Procedure followed for PCA :

PCA

* Convert the colour images to pixels after grayscaling and store it in a 2d array → $d \times N$ (features × no. of img)
* Center the data by subtracting the mean.
* Compute covariance matrix

$$Cov = \frac{1}{N} \sum_{i=1}^{N} \underset{N\times 1}{(x-\mu)} \underset{d\times N}{(x-\mu)^T} = \underset{N\times N}{Cov}$$

* Compute the eigen vectors of cov

let v be the eigen vectors of XX^T & u for X^TX

$$X^TX u = \lambda u$$

$$XX^T v = \lambda v$$

Multiply X^T on both sides

$$X^TX(X^Tv) = \lambda(X^Tv)$$

$$X^TX\,\underset{\downarrow}{u} = \lambda\,\underset{\downarrow}{u}$$

$\therefore$ u can be replaced by X^Tv

* Eigenvector $[N \times N]$ out of which $\$N$ are selected for reconstruction.
* Here Z represents the co-efficients of all components for all imgs.

$$Z = \underset{d\times N}{X^T} * \text{eigen vectors} = \underset{N\times \$N}{\quad} = \underset{d\times \$N}{Z}$$

* New features : New

$$New = X_{N \times d} \cdot Z_{d \times \$N} = New_{N \times \$N}$$

* For reconstruction multiply the new features with Z^T.

$$\widetilde{old} = New_{N \times \$N} * Z^T_{\$N \times d} = \widetilde{old}_{N \times d}$$

* Mean Square error :

· Add the mean to the $\widetilde{old}$ and subtract it pixel wise from original x to get mean square error.

$$MSE = \frac{\sum_{i=1}^{N} \sum_{j=1}^{d} [\widetilde{old}[i][j] - x[i][j]]^2}{N * d}$$

* $\$N = 32$, from the mean square error plot seemed most appropriate for reconstruction · as it has least mse.

For reconstructing images the $\widetilde{old}$ matrix must be reshaped into a 3d matrix and converted to image using PIL.

Original images and their corresponding reconstructed images for N=32

1)

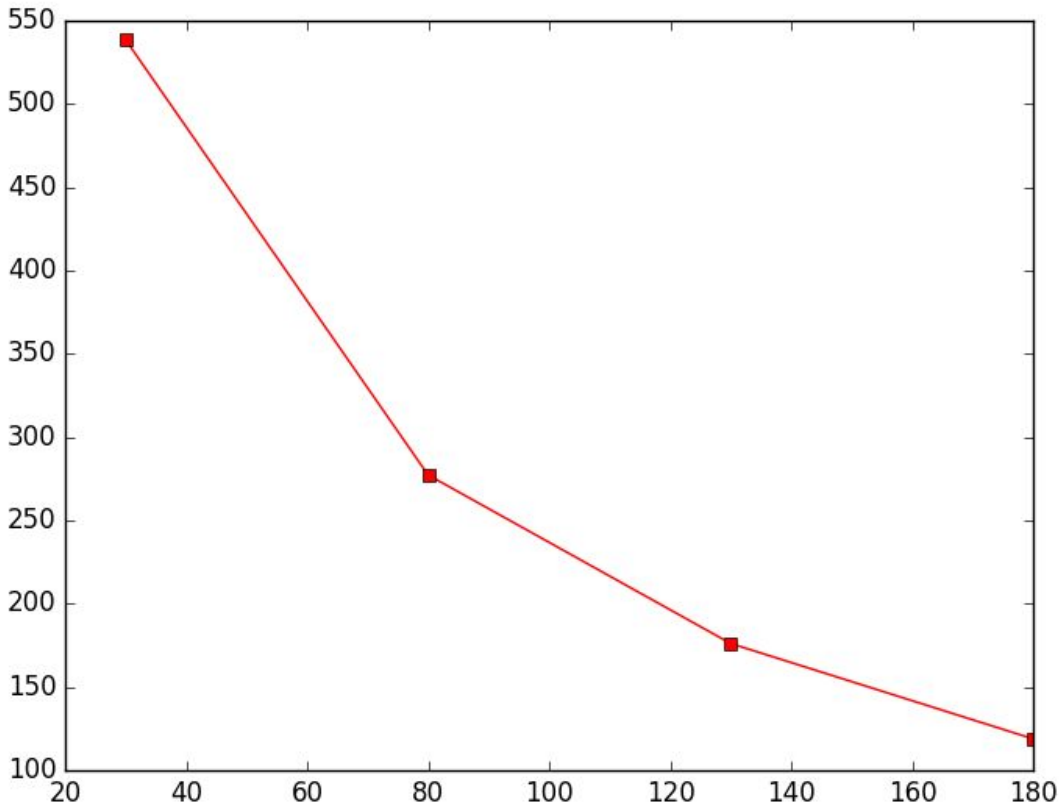
2)

3) 

4) 

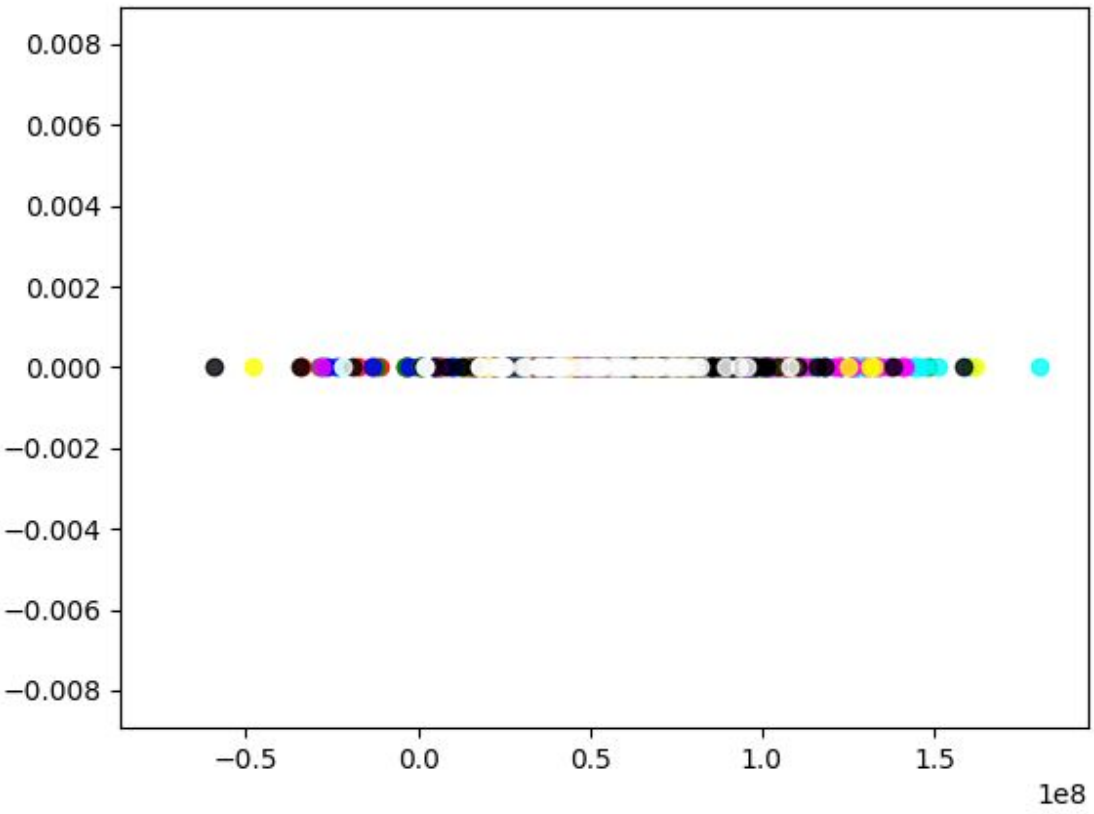
5) 

6) 

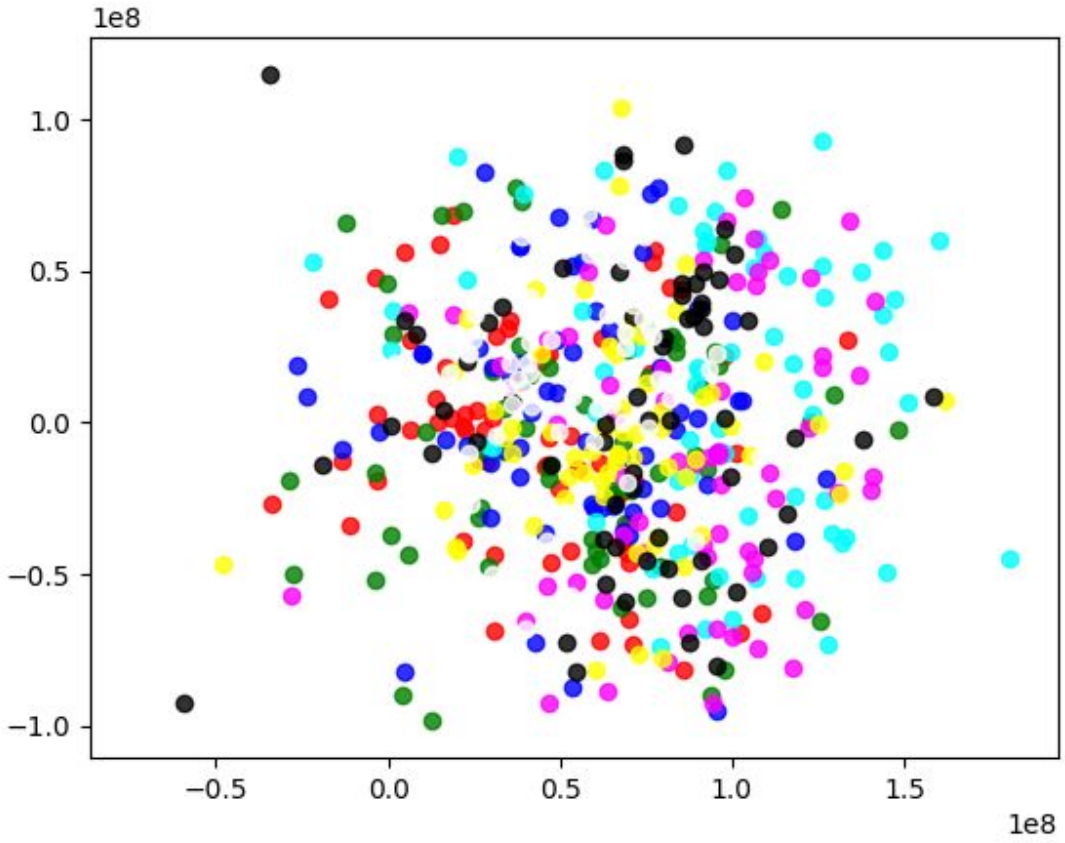


x-axis represents the no.of eigenvectors chosen
Y-axis represents the mean square error

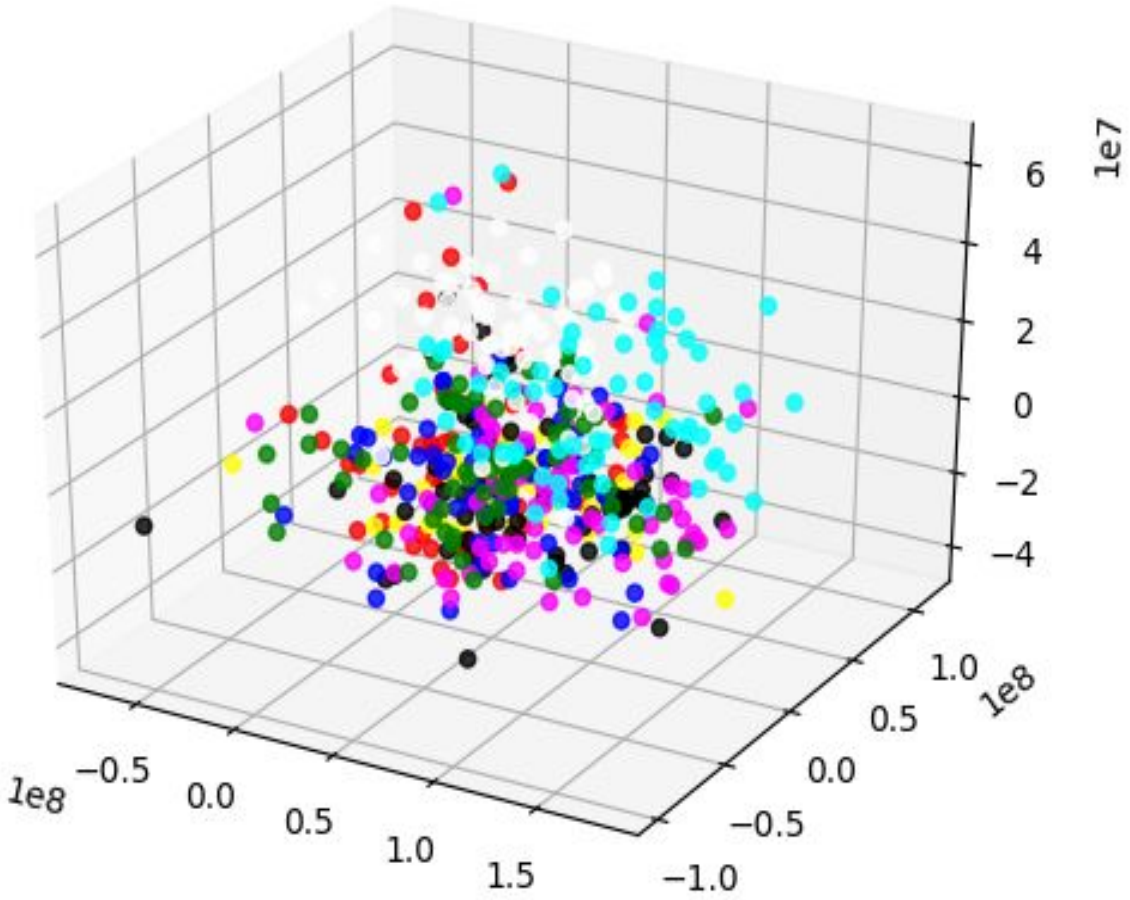
Mean Square Error Plot



Scatter plot for image clustering in 1D



Scatter plot for image clustering in 2D



Scatter plot for image clustering in 3D