results :



As the two eigenvalues for the 2 best principal components are significantly higher than the other 4, the two first PCA images have the most contrast because most of the variance of the original images are contained within these two images.



It is also because the two best principal components (used for the reconstruction) encapsulate most of the image variance (and therefore information), that the reconstructed image are very similar to the original images. Only the sixth image is pretty different, being due to its original blurriness, being lost in the reconstruction because the two PCA images are pretty sharp.

3) The change in the MATLAB code to obtain reconstructed images using different numbers of principal components are regarding this line :

```
Ak = A(1:2,:);
```

Ak is a truncation of the original A matrix containing the eigenvectors of the covariance matrix. Here, it is truncated to keep only the two first eigenvectors, corresponding to the two best principal components (eigenvalues). To change how many principal components we keep, we can change the 2 into any number from 1 (keep only the first PC) to 6 (keep all PC)

Here are the MSE and PSNR results with reconstructed image using :

- 2 best principal components :

Variable1	values1	Variable2	values2
{'Error 1'}	150.6	{'PSNR 1'}	26.353
{'Error 2'}	159.42	{'PSNR 2'}	26.105
{'Error 3'}	173.45	{'PSNR 3'}	25.739
{'Error 4'}	68.672	{'PSNR 4'}	29.763
{'Error 5'}	117.13	{'PSNR 5'}	27.444
{'Error 6'}	1060.1	{'PSNR 6'}	17.877

- 3 best principal components :

Variable1	values1	Variable2	values2
{'Error 1'}	76.424	{'PSNR 1'}	29.299
{'Error 2'}	32.383	{'PSNR 2'}	33.028
{'Error 3'}	34.499	{'PSNR 3'}	32.753
{'Error 4'}	66.071	{'PSNR 4'}	29.931
{'Error 5'}	116.75	{'PSNR 5'}	27.458
{'Error 6'}	2.64	{'PSNR 6'}	43.915

- 4 best principal components :

Variable1	values1	Variable2	values2
{'Error 1'}	57.301	{'PSNR 1'}	30.549
{'Error 2'}	28.444	{'PSNR 2'}	33.591
{'Error 3'}	34.065	{'PSNR 3'}	32.808
{'Error 4'}	1.7288	{'PSNR 4'}	45.753
{'Error 5'}	3.1291	{'PSNR 5'}	43.177
{'Error 6'}	0.64597	{'PSNR 6'}	50.029

When we use more principal components to reconstruct the image, we use more of the original information carried by the image, therefore by using most principal components, errors generally decrease and PSNR increase.

In addition, the best two principal components are responsible for 88.5% of the variance of the image, while the best three 97.8% (9.3% increase) and the best four 99.1% (1.3% increase). Therefore, the gain in the PSNR and the decrease in the error is more significant between the two-PC and three-PC cases than between the three-PC cases and four-PC cases : For image 1 between two-PC case and three-PC case there is a decrease of 74.2 in the error and an increase of 2.9 in the PSNR, while between the three-PC case and the four-PC case there is a decrease of 19.1 in the error and an increase of 1.25 in the PSNR.

We can also note the particular case of the sixth image, for which error drops significantly between the two-PC case (1060.1) and the three-PC case (2.64). A possible explanation is that the blurriness that was lost in the reconstruction with two PC, causing most of the image error, was mostly contained within the information of the third PCA image, causing the reconstruction with three PC being very close to the original one.