# ANALYSIS REPORT FOR PCA AND LDA

RATNESH KUMAR RAI(MIT2019098)

* PCA ON SATELLITE IMAGES OF RIVER

Dataset:

We were given four images of a river taken by satellite in each blue, green, red, and Intensity.

At First, we did data pre-processing i.e. zero out mean, we **did not** the normalize the data as all data where given on same scale i.e. 0 -256. Then applied the PCA to find the covariance matrix, and eigen vectors as asked in the question.

Result:

**Figure 2**



**Figure 1**

**Figure 4**

**Figure 3**

**Figure 3**



**Conclusion:**

**As** we can observe, Fig.1, Fig 2(R, G band) cannot be used for river – non river classification, as the boundaries of river not clearly visible and hence can’t be ignored. But Fig.3 and Fig. 4 (B and I band) carry most significant information and can be used to classify it although Fig 4, is the best fit for image classification.

* PCA ON FACE CLASSIFICATION USING EIGEN FACES:
  + Dataset:
    - For this assignment we used a images of 10 subjects, each having 10 pictures, in different scenarios like different facial expression, different hair style, with and without glasses etc.
    - We classified total Dataset into 60% for training and 40% for testing of each individual.
    - So, to in sort, we have 4 set of images each subject of total 10 subjects i.e. 40 images for testing, remaining for training.
    - For, reducing the complexity we are only taking the grayscale value, for our project.
    - Each image is of dimension 112x92 pixels. So, We have total (10,304) features.
  + **Result**:

|  |  |
| --- | --- |
| (No. Of Feature Vector)K | Accuracy(%) |
| 1 | 15 |
| 2 | 52.5 |
| 5 | 82.5 |
| 8 | 92.5 |
| 10 | 97.5 |
| 50 | 97.5 |

* For imposter Images, which do not belong to any class we still are getting false positive result, i.e. it’s predicting that it belongs to one of the class and this behaviour is expected as we didn’t set any threshold for what minm distance, we can say either it is co-related or not. We can classify it as one of the limitations of Eigen Faces.
* Threshold can be set as 22% of max\_distance , for our current dataset, again it depends on one’s own requirement on the basis False Acceptance Rate(FAR) Vs False Matching Rate(FMR), strict threshold result to good FAR but poor FMR,
  + on 22% as threshold we have 18% FAR and 2% FMR
  + on 15% as threshold we have 8% FAR and 7% FMR
  + **Conclusion:**
    - The above result shows us that out of 10,304 pixels, there is only 10 major pixels, which hold maximum significant information for our face classification, that can accurately classify a person up to 97.5%.
    - Thus, we learned that PCA can largely reduce dimension of data for faster and optimised calculation.
* PCA and LDA ON FACE CLASSIFICATION USING EIGEN FACES:
  + Dataset is same as above mentioned.
  + **Result**:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **PCA(without threshold)** | | **PCA+LDA using Euclidean distance** | | **PCA+LDA using Mahalanobis distance** | |
| k | Acc. | k | m | Acc. | k | m | Acc. |
| 5 | 82.50 | 5| 3 | 57.50 | 5 | 3 | 62.50 |
| 10 | 97.50 | 10| 5 | 75.00 | 10| 5 | 82.50 |
| 15 | 97.50 | 15| 7 | 70.00 | 15| 7 | 97.50 |

* + - Comparison on PCA vs LDA+PCA

k = no. of principal component for PCA, m = no. of principal component for LDA, Acc. = Accuracy in %

Thus, we observed that for the given dataset PCA+LDA using Mahalanobis distance performs as the best, and is the best suit for face data. It requires only 7 features to perform as best as PCA over 10 features.

For comparison betwn PCA and PCA+LDA using Euclidean distance we will choose PCA as we get better accuracy.

* + **Conclusion:**
    - The above result shows that how a 10,304 dimensional data is reduced to only 7 dimensional data using PCA + LDA, which in turns shows that PCA+LDA preserve most of information of dataset to provide best result even if compared to PCA only, if one time computation cost is not considered then we should always go for PCA+LDA for face classification.
    - Mahalanobis Distance performs better compared to Euclidean distance as it takes into account of the scale of the data, thus giving better result for face data.