Face Recognition Report

1. Pre-processing and Data Splitting

The dataset was pre-processed by normalizing each face image vector to unit length. Each vector was divided by its magnitude. The data was then split into training and testing sets for each subject, with 150 images for training and 20 for testing per subject.

2. k-NN Classifier Implementation and Evaluation

Evaluation with Different k Values:

For k=2 and using Euclidean distance, you achieved a high accuracy of 99%. This indicates that using only 2 neighbors and Euclidean distance resulted in a very effective classifier for this dataset.

However, using cosine similarity as the distance metric resulted in significantly lower accuracies across all values of k. This could be due to the nature of cosine similarity, which measures the angle between vectors rather than their magnitude.

Decreasing the number of training images led to a decrease in accuracy across all configurations, indicating the importance of sufficient training data.Evaluation with Fewer Training Images:
When using fewer training images (100 instead of 150 per subject), the accuracies decreased further, especially for cosine similarity.

3. Comparison with Sklearn Classifiers

- Support Vector Machine (SVM) achieved a perfect accuracy of 100%, demonstrating its effectiveness for this task.
- Gaussian Naive Bayes (GaussianNB) achieved an accuracy of 85%, slightly lower than SVM but still performing well.

4. Dimensionality Reduction and Visualization

Principal Component Analysis (PCA) was applied to reduce the dimensionality of the data to 3 components. The resulting data was visualized in a 3D scatter plot, showing the distribution of classes in the reduced feature space.

Conclusion

- The k-NN classifier performed well with a small value of k and using Euclidean distance, but performed poorly with cosine similarity.
- SVM showed superior performance compared to k-NN and GaussianNB, achieving a perfect accuracy.
- PCA visualization provided insights into the structure of the data in a reduced-dimensional space.