Face recognition by Principal Components Analysis using ORL database

Liang Cheng
1024041117
Nanjing University of Posts and Telecommunications
School of Computer Science
Nanjing, China

***Abstract*—This paper primarily discusses the importance of Principal Component Analysis (PCA) in data processing, using its application in face recognition as an example. PCA, as a dimensionality reduction technique, can transform high-dimensional facial image data into a lower-dimensional feature space, thereby reducing computational complexity and improving recognition efficiency. The paper first introduces the basic principles of PCA and its application in image processing, then analyzes the characteristics of facial images and how PCA is used to extract facial features. Through experiments on multiple facial datasets, the study shows that PCA performs well in face recognition, effectively distinguishing different individuals. Finally, the paper discusses the suggests the possibility of improving recognition accuracy by combining PCA with other algorithms, such as Support Vector Machines (SVM) or deep learning techniques. Overall, PCA provides an efficient and practical solution for face recognition, especially in resource-constrained scenarios.**

***Index Terms*—principal components analysis, face recognition.**

I. INTRODUCTION

Principal Component Analysis (PCA) was first introduced by the mathematician and statistician Karl Pearson in 1901 as a method for analyzing and simplifying complex datasets [1]. Initially, PCA was a technique for extracting patterns and reducing dimensionality in multivariate data. Pearson's original concept focused on identifying a set of orthogonal axes that could explain the maximum variance in the data, laying the groundwork for modern applications of PCA [2].

Today, PCA has become a fundamental tool in machine learning and data processing, serving not only for dimensionality reduction but also for feature extraction, data compression, and noise reduction. It is widely applied in dimensionality reduction and data analysis across various fields such as image processing, machine learning, and pattern recognition. PCA transforms high-dimensional data into a smaller, more manageable set of features while preserving the essential variance of the original data. This makes it particularly useful when dealing with complex datasets, where computational efficiency and model performance are crucial.

In the context of face recognition, PCA has proven to be an effective method for extracting key features from facial images, reducing the complexity of the recognition process. By projecting facial images onto a lower-dimensional space, PCA identifies the most significant components that capture the variance in facial structures, making it easier to compare and recognize individual faces. The technique's ability to handle large sets of image data while maintaining accuracy has made it a foundational tool in many computer vision applications.This paper explores the application of PCA in face recognition, discussing its underlying principles, benefits, and challenges.

II. PRINCIPAL COMPONENT ANALYSIS

Principal Component Analysis (PCA) is a widely used technique for dimensionality reduction and data preprocessing. In this section, we discuss the application of PCA in preparing data for machine learning algorithms by reducing its dimensionality, while retaining the most significant features.

A. Overview of PCA

PCA is a statistical method that transforms the data into a new coordinate system [3], where the axes (called principal components) are ordered according to the variance of the data along those axes [4]. The first principal component captures the direction of maximum variance, while each subsequent component captures the direction of the next largest variance, subject to the constraint that the components are orthogonal to each other.

The primary goal of PCA is to reduce the dimensionality of large datasets while preserving as much of the data's variance as possible. This process makes it easier to visualize high-dimensional data and speeds up computations by removing less significant features. PCA is often used in data preprocessing to improve the performance of machine learning models by reducing overfitting and increasing interpretability.

B. Mathematical Foundations of PCA

Given a dataset $\mathbf{X} \in \mathbb{R}^{n \times p}$, where n is the number of samples and p is the number of features, PCA performs the following steps:

1) Data Centering: First, the data is centered by subtracting the mean of each feature:

$$\bar{\mathbf{X}} = \mathbf{X} - \mu$$

where μ is the mean vector of the features:

$$\mu_j = \frac{1}{n} \sum_{i=1}^{n} X_{ij}$$

2) Covariance Matrix Computation: Next, the covariance matrix $\mathbf{C}$ is computed to capture the relationships between features:

$$\mathbf{C} = \frac{1}{n-1}\bar{\mathbf{X}}^\top \bar{\mathbf{X}}$$

3) Eigen Decomposition: The eigenvectors and eigenvalues of the covariance matrix are computed through eigen decomposition:

$$\mathbf{C} = \mathbf{V}\Lambda\mathbf{V}^\top$$

where $\mathbf{V}$ is the matrix of eigenvectors and Λ is the diagonal matrix of eigenvalues. The eigenvectors correspond to the principal components, and the eigenvalues represent the variance along each component.

4) Principal Components Selection: Select the top eigenvectors corresponding to the largest eigenvalues. These eigenvectors form the principal components. Select the top k eigenvectors corresponding to the largest eigenvalues to form a new basis matrix $\mathbf{V}_k$.

5) Data Projection: The original data is then projected onto the selected principal components to reduce its dimensionality:

$$\mathbf{Z} = \bar{\mathbf{X}}\mathbf{V}_k$$

where $\mathbf{Z}$ represents the transformed data in the reduced dimensionality space.

C. Geometric Interpretation of PCA

Geometrically, PCA finds the directions along which the data varies the most. These directions are represented by the principal components, and the data is projected onto these components. The first principal component is the direction of maximum variance, the second is orthogonal to the first and captures the second-largest variance, and so on.

By projecting the data onto a smaller number of principal components, PCA effectively reduces the dimensionality while retaining the most important features. This is particularly useful in high-dimensional datasets, where removing redundant or less significant features can improve both the efficiency and the performance of machine learning algorithms.

D. Application of PCA in Data Preprocessing

PCA is often applied as a preprocessing step in machine learning tasks to achieve the following benefits:

1) Dimensionality Reduction: In many real-world datasets, especially those with a large number of features, PCA can be used to reduce the number of dimensions while preserving the most important information. This helps to mitigate the curse of dimensionality and reduces the complexity of subsequent machine learning models.

2) Noise Reduction: By focusing on the most significant principal components, PCA can help eliminate noise and irrelevant features, which may be present in the smaller eigenvalues. This helps improve the quality of the data and, in turn, enhances the performance of machine learning models.

3) Data Visualization: In datasets with more than two or three dimensions, PCA can be used to project the data into two or three principal components, making it easier to visualize the structure and relationships within the data. This is particularly useful for exploratory data analysis and understanding complex datasets.

E. Summary

PCA is a powerful tool for data preprocessing, enabling dimensionality reduction, noise filtering, and data visualization. By transforming the data into a new set of orthogonal features, PCA simplifies the data structure while retaining its essential characteristics. As a result, it can significantly improve the performance of machine learning models and facilitate the analysis of high-dimensional datasets.

III. DATA PREPROCESSING

Dimensionality reduction is a crucial technique in data preprocessing that simplifies datasets by reducing the number of input variables or features. By focusing on the most informative features, dimensionality reduction enhances the efficiency of machine learning models, reduces computational costs, and helps mitigate the curse of dimensionality. This chapter discusses the significance of dimensionality reduction, the techniques used, and their applications in various domains.

A. Importance of Dimensionality Reduction

The process of dimensionality reduction addresses challenges associated with datasets containing a large number of features. High-dimensional data can lead to overfitting, increased computational complexity, and difficulty in visualization. Dimensionality reduction techniques help in addressing these issues by retaining the essential information while removing redundant or less informative features.

Reducing the number of features can improve model performance by focusing on the most relevant data. Additionally, lower-dimensional data reduces memory requirements and accelerates model training and prediction speeds. Furthermore, dimensionality reduction techniques such as Principal Component Analysis (PCA) help visualize high-dimensional data in two or three dimensions.

B. Techniques for Dimensionality Reduction

To reduce the dimensionality of the dataset, PCA was implemented in this paper. PCA can effectively reduce the dimensionality of the dataset and improve model performance. Figure 1 shows the effect of dimensionality reduction on facial images.

We can see that the reconstructed image after dimensionality reduction still retains the main features. We can consider an image as an $n \times p$ array in grayscale space or an $n \times p \times 3$ array in RGB space. Taking a grayscale image as an example, PCA can be used to reduce the $n \times p$ matrix to an $n \times l$ matrix ($l < p$), achieving data compression.

It is evident that the more principal components retained, the clearer the reconstructed image becomes. When retaining 10 principal components, the reconstructed image has

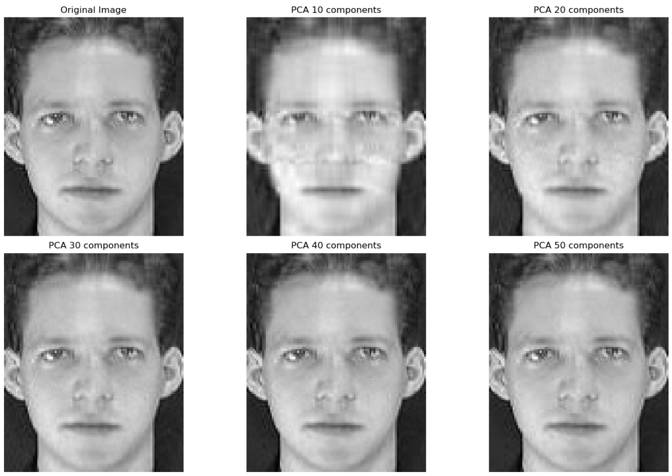

Fig. 1. Before and After Applying PCA for Dimensionality Reduction.

already achieved relatively high clarity. At this point, the dimensionality-reduced data is approximately 20% of the original data size.

When reducing an $n \times p$ image to $n \times l$ ($l < p$) dimensions, two smaller matrices need to be retained: a principal component matrix of size $p \times l$ and a matrix of the new image data of size $n \times l$. Therefore, if the mean vector and standard deviation vector, which contribute minimally to the data size, are ignored, the data compression ratio can be approximated by:

$$\text{Compression Ratio} = \frac{p \times l + n \times l}{n \times p}$$

For grayscale image compression, when $n = 512$, $p = 512$, and $l = 50$, the data compression ratio is approximately 19.53%. For color image compression, when $n = 512$, $p = 512 \times 3$, and $l = 50$, the data compression ratio is approximately 13.02%. The face image shown in Figure 1 has a resolution of 92×112, so the dimensionality-reduced data is approximately 20% of the original data size.

C. Summary

Dimensionality reduction is a valuable tool in data preprocessing, especially for high-dimensional data. By employing techniques such as PCA, it is possible to reduce the number of features, enhance computational efficiency, and improve the performance of machine learning models. However, it is crucial to carefully evaluate the results and consider the potential trade-offs in terms of information loss and model accuracy.

IV. Face Recognition

Face image recognition is a fundamental task in computer vision, with applications in security, human-computer interaction, and biometrics. In this chapter, we explore the implementation of a face recognition system using Principal Component Analysis (PCA) for dimensionality reduction, followed by the use of a classifier to identify individuals based on their facial features. PCA allows for the extraction of the most significant features of the face images, enabling a more efficient recognition process by reducing computational complexity.

A. Face Image Dataset

For the purpose of this study, we utilized the well-known ORL Face Database [5], which contains grayscale images of 40 subjects, each having 10 different images taken under varying lighting conditions and facial expressions [6]. The dataset has a resolution of 112x92 pixels per image, providing a high-dimensional feature space for analysis.

B. Feature Extraction

PCA is employed to reduce the dimensionality of the face images while preserving the most important features. The process begins with the construction of the covariance matrix of the standardized image data. The eigenvalue decomposition of this matrix allows us to identify the principal components, which represent the directions of maximum variance in the data. The top k eigenvectors corresponding to the largest eigenvalues are selected as the new basis for the transformed data. The original face images are then projected onto this lower-dimensional space to obtain the reduced features.

We used the feature set after dimensionality reduction to train a face recognition classifier. This paper applies three classification techniques for face recognition:

1) Euclidean Distance Classifier: The Euclidean Distance Classifier calculates the distance between the test image and each training image in the reduced feature space. The class of the nearest image (with the smallest distance) is assigned to the test sample. The distance is computed using:

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^{k} (x_i - y_i)^2}$$

where $\mathbf{x}$ and $\mathbf{y}$ are the reduced feature vectors of the test and training images, respectively.

2) k-Nearest Neighbors (KNN): The KNN classifier assigns a test image to the majority class among its k-nearest neighbors, and the Euclidean distance was used as the distance metric.

3) Support Vector Machine (SVM): The SVM classifier was applied with a linear kernel to maximize the margin between classes in the transformed space. SVM is particularly effective for high-dimensional data, making it suitable for the reduced feature set obtained through PCA.

C. Recognition Accuracy

The face recognition system made predictions on a set of images from a face dataset. The recognition accuracy of the system is measured by the percentage of correctly recognized faces in the entire test set. We observed the changes in recognition accuracy after varying the number of images used for training (K). As shown in Figure 2, the results indicate that after applying PCA for dimensionality

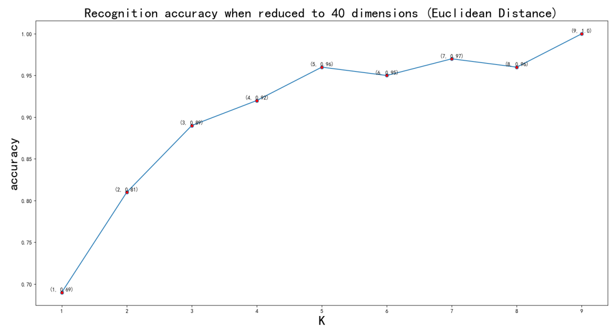

Fig. 2. Accuracy when reduced to 40 dimensions (Euclidean Distance).

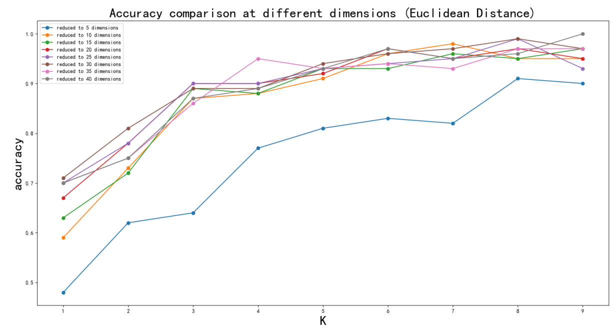

Fig. 3. Accuracy comparison at different dimensions (Euclidean Distance).

reduction, the recognition accuracy generally increases as the number of training images increases, which is consistent with general patterns. When sufficient training data is provided, the recognition accuracy can exceed 90%. This suggests that PCA dimensionality reduction does not significantly damage the primary features of the data while greatly reducing data redundancy. Compared to using the original high-dimensional data, PCA dimensionality reduction can significantly improve efficiency.

D. Impact of Dimensionality Reduction

In this study, a systematic evaluation of the recognition system's performance was conducted, focusing on the impact of the number of principal components—that is, the dimensionality (k) of the feature space after reduction—on the system's recognition accuracy. As a classic unsupervised linear transformation method, the core objective of Principal Component Analysis (PCA) is to achieve dimensionality reduction by projecting the original high-dimensional data onto a lower-dimensional orthogonal subspace, while maximizing the preservation of data variance.

The experimental results revealed a distinct trend: as the number of selected principal components (k) was incrementally increased, the system's recognition accuracy showed a significant upward trajectory. This phenomenon is anticipated, as each additional principal component corresponds to an eigenvalue, ordered by magnitude, and represents a portion of the data's variance. The principal components added in the initial phase correspond to larger eigenvalues, capturing the primary structural and most discriminative information within the original data, thus contributing most significantly to the enhancement of the classifier's performance.

However, this performance improvement did not exhibit limitless linear growth. Once the number of principal components reached a certain critical point, further increases in k yielded negligible gains in recognition accuracy, with the performance curve approaching an asymptotic plateau. The underlying reason for this "plateau" phenomenon is that the eigenvalues corresponding to subsequently added components progressively decrease. These later components primarily capture minor variations or even random noise from the data. Incorporating these components, which have a low signal-to-noise ratio, into the feature vector not only fails to provide effective supplementary information for the model's decision boundary but may also introduce noise, potentially having a slight adverse effect on the model's generalization capabilities.

In summary, these experimental findings robustly demonstrate the efficacy of PCA in feature extraction and dimensionality reduction. By selecting an optimal number of principal components (k), it is possible to substantially reduce the dimensionality of the feature space with almost no sacrifice of critical recognition information. This not only significantly enhances the training and inference speed of subsequent classification models and reduces storage and computational complexity, but also mitigates the risk of the "Curse of Dimensionality" to some extent by filtering out noise-associated dimensions. Ultimately, this analysis validates that PCA enables an optimized trade-off between computational efficiency and recognition accuracy, providing both theoretical and empirical support for constructing efficient and robust recognition systems.

E. Comparison with different Methods

To further analyze the effectiveness of the PCA-based approach, I implemented different classifiers for face image recognition. It was observed that for various classification methods, the data processed with PCA still achieved high recognition accuracy, demonstrating the advantage of PCA in reducing the feature space and enhancing the classifiers' ability to distinguish faces. Moreover, PCA provided faster computation times for both training and testing, highlighting its efficiency in face recognition tasks.

Notably, compared to other methods, the recognition accuracy was lower when using KNN as the classifier, especially when the data dimensionality was low. As the dimensionality increased, this issue gradually diminished, and by 40 dimensions, the accuracy of the KNN method was comparable to that of other methods. This phenomenon is likely due to the insufficient number of features in the dataset. Since the KNN algorithm relies on calculating the distance between the sample to be classified and all training samples, having too few features may lead to underfitting.KNN does not explicitly

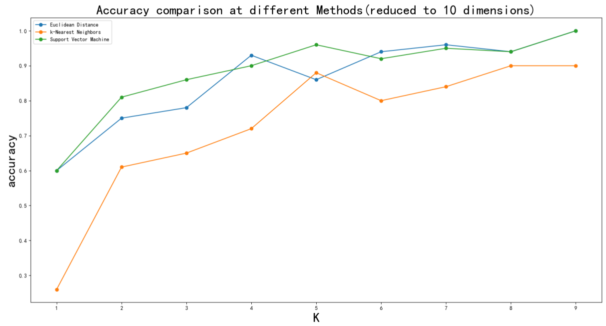

Fig. 4. Accuracy comparison at different Methods(reduced to 10 dimensions).

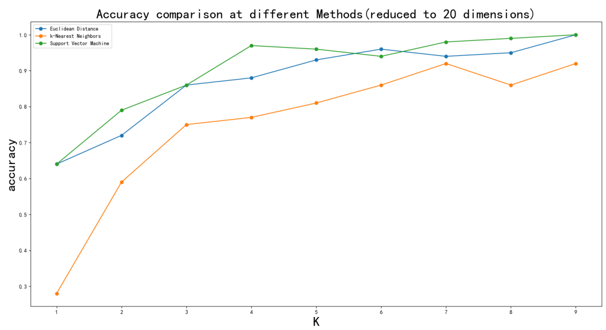

Fig. 5. Accuracy comparison at different Methods(reduced to 20 dimensions).

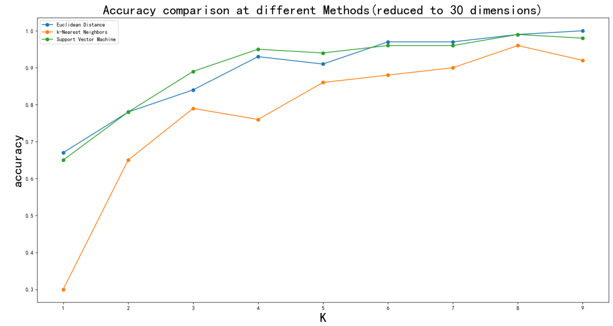

Fig. 6. Accuracy comparison at different Methods(reduced to 30 dimensions).

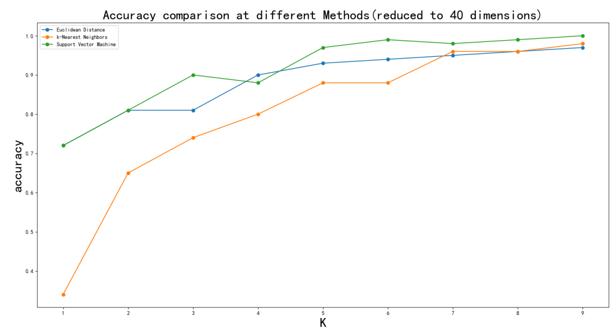

Fig. 7. Accuracy comparison at different Methods(reduced to 40 dimensions).

filter out noise during training. When noise or outliers are present in the data, KNN's predictions can be significantly affected, especially in low-dimensional spaces. Noisy samples can influence the selection of nearest neighbors, leading to a reduction in accuracy.

In addition, KNN is a *non-parametric* algorithm, meaning it does not build an explicit model during training. Instead, it stores the entire training set and uses it during classification. This means that the classification decision depends entirely on the distribution of the training samples. If the training data is imbalanced or contains significant noise, KNN's performance may be affected. Unlike methods such as Support Vector Machines (SVM), KNN does not optimize the decision boundaries globally.

F. Summary

This chapter presented a face image recognition system using PCA for dimensionality reduction, followed by classification with Euclidean Distance, KNN, and SVM. The results demonstrated that PCA is an effective technique for improving the performance and efficiency of face recognition systems by reducing the data dimensionality while retaining essential features. The system achieved high accuracy, with dimensionality reduction significantly decreasing computational complexity without noticeable performance loss.

V. Conclusion

In this work, Principal Component Analysis (PCA) has been applied to face image recognition to reduce the high-dimensional data while retaining the essential features for classification. PCA, by extracting the principal components of the data, effectively compresses the information and removes redundancy. This dimensionality reduction significantly reduces computational complexity, making it easier and faster to process large datasets, which is essential in face recognition tasks where datasets often contain thousands of features.

We first performed the standardization of the face images and calculated the covariance matrix, followed by eigenvalue decomposition to extract the principal components. By projecting the face images onto the subspace spanned by the principal components, we achieved reduced dimensionality while retaining most of the information necessary for effective classification. The reduced dataset was then used for training a classifier, such as a Support Vector Machine (SVM) or k-Nearest Neighbors (k-NN), which showed high accuracy in identifying individuals from the test images.

The results demonstrated that PCA is an efficient and reliable technique for face image recognition, particularly in scenarios where computational resources are limited. However, PCA has some limitations, such as its linearity assumption, which may not fully capture the nonlinear nature of facial features.

In conclusion, PCA serves as an effective preprocessing step for face image recognition, providing a solid foundation for building scalable and efficient face recognition systems. The method can be further enhanced by exploring hybrid models that incorporate nonlinear feature extraction techniques.

REFERENCES

[1] K. Pearson, ”On Lines and Planes of Closest Fit to Systems of Points in Space,” *Philosophical Magazine*, vol. 2, no. 11, pp. 559-572, 1901.

[2] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed., New York, NY, USA: Springer, 2002.

[3] J. Shlens, ”A Tutorial on Principal Component Analysis,” arXiv preprint arXiv:1404.1100, 2014.

[4] H. Abdi and L. J. Williams, ”Principal component analysis,” *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 4, pp. 433-459, July/Aug. 2010, doi: 10.1002/wics.101.

[5] M. A. Turk and A. P. Pentland, ”Face recognition using eigenfaces,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, San Juan, Puerto Rico, 1991, pp. 586-591, doi: 10.1109/CVPR.1991.139758.

[6] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, ”Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711-720, July 1997, doi: 10.1109/34.598228.