Assignment 4 (Part 2 - PCA)

By- Akhil singh
Roll no.:20171210

INTRODUCTION :
Principal component analysis (PCA) is a procedure which can be used to reduce image sizes and this uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components . Let us suppose there are ‘n’ observations with ‘p’ variables, then the number of
distinct principal components is minimum of (n − 1 , p). This transformation is defined in such a way that the first principal component has the largest possible variance, and each succeeding
component in turn has the highest variance possible under the constraint that it is orthogonal to the preceding components. The resulting vectors are an uncorrelated orthogonal basis set.

Procedure:
The problem in the dataset was that if we simply calculated the basis then the matrix dimensions would’ve been [256x256x3][256x256x3] , performing calculations on this huge matrix is infeasible, to avoid this problem we use a different approach.Now suppose we have a matrix A o
n which we want to perform PCA on, suppose that the columns of A have already been normalized to have zero mean, so that we just need to compute the eigenvectors of the covariance matrix ATA. Now if A is an m × n matrix, with n >> m , then ATA is a very large n × n matrix. So instead of computing the eigenvectors of ATA ,what we do is that we like to compute the eigenvectors of the much smaller m×m matrix (which in our case is 520x520) AAT.Assuming we can figure out a relationship between the two. So we figure out how are the eigenvectors of A T A related to the eigenvectors of AAT.
Let v be an eigenvector of AAT with eigenvalue λ. Then

$AA^T v = \lambda v$

$A^T (AA^T v) = A^T (\lambda v)$

$(A^T A)(A^T v) = \lambda(A^T v)$

In other words, if v is an eigenvector of AAT , then ATv is an eigenvector of A T A, with the same eigenvalue. So when performing a PCA on A , instead of directly finding the eigenvectors of ATA
(which may be very expensive), it's easier to find the eigenvectors v of AAT and then multiply these on the left by AT to get the eigenvectors ATv of ATA

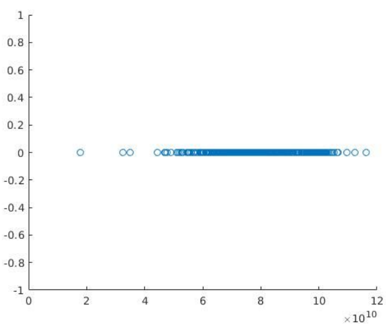
Conclusion :

We get the basis and we use the basis to reduce the dimension of the dataset to 35.

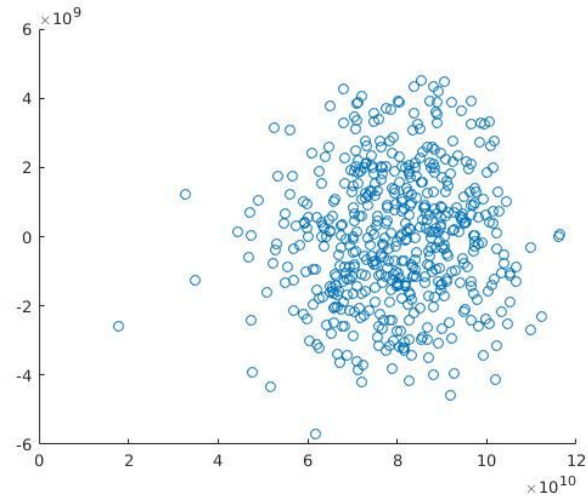
Original **Reconstructed**


Using scatterplots to examine how the images are clustered in the 1D, 2D and 3D space using the required number of principal components.

1D



2D



3D