Image Compression using PCA

Anurag Karehonna
ECE
02FE21BEC009

Basavesh Patil
ECE
02FE21BEC018

Diya Bandi
ECE
02FE21BEC031

Prerana Kolkar
ECE
02FE21bEC060

Guide : Dr.Dattaprasad Torse

January 10, 2024

Abstract

Image compression plays a pivotal role in managing and transmitting vast amounts of visual data efficiently. This project explores the application of Principal Component Analysis (PCA) as a technique for image compression. PCA, a widely used dimensionality reduction method, aims to represent data using a reduced set of orthogonal components that capture the maximum variance in the dataset. The project methodology involves implementing PCA on images to extract the most significant features, allowing for efficient representation with minimal loss of information. By analyzing the eigenvalues and eigenvectors obtained through PCA, a subset of principal components is selected to reconstruct the original image. The experiment evaluates the trade-off between compression ratio and image quality, measuring the impact of varying the number of principal components used for reconstruction. Additionally, the project investigates the computational efficiency of PCA-based compression compared to other existing techniques.

Keywords: Image Compression, Principal Component Analysis (PCA), Dimensionality Reduction, Eigenvalues, Eigenvectors, Compression Ratio, Image Quality, Computational Efficiency.

1 Introduction

Principal component analysis (PCA) is representing a data in an eigenvector with eigenvalues which is mean a new coordinate system; therefore PCA is a linear transformation [1]. The efficiently represented faces of images using PCA called Eignface which now become a standard and a common performance benchmark in face recognition [2]. PCA produced dimensionality reduction therefore, it suitable to use for lossy compression, while retaining those characteristics of the dataset which contribute most to its variance [3]. Many algorithms based on various principals leading to the image compression by reducing volume data of image [4]. Lossy Algorithms are based on the image color reduction, but reconstructed image are obviously used for some applications. Color image is converted to gray-scale (intensity) image. The resulted image is reducing the data volume belongs to the most common algorithms [5] This project delves into the utilization of PCA as a methodology for image compression, aiming to explore its efficacy in reducing image data size while maintaining acceptable visual quality. By analyzing the trade-offs between compression ratio and image fidelity and comparing the computational efficiency against traditional compression techniques, this study aims to provide insights into the practical applications of PCA in the realm of image processing and compression

So, can we reduce the image with the application of PCA? It is possible to reduce the image using PCA because it carries uncorrelated data

2 Princiapl component analysis

Principal Component Analysis (PCA) is a fundamental technique used in data analysis and dimensionality reduction. It's employed in various fields such as statistics, machine learning, image processing, and more Data analysis is a sub-area of statistics concerned with the description of joint data [5]. These methods seek to give the links that may exist between the different data and to draw statistical information there from which makes it possible to more succinctly describe the main information contained in these data. We can also try to classify the data into different more homogeneous subgroups. The purpose of these methods is to synthesize the large tables to provide a simplified presentation

Multifactorial methods [6] provide graphical representations that are the best possible summary of the information contained in a large data table. Therefore, we must accept to certain degree a loss of information in order to gain readability. Depending on the phenomena we want to study and the nature of the data table we have, we will apply a multi-factorial method. Indeed, there are many factorial method of data analysis that are based on the same mathematical theories

2.1 Key steps in Principal component analysis

2.1.1 Data Standardization:

PCA works best when the data is standardized (mean-centered and scaled) to ensure that each variable contributes equally to the analysis.

2.1.2 Covariance Matrix Computation:

PCA computes the covariance matrix of the standardized data. This matrix summarizes the relationships between different variables, capturing their joint variability.

2.1.3 Eigenvalue Decomposition:

The covariance matrix is then decomposed into its eigenvectors and eigenvalues. Eigenvectors represent directions or principal components, while eigenvalues denote their respective magnitude or importance.

2.1.4 Dimensionality Reduction:

By choosing a subset of the principal components, PCA allows the transformation of the data into a lower-dimensional space. This transformation retains most of the variability present in the original data

2.2 Objectives of PCA

There are several different approaches to PCA, but all agree on the conditions for its application and its general objective.

This method applies to quantitative data sets of at least two variables. Since this is a multifactorial data analysis method, its purpose is to summarize this dataset. This is done in the following ways:

- Provide simple and readable tools for representing the information processed, making it possible to bring out from the raw data the possible links existing between the variables (in terms of correlation),
- Give indications on the nature, the strength and the relevance of these links, in order to facilitate their interpretation and discover which are the dominant trends of the dataset.
- Effectively reduce the number of dimensions studied (and thus simplify the analysis), by seeking to express the original set of data as faithfully as possible thanks to the relationships detected between the variables.
- Identify Key Features: PCA helps in identifying the most influential features or variables within a dataset. It achieves this by determining the components that contribute the most to the variance observed in the data

3 The PCA problem

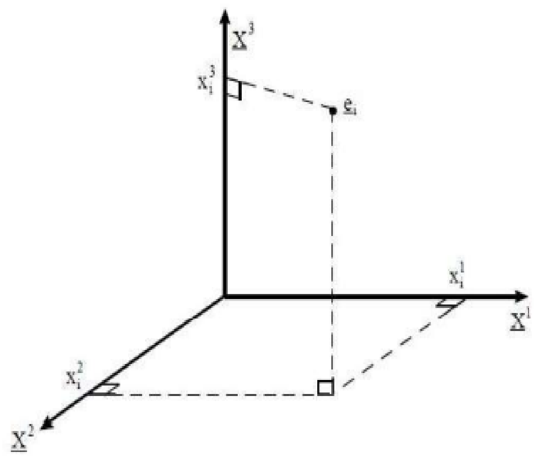

Figure 1: Represenation of an individual in dimesnion 3

4 Block Diagram

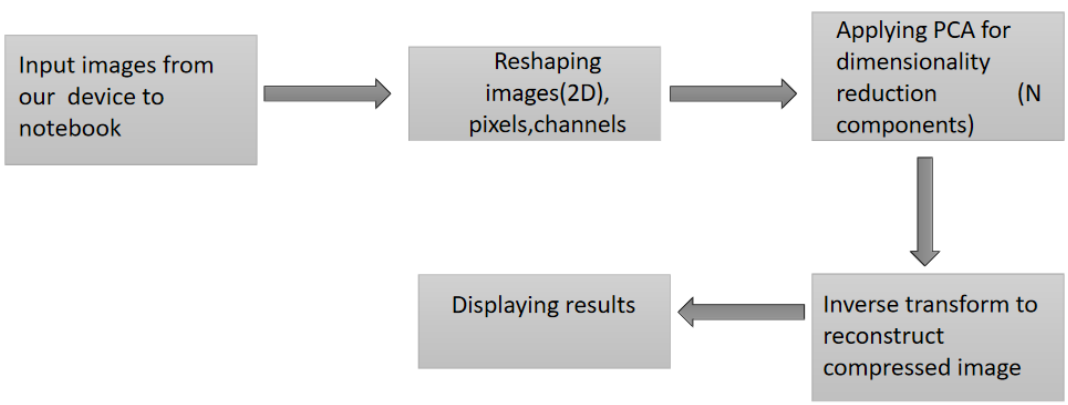

Figure 2: Block diagram

5 Results

5.1 output when components = 10

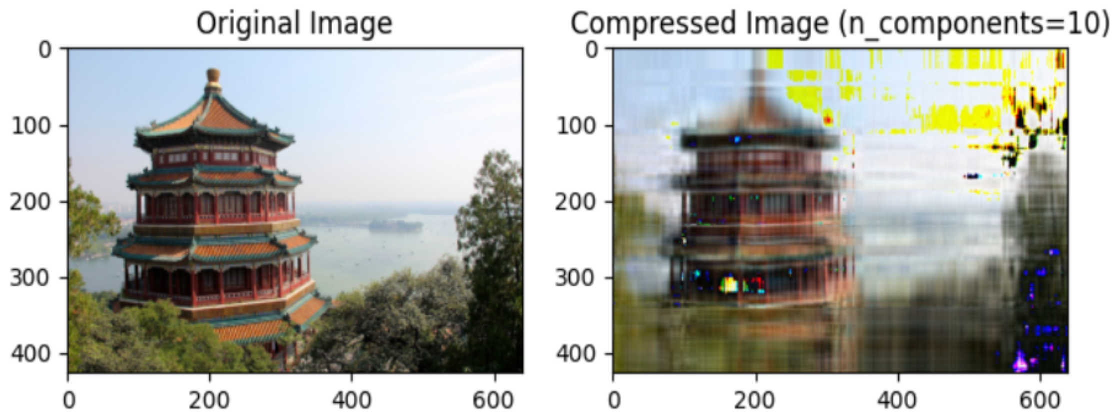

Figure 3: compressed image

5.2 output when components = 200

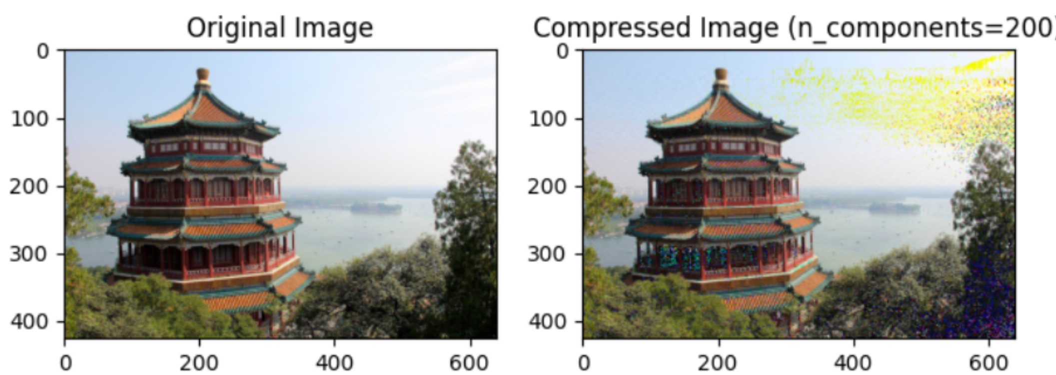

Figure 4: compressed image