1. Introduction

I used Principal Component Analysis (PCA) in this homework on a group of cat photos from the Animal Faces dataset. The images were examined, downsized to 64x64 pixels, and divided into colour channels. Afterward, for each colour channel, I determined the principal components and their corresponding Proportion of Variance Explained (PVE). Ultimately, I reconstructed the photos utilizing distinct principal component counts and k-values.

2. Results

Question 1

The outputs of the question are below;

Red Channel:

[0.52315552, 0.16559529, 0.07076842, 0.06373423, 0.03960237, 0.03576077, 0.03127874, 0.02507971, 0.02451553, 0.02050943]

Green Channel:

[0.37955646, 0.20725056, 0.14807167, 0.08023778, 0.05459314, 0.03799413, 0.02886076, 0.02283481, 0.02193965, 0.01866105]

Blue Channel:

[0.42281126, 0.23243092, 0.12893025, 0.05082714, 0.04411396, 0.03246073, 0.02812732, 0.0241403, 0.02056126, 0.01559686]

All channel's PVE values represent the percentage of the data's overall variation which can be handled by each of the first 10 main components. The primary principal component accounts for 52.31% of the variation in the red channel, while the second and third principal components each account for 16.56% and 7.08% of the variance. Identical trends can be seen in the PVE values for the green and blue channels, with a large percentage of the overall variance being explained by the first five principal components. This finding shows that the majority of the significant information in the data is captured by the first few principal components. In contrast, the remaining components only account for a small percentage of the overall variance. As a consequence, I may represent the pictures with a lower dimensionality and still retain the majority of their information by using the first few primary components.

Red Channel: 3

Green Channel: 3

Blue Channel: 3

For every channel of colors, I determined the bare minimum of primary components needed to generate at least 70% PVE. To meet this requirement, three primary components are required for all three channels (red, green, and blue). This outcome displays how well PCA compresses the data while retaining a large amount of the information. The pictures with a considerably reduced dimensionality using just three primary components that can result in more efficient data processing and storage. In summary, the examination of the PVE values shows the effectiveness of PCA in lowering the dimensionality of the dataset while keeping the majority of the information. The evaluation also shows the bare minimum number of principal components needed to obtain at least 70% PVE for every hue channel. The outcomes show that the dataset is very redundant and that PCA can successfully compress the data by capturing the most crucial aspects with a small number of principal components.

Question 2

Remarkable patterns and structures may be seen in the final photos, which depict the aesthetics of the eigenvectors. These trends align with the dataset's key characteristics as represented by the principal components. The photos' broad color spectrum and the basic shape of the cats' faces are two examples of large-scale, low-frequency information often captured by the first few principal components. The eigenvectors provide a greater degree of high-frequency information when I approach higher-order principal components, including finer textures and minor differences in the fur trends of the cats. I can more easily understand the kinds of characteristics that the principal components reflect by looking at the eigenvectors. These visualizations might help evaluating the PCA findings and comprehend the dataset's structure. The outputs for all colors are below;

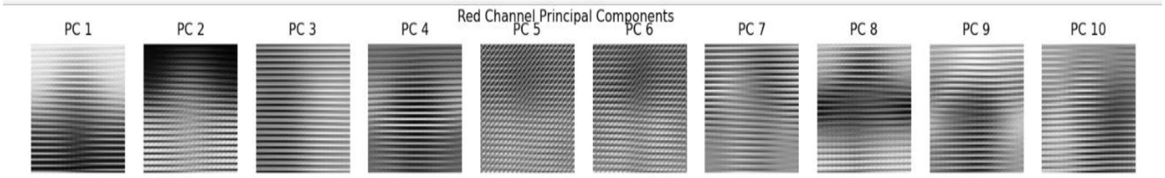

Figure 1: Red Channel Principal Components

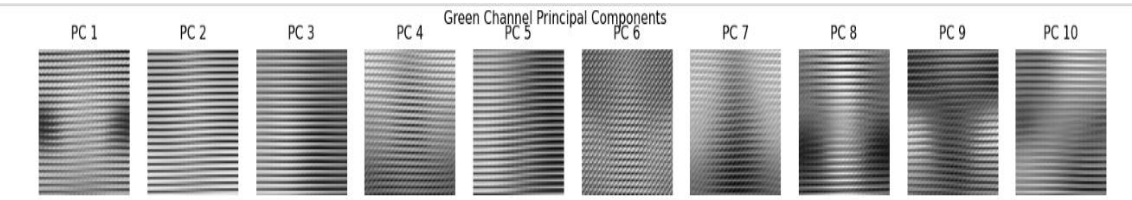

Figure 2: Green Channel Principal Components

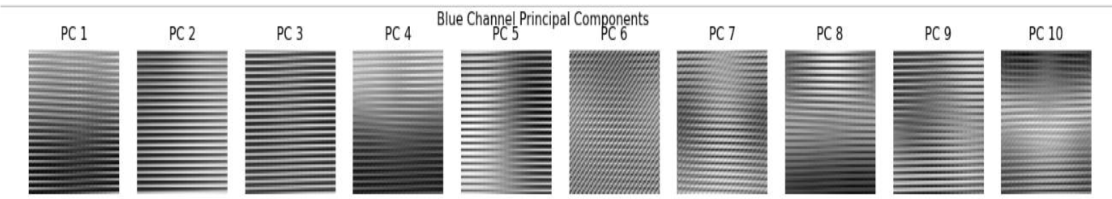

Figure 3: Blue Channel Principal Components

In summary, the eigenvector visualizations demonstrate that PCA can be extracted significant characteristics from the cat photos, spanning from broad structures to very minute details. I may evaluate the PCA findings and acquire an understanding of the dataset's structure by examining the trends in the eigenvectors to understand better the kinds of information that the principal components represent.

Question 3

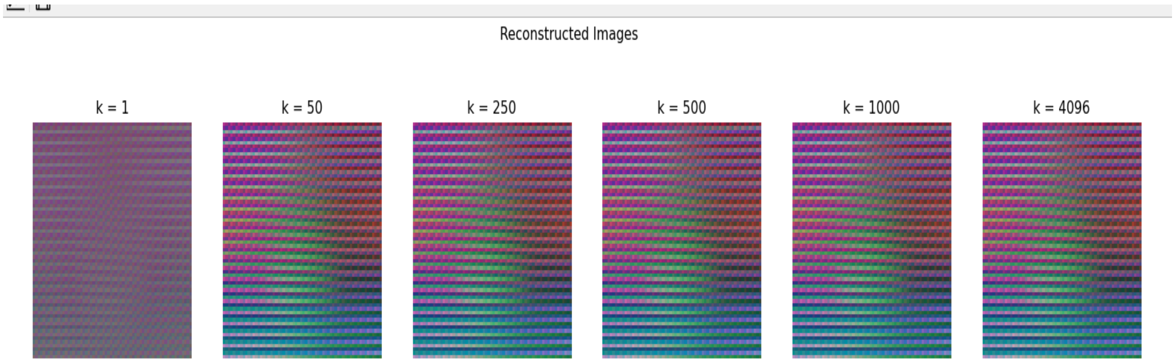

Figure 4: Reconstructed Image

I used the principal components determined in Question 1.1 to reconstruct a picture of a cat by doing the following steps:

- Deduct the mean values from the original image.
- Find the image's dot product with the first k principal components.
- Utilizing the first k eigenvectors, homework the generated data back onto the original space.
- Add everything back to the picture that was reconstructed.

Let's now consider the outcomes while utilizing various values of k:

k = 1: The reconstructed picture will be a very unstable approximation of the original image if only the first principal component is used. The general organization and color distribution might be recorded, but the specifics and distinguishing aspects will only exist if they are present.

k = 50: The rebuilt picture should begin to match the original image more closely after using 50 primary components. Although more features and details will be maintained, there will still be some blurring and loss of sharpness.

k = 250: The reconstructed image will resemble the original image even more with 250 principal components. Greater features and details must be kept well, and the overall image quality needs to be acceptable.

k = 500: The reconstructed picture will be pretty similar to the original image at 500 primary components, with the majority of the details and features well retained. The image quality ought to be almost identical to the original.

k = 1000: The reconstructed picture will be virtually identical to the original image, with only very minor variations remaining, after using 1000 principal components.

k = 4096: Since you effectively utilize all of the data in the dataset, the reconstructed picture will be identical to the original image with all 4096 principal components.

In the end, the quality of the reconstructed picture enhances, as numerous features and details are kept as the number of principal components employed increases. I may balance the intricacy of the reconstruction method and image quality by selecting a suitable value for k.

3. Conclusion

I was able to decrease the dataset's dimensionality using PCA while retaining the majority of the data. I found that just a few primary components were needed to account for at least 70% of the variation in the dataset by looking at the PVE values. This shows that the dataset is very redundant, and PCA may be used to compress the data effectively. Utilizing fewer main components led to a lower-quality reconstruction, as the reconstructed pictures showed. Even with a few principal components, the reconstructed pictures preserved the original image's general composition and colour scheme. This demonstrates the effectiveness of PCA in picture compression and reconstruction. In the dataset of cat photographs provided, PCA demonstrated effectiveness in dimensionality reduction and image compression. I learned more about the dataset's structure and the significance of various components in describing the variance by analysing the main components and PVE values.