

# Shivani\_workbook

November 9, 2021

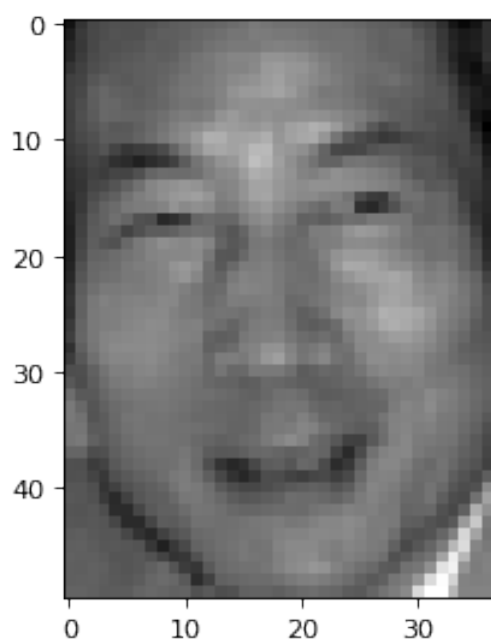
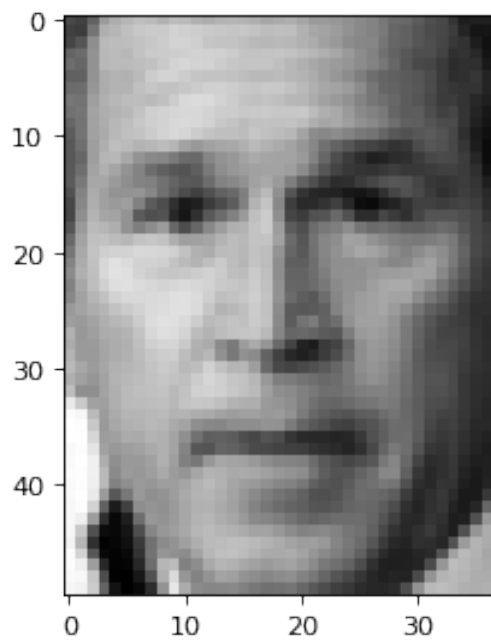
```
[1]: %load_ext autoreload
      %autoreload 2
      %matplotlib inline
```

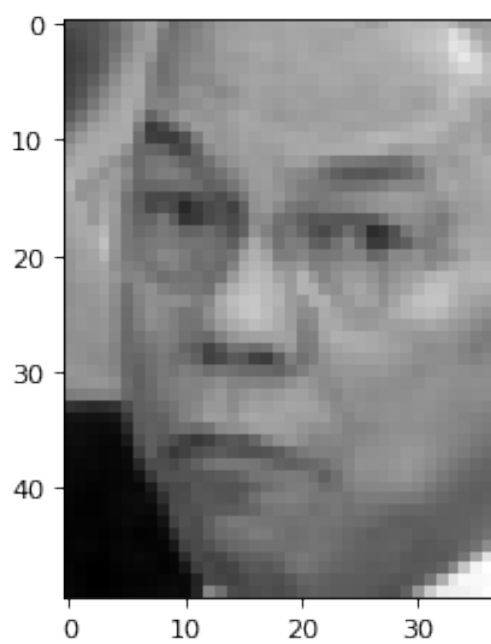
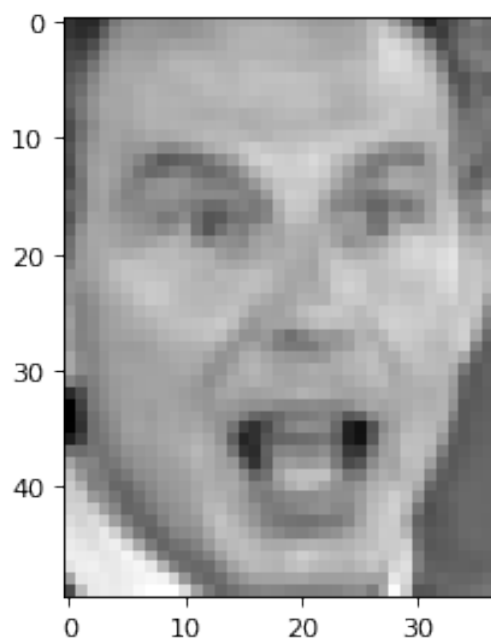
```
[2]: import sys
      from pathlib import Path
      import matplotlib.pyplot as plt
      #plt.rcParams["figure.figsize"] = (2, 6)
      #plt.rcParams['figure.dpi'] = 600
      if 'faces' not in sys.path:
          sys.path.append('faces')
```

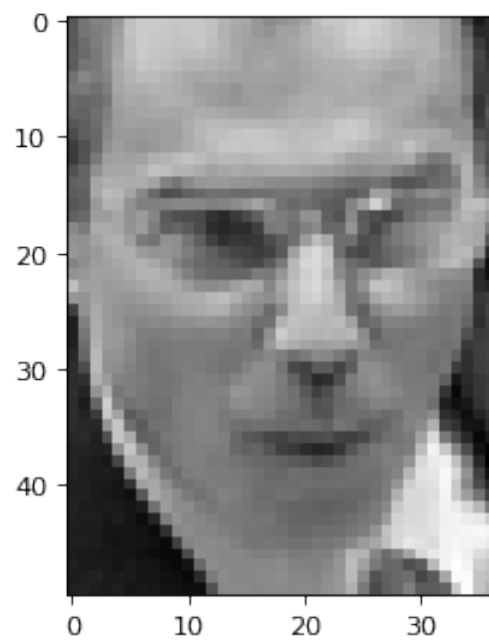
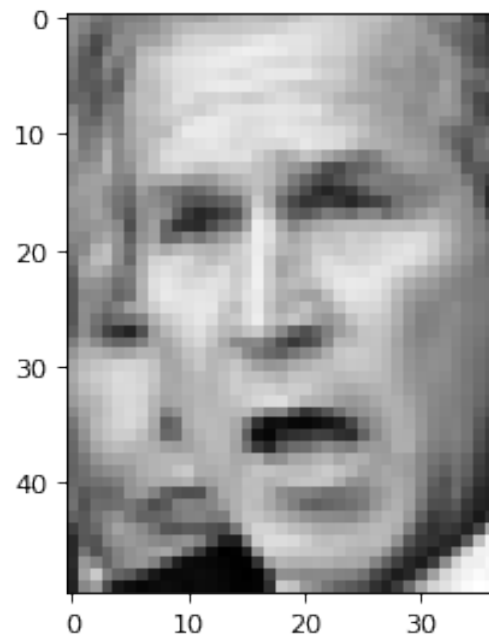
```
[3]: import faces
      from faces import main
```

```
[52]: main()
```

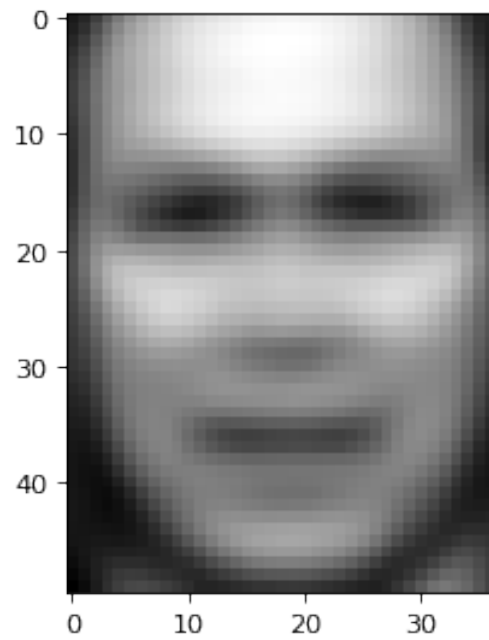
```
Total dataset size:
    num_samples: 1706
    num_features: 1850
    num_classes: 16
Random 6 images...
```







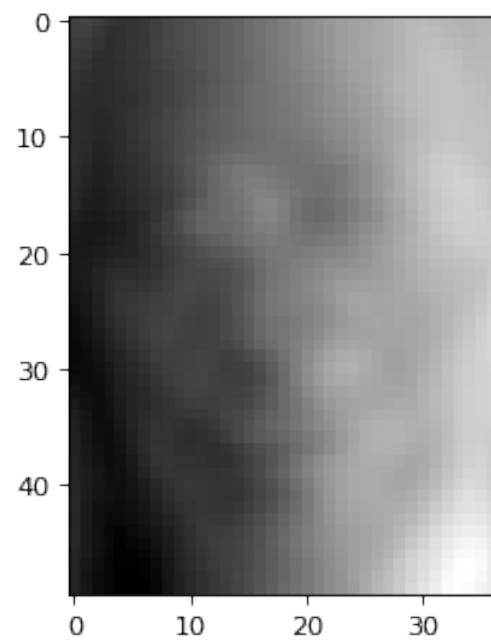
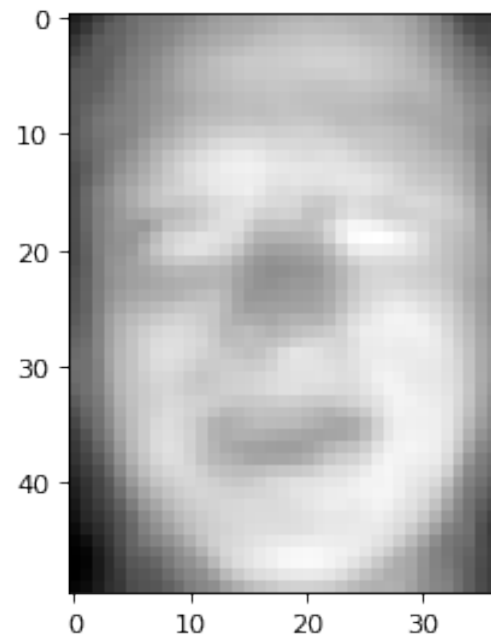
average of all images...



Top 12 eigen vectors...



for 2 eigen faces...



Applying PCA and reconstructing ut back  
for L of 1



for L of 10





for L of 50



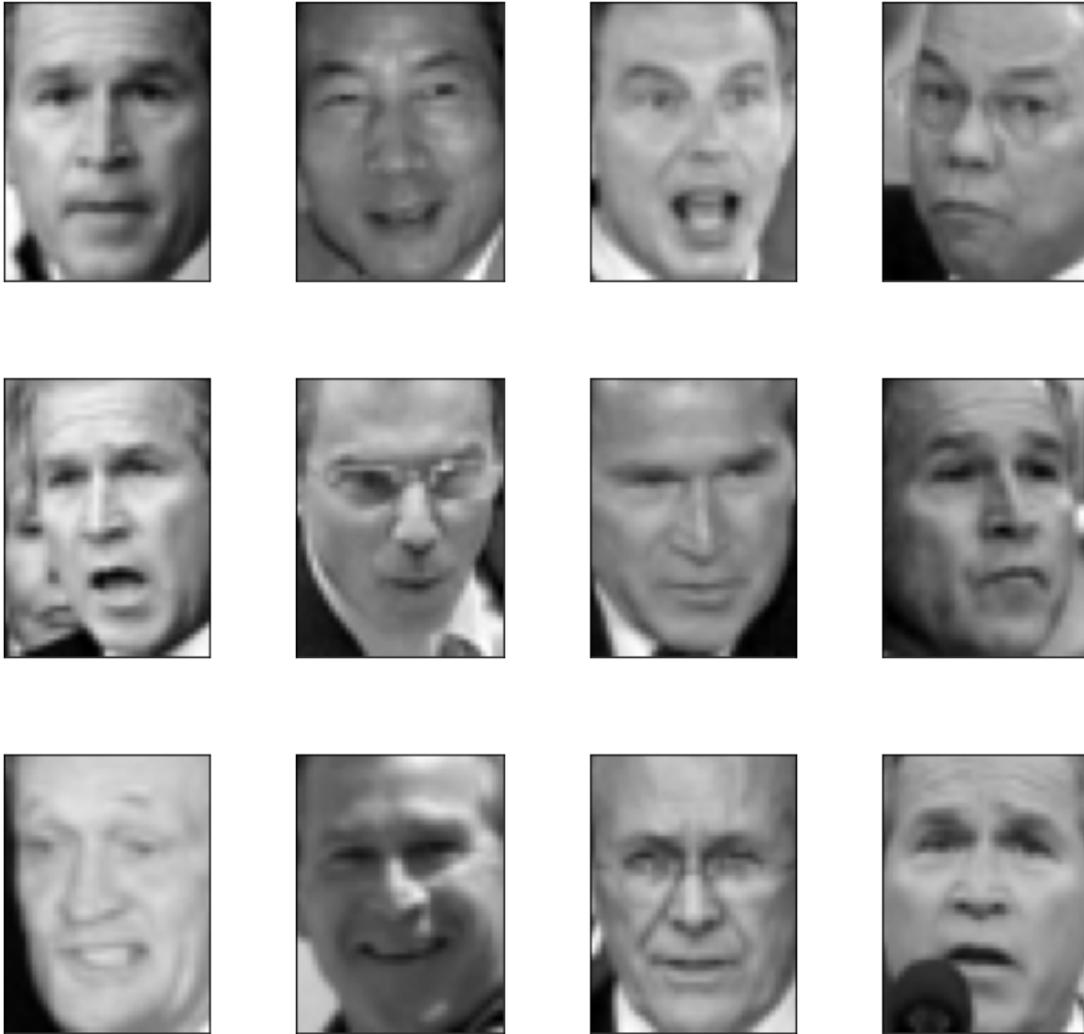
for L of 100



for L of 500

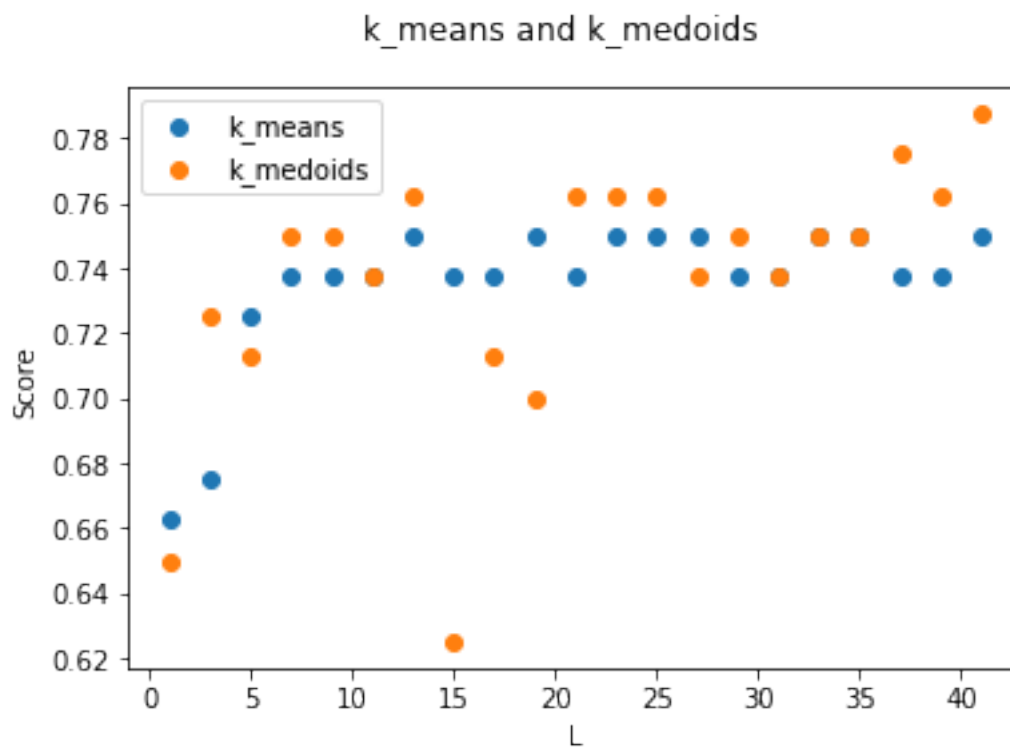


for L of 1288

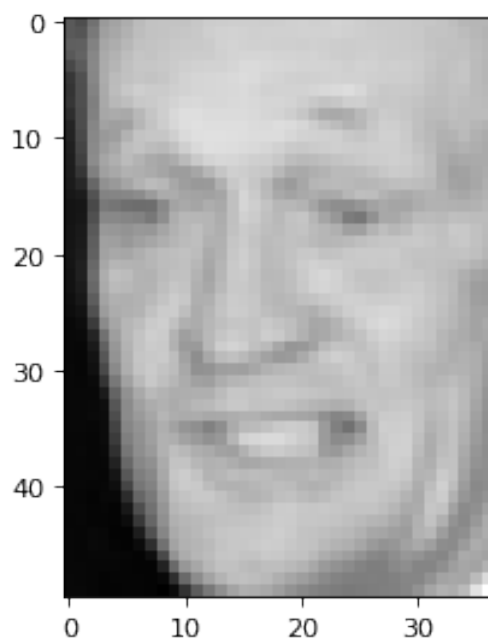


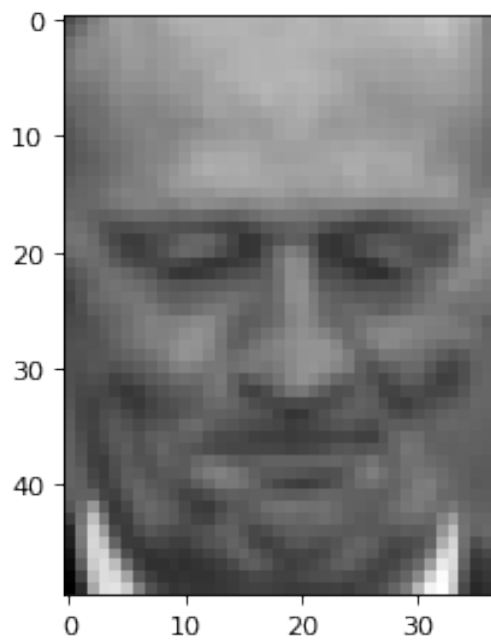
Mean, Min and Max for Kmeans are : 0.505 0.4875 0.51875

Mean, Min and Max for kmedoids are : 0.490625 0.46875 0.51875

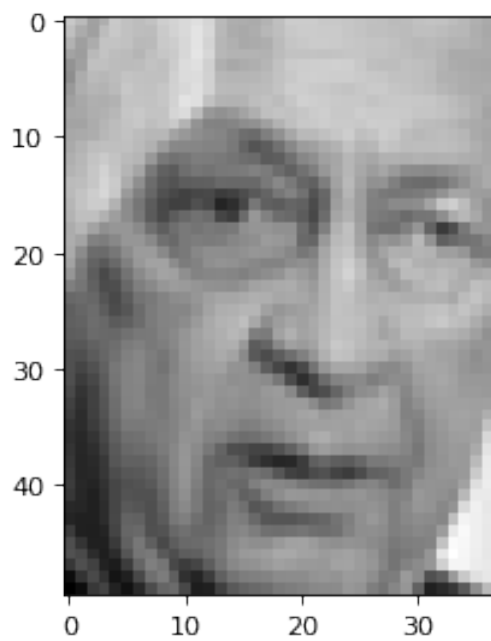


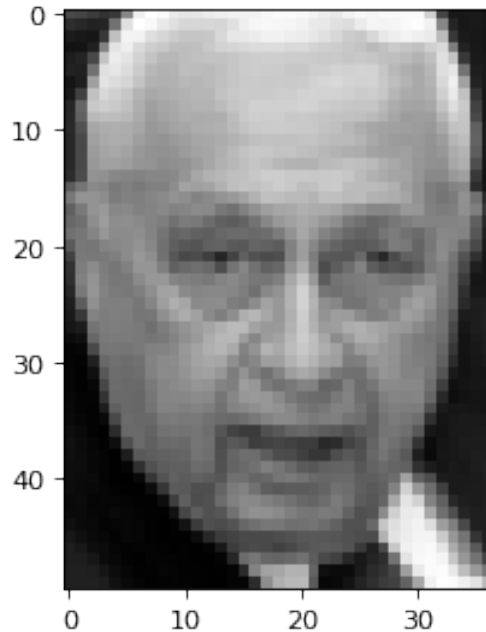
Best pair with the score : (7, 13) 1.0





Worst pair with the score : (0, 3) 0.5





### 0.1 4. PCA and image reconstruction

1. Average Face just gave an idea about how the overall face would be or look like, but without main features like eyes, nose etc.
2. The eigen faces obtained are not clear, it is better than average face, but does not give much information clearly on faces. Top eigen faces because, eigen values represents the correlation between the variables and the chosen ones are that represent the lowest dimension.
3. As the value of  $L$  increases the image becomes more clearer and is more easy to recognize.

### 0.2 6. Clustering Faces

1. In terms of performance Kmeans and Kmedoids are almost the same, but kmeans seems to be slightly better than kmedoids. kmedoids takes more time to run when compared to that of kmeans.
2. As the value of  $l$  increases, the score of kmeans and kmedoids increases. That is because, as  $l$  increases the value of  $l$ , the number of dimensions/correlation of variables considered increases and hence the kmeans and kmedoids will be able to cluster more properly and increase the purity and score increases.
3. The image with the best pair, has score 1, which indicates that the cluster purity is good, which shows that they can be easily clustered, but for worst pair, has score of 0.5, which indicates that the cluster purity is not great and it is difficult to cluster as they are very similar.

#### 7. Experimental Design

provided the dataset, compute the distance between the centroids of each cluster, that is try to get the euclidean distance or manhattan distance. Compute a matrix between each cluster in comparison with others. check if the minimum distance encountered will provide the set of images that are very similar and the maximum distance encountered will provide the set



the set of images that are totally different. compare that images obtained with the above experiment.

(Assumption : the centroids that are closer will have kind off similar faces)