Project 5: Face Recognition Using Principal Component Analysis and Eigenfaces in MATLAB

Matteo De Angelis   
*College of Engineering*  
*University of Miami*Miami, USA  
mld163@miami.edu

Carter Falkenberg  
*College of Arts and Sciences  
University of Miami*  
Miami, USA  
ckf25@miami.edu

Kelly Rojas  
*College of Engineering*  
*University of Miami*Miami, USA  
knr20@miami.edu

*Abstract* — This project implements a face recognition system using the eigenfaces method based on Principal Component Analysis (PCA). The system is trained to detect and recognize facial images by projecting them into a reduced eigenspace and comparing them against known individuals. Two testing sets are used to evaluate the system: one containing known faces and another with unknown or non-face images. Key performance metrics include recognition accuracy, false positives, and the effect of varying the number of eigenfaces. The results demonstrate the effectiveness of PCA in dimensionality reduction and the trade-offs between recognition performance and computational efficiency. The project is implemented in MATLAB using grayscale face images, emphasizing theoretical understanding and practical application of eigenface-based recognition.

Keywords — Face Recognition, Principal Component Analysis (PCA), Eigenfaces, MATLAB, Image Classification, Dimensionality Reduction

# Introduction

Face recognition is critical in computer vision, with widespread applications in human-computer interaction, biometrics, and even security. Among the various techniques developed for face recognition, the eigenfaces method is highlighted due to its simplicity and effect8iveness in dimensionality reduction and feature extraction[1]. This project focuses on implementing the eigenfaces approach using PCA to develop a recognition system in MATLAB.

The system is trained on a dataset of facial images to compute the eigenfaces from PCA, and then tested on two sets of images: T1 (known faces), and T2 (unknown or non-faces). The primary objectives are to classify whether a test image is a face, and if it is a face, to identify it as a known individual from the training set. Only a subset of the most significant eigenfaces (those with the highest eigenvalues) is used, allowing analysis of recognition performance versus the number of retained components. Furthermore, the project investigates the system’s accuracy, failure cases, and sensitivity to the number of principal components. A simplified explanation of this approach is commonly found in educational resources that explain PCA’s role in facial classification[3].

# Theory

The eigenfaces methos is based on PCA, a statistical technique that transforms high-dimensional data into a lower-dimensional spaces while preserving the most important variance[2]. In the context of face recognition, each grayscale image is considered a high-dimensional vector, and PCA helps find a new set of orthogonal basis vectors (called eigenfaces) that best describe the dataset[1],[4].

First, the training images are mean-centered and then arranged into a matrix where each column represents a face vector. PCA is applied to this matrix to compute eigenvectors (eigenfaces) and their corresponding eigenvalues. The top eigenfaces (those with the largest eigenvalues) are retained, capturing the key variations among facial features[4].

A test image is projected into the reduced eigenspace to obtain a feature vector (its PCA coefficients). Recognition is then performed by comparing this vector to the feature vectors of training images using a distance metric like Euclidean distance. If the minimum distance is below a pre-defined threshold, the test image is classified as a known face. Otherwise, it is rejected as either an unknown or a non-face image. Additionally, PCA reduces noise and redundancy in facial data, and its comparison with other dimensionality reduction techniques further demonstrate its usefulness in features extraction[2],[3].

# Methodology

This project implements a face recognition system using the eigenfaces technique based on Principal Component Analysis (PCA). The purpose is to identify the provided images as either known faces, unknown faces, or non-faces. The methodology is divided into five parts: data pre-processing, PCA and eigenface computation, feature projection, classification, and evaluation.

## Data Preparation and Pre-Processing

The facial images were categorized into training and testing datasets, with the testing images divided into two sets, T1 (known faces) and T2 (unknown faces or non-faces). All the images were resized to a fixed resolution of 112x92 pixels for standardization. Color images were converted to grayscale, and each image was vectorized into a column vector using a custom pre-processing function. The resulting matrices trainData, testDataT1, and testDataT2 contained the flattened image vectors as columns.

## Eigenface Generation via PCA

PCA was applied to the mean-centered training data. The mean face vector was computed and subtracted from each of the training image vectors. Afterward, Singular Value Decomposition (SVD) was used to compute the eigenfaces from the training data. Furthermore, the top k eigenfaces corresponding to the largest eigenvalues were selected, where k varied to analyze the performance trends. Each of the training images were projected into the reduced-dimensional eigenspace to obtain the PCA coefficients, and which helped as feature representations.

## Feature Projection and Classification

The test images were also mean-centered and projected into the eigenface space using the selected eigenfaces. Furthermore, the Euclidean distance was calculated between each test image’s projection and the training projections for classification purposes. If the minimum distance was below an empirically selected threshold of three thousand, the images were classified as faces. Otherwise, they were not considered a match. Moreover, for the T2 images, the distances exceeding the threshold were classified as non-faces.

## Performance Evaluation

The accuracy of this project was assessed by computing the recognition rate for T1 (true positives) and the rejection rate for T2 (true negatives) across multiple k values. Furthermore, the accuracy information was plotted as a function of the number of eigenfaces to help with identifying the optimal dimensionality for balancing recognition and rejection performance.

## Analysis of Failure Cases

The failure cases from the T1 test set were visualized to assist with understanding misclassifications. The visualization provided the test faces that were classified as non-faces incorrectly due to the large distances from the training samples in the reduced eigenspace. Moreover, the analysis of these cases provided information on improvements for the threshold selection and overall model robustness.

# Results

Results.

Text Box to add picture and description and keep it in the same column together

# Conclusion

Conclusion.

##### References

1. M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991. [Online]. Available: <https://www.face-rec.org/algorithms/PCA/jcn.pdf>
2. S. Raschka, "PCA vs LDA," *Python Machine Learning Book - GitHub FAQ*, 2016. [Online]. Available: <https://github.com/rasbt/python-machine-learning-book/blob/master/faq/lda-vs-pca.md>
3. A. Mishra, "Face Recognition using PCA (Eigenfaces)," *YouTube*, Sep. 18, 2021. [Online]. Available: <https://www.youtube.com/watch?v=g4Urfno4aTc>
4. Wikipedia contributors, "Eigenface," *Wikipedia, The Free Encyclopedia*, 2024. [Online]. Available: <https://en.wikipedia.org/wiki/Eigenface>