Face Recognition with PCA: An Eigenfaces-Based Implementation

Berker Eriş
Artificial Intelligence and Data Engineering Department
Istanbul Technical University
Istanbul, Turkey
erisb22@itu.edu.tr

***Abstract*—This paper describes how to use Eigenfaces, a PCA-based method, in face recognition. ORL face set is used to evaluate the performance of this method. Data preprocessing, PCA computation, face reconstruction, classification and noise testing are implemented from scratch in Python. Experimental results show that the number of eigenfaces (M) directly affects the quality of the application. In addition, the performance of the system under varying levels of Gaussian and Salt-and-pepper noise is compared.**

***Index Terms*—Face Recognition, PCA, Eigenfaces, Dimensionality Reduction, Noise Robustness.**

I. Introduction

Face recognition is a widely studied topic. One of the first approaches is the Eigenfaces method introduced by Turk and Pentland [1]. This method is based on Principal Component Analysis (PCA).

The basic idea of Eigenfaces method is to treat face images as high-dimensional vectors and identify the most significant components. These components, known as Eigenfaces, form a basis for face representation and recognition.

This project aims to implement the Eigenfaces method from scratch using the ORL face dataset.The project also shows the importance of the number of selected eigenfaces (M) on the reconstruction and recognition accuracy. Also, the robustness of the approach is analyzed using Gaussian and Salt-and-pepper distortions.

II. Methods

This section describes the implementation pipeline of the Eigenfaces method, which includes data preprocessing, Principal Component Analysis (PCA), face reconstruction, classification, and noise robustness testing.

A. Data Preprocessing

The ORL face dataset, consisting of 40 individuals with 10 grayscale images each (92×112 pixels), is used. Each image was reshaped into a vector of size 10,304 and stacked into a data matrix $X \in \mathbb{R}^{400\times 10304}$. The mean face μ was computed as:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$$

where x_i is the i-th face vector and $N = 400$. Images were normalized by subtracting the mean face:

$$\hat{x}_i = x_i - \mu$$

B. Principal Component Analysis (PCA)

To reduce computational cost, PCA used the smaller covariance matrix $C = XX^T \in \mathbb{R}^{N\times N}$ instead of X^TX. Eigenvalues λ_i and eigenvectors v_i of C were computed and projected into the original space to get the eigenfaces:

$$u_i = X^T v_i$$

The eigenfaces were then normalized to unit length. The top M eigenfaces which have the largest eigenvalues were used for dimensionality reduction.

C. Face Reconstruction

Given a normalized face $\hat{x}$ and the eigenfaces $U = [u_1, u_2, ..., u_M]$, the projection (or weight vector) w was computed as:

$$w = U^T \hat{x}$$

The face was then reconstructed from its lower-dimensional representation:

$$\tilde{x} = Uw + \mu$$

Reconstruction quality was evaluated using the Mean Squared Error.

D. Face Classification

All face images are projected into eigenspace and classified. For each class k, the mean vector $\bar{w}_k$ is calculated from the training set projections.

E. Robustness to Noise

To test robustness, two types of noise were applied to a subset of test images:

- **Gaussian noise:** pixel values were altered by adding noise sampled from $\mathcal{N}(0, \sigma^2)$ with varying σ values.
- **Salt-and-Pepper noise:** a percentage of randomly selected pixels were set to either 0 or 255.

A fixed number of eigenfaces ($M = 50$) was used for both noise types.

III. Results

This section includes the experimental findings of the project. Each result is supported by both quantitative metrics and visual illustrations.

A. Mean Face Comparison (Task 1)

Experiment: Two types of mean faces were computed: one using all 400 images and another using only the first image of the first 10 individuals. As shown in Fig. 1, the mean face computed from the full dataset is smoother and less biased.

Fig. 1. Left: Mean face from all 400 images. Right: Mean face from 10 selected images.

B. Eigenfaces and Variance Analysis (Task 2)

PCA was applied to the dataset to compute eigenfaces. The top 10 eigenfaces (largest eigenvalues) are visualized. Fig. 2 shows representative eigenfaces from this set and highlights how different eigenfaces capture global and local facial features.

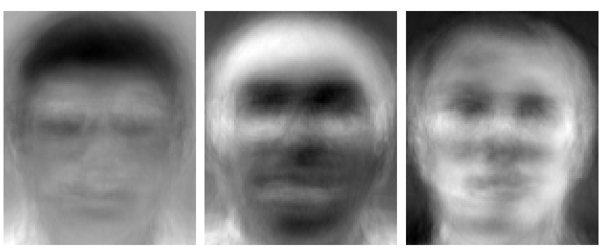

Fig. 2. Top 3 eigenfaces ranked by eigenvalue magnitude.

The distribution of eigenvalues is shown in Fig. 3, where the first few components dominate, indicating that most variance is captured in early directions.

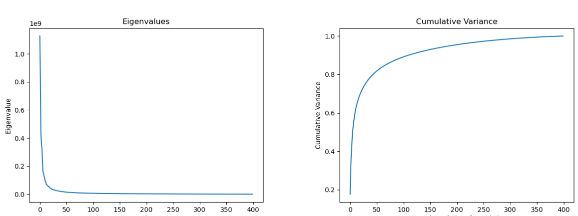

Fig. 3. Left: Sorted Eigenvalues. Right: Cumulative Variance vs. Number of Eigenfaces.

Experiment: To demonstrate the effect of increasing M, reconstructions of the initial training face were generated using different numbers of eigenfaces. Fig. 4 presents these results. As M increases from 10 to 300, the face becomes sharper and more realistic.

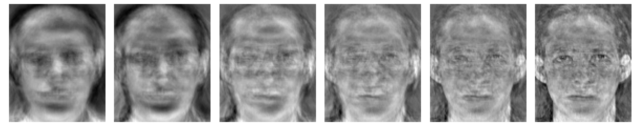

Fig. 4. Face reconstructed from eigenfaces using the first image of the dataset. M increases from left to right: 10, 20, 50, 100, 200, 300.

C. Face Reconstruction (Task 3)

The faces of individuals s10 and s11 were reconstructed for varying values of M. As seen in Fig. 5, the reconstruction quality is directly proportional to M.

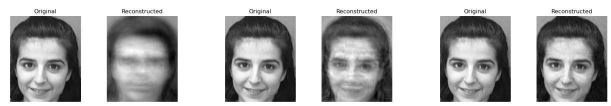

Fig. 5. Reconstructions for increasing M. Left to right: M=10, M=100, M=300.

Experiment: The minimum M value required to obtain $MSE < 500$ was found to be 50.

D. Classification Accuracy (Task 4)

Classification accuracy improves with increasing M, as shown in Fig. 6. Experiment: The best performance was obtained at $M = 200$ with 98.75% accuracy.

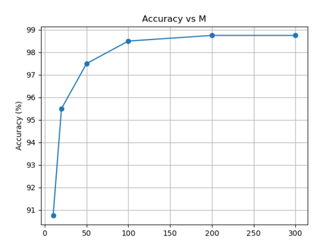

Fig. 6. Recognition Accuracy vs. Number of Eigenfaces.

Representative confusion matrices for $M = 10$, $M = 100$, and $M = 300$ are shown in Fig. 7.

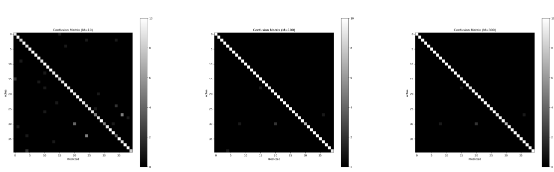

Fig. 7. Confusion matrices for $M = 10$, $M = 100$, $M = 300$.

E. Noise Robustness (Task 5)

Classification under noise was tested with $M = 50$. Fig. 8 shows that while Gaussian is robust up to a certain noise level, Salt-and-pepper shows a linear decrease in accuracy as noise increases.

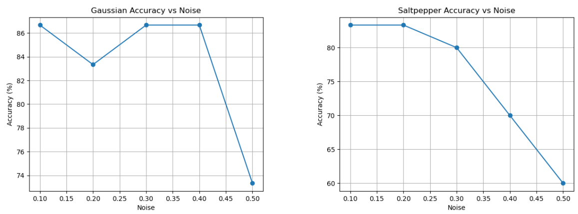

Fig. 8. Recognition accuracy under Gaussian (left) and Salt-and-Pepper (right) noise.

Sample distorted images are shown in Fig. 9.

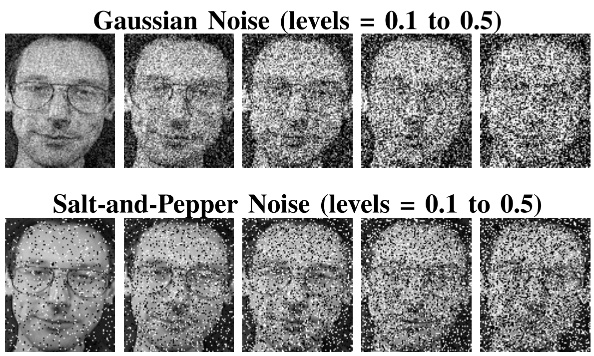

Fig. 9. Face image corrupted with increasing levels of Gaussian (top row) and Salt-and-Pepper (bottom row) noise. Levels range from 0.1 to 0.5 from left to right.

IV. Discussion

Experimental results show that Eigenfaces method is effective for face recognition. Important conclusions can be drawn from the findings:

A. Impact of Eigenface Count (M)

The number of eigenfaces (M) plays a critical role in both reconstruction and classification. As shown in Fig 5, low values of M fail and result in inefficient reconstructions. In contrast, $M \geq 100$ produces very efficient reconstructions with very low MSE.

The additive variance plot shows that more than 95% of the data variance can be preserved with less than 100 eigenfaces, proving that PCA is quite effective.

B. Recognition under Noise

The system shows relatively good robustness against Gaussian noise. Salt-and-Pepper noise was more destructive. (Figure 8).

This difference may be due to Salt-and-Pepper noise degrading basic local features such as eyes and mouth, while Gaussian noise introduces more distributed distortions.

C. Confusion Matrix Interpretation

For the confusion matrices (Figure 7), as M increases, off-diagonal errors decrease, indicating that higher dimensional projections classify faces better.

D. Limitations

Although Eigenfaces work well under the tested conditions, performance may degrade under real-world variations, such as lighting changes.

V. Conclusion

This work included the application of Eigenfaces method for face recognition using Principal Component Analysis. Operations such as image vectorization, PCA analysis, reconstruction and classification were applied.

Experimental results show that PCA effectively reduces the dimensionality of face images while preserving their essential features. The reconstruction and classification accuracy depend on the number of eigenfaces. The optimal balance between performance and efficiency is achieved at $M = 200$ with a classification accuracy of 98.75%.

Additionally, robustness to noise was evaluated under both Gaussian and Salt-and-Pepper distortions. While the system tolerated moderate Gaussian noise, it was more sensitive to Salt-and-pepper noise.

Overall, the Eigenfaces approach is an efficient approach for face recognition.

References

[1] M. A. Turk and A. P. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.