Introduction

This project explores the use of Principal Component Analysis (PCA) for facial recognition, with a focus on how effectively PCA can extract meaningful facial features that distinguish individuals. Specifically, the project investigates how many principal components are needed to achieve accurate classification and what challenges arise when applying PCA to real-world facial images.

Facial recognition is an important application of machine learning and pattern recognition, with uses in security, authentication, and human-computer interaction. However, facial image data is often high-dimensional and noisy, making dimensionality reduction techniques such as PCA especially valuable. Understanding the strengths and limitations of PCA helps illustrate both its power and its shortcomings in real-world scenarios.

The dataset used in this project is Kaggle's Labeled Faces in the Wild (LFW) dataset, which contains over 13,000 facial images of more than 5,000 individuals. To ensure reliable learning, the dataset was filtered to include only individuals with a sufficient number of images.

To address the research questions, the project applies PCA to extract eigenfaces, followed by a Support Vector Classifier (SVC) trained on the reduced feature representations. Model performance is evaluated using classification accuracy and a confusion matrix.

Methods

The dataset was first filtered to include only individuals with between 20 and 30 images, ensuring that each class had enough samples for meaningful training. Images were converted to grayscale to simplify the feature space and reduce computational complexity. Each image was then flattened into a one-dimensional vector of pixel intensity values to allow PCA to be applied.

Principal Component Analysis was performed on the flattened image vectors to reduce dimensionality and extract the most informative components. These principal components represent eigenfaces, which capture common visual patterns across the dataset. Different numbers of principal components were tested to examine how dimensionality reduction affects classification performance.

After PCA transformation, the reduced feature vectors were used to train a Support Vector Classifier (SVC). The dataset was split into training and testing sets to evaluate generalization performance. The SVC was chosen due to its effectiveness in high-dimensional spaces and its common use in facial recognition pipelines.

Results

The PCA process produced eigenfaces that highlighted key facial structures such as eyes, nose, and overall face shape, while removing background noise and minor variations. The reconstructed images were blurry but still preserved some type of structure. Images in the dataset varied in lighting, angles, quality, among other things that lowered the accuracy.

Classification accuracy increased as the number of principal components increased, up to a point. Our model had an accuracy high of 34.97%, which is not the most accurate due to the complications in the images. This was a key limitation since PCA does not model things like pose, expression, or lighting. Ultimately, the classification was negatively affected in its accuracy.

Discussion

The results show that PCA is effective at extracting meaningful facial features and reducing dimensionality while preserving identity-related information. Using too few components results in loss of discriminative detail, while using too many components offers limited additional benefit and increases computational cost.

The combination of PCA and SVC worked well overall, demonstrating a classic and effective facial recognition pipeline. However, PCA's linear nature limits its ability to handle complex variations in real-world facial images, such as changes in lighting or facial expression. In our case we saw it with its low accuracy.

If more time were available, potential extensions include experimenting with nonlinear dimensionality reduction methods, incorporating face alignment techniques, or comparing PCA-based methods with convolutional neural networks.

Conclusion

This project demonstrated how Principal Component Analysis can be used to reduce the dimensionality of facial image data and extract meaningful features for classification. By combining PCA with a Support Vector Classifier, the project explored the relationship between the number of principal components and classification accuracy. The results highlight both the effectiveness and limitations of PCA when applied to real-world facial recognition tasks.