IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

6th National Conference on Science, Technology and Communication Skills – NCSTCS 2K23

Narula Institute of Technology, Agarpara, Kolkata, India

Vol. 10, Special Issue 3, September 2023

Image Matrix Reduction Using Principal Component Analysis by Finding a New Variable

Puja Supakar[1], Shilpi Pal[2], Snehangshu Mahapatra[3] and Saket Srivastava[4]

Faculty, Department of Basic Science and Humanities, Narula Institute of Technology, Kolkata, India[1,2]

Student, Department of Computer Science and Engineering, Narula Institute of Technology, Kolkata, India[3,4]

Abstract: In this paper, we have shown how we can reduce the dimension of an image matrix. Dimensionality Reduction is a statistical and Machine Learning-based technique wherein we try to reduce the number of features in our dataset. This helps us in obtaining a dataset with optimal number of dimensions. We have used Principal Component Analysis (PCA) to reduce the dimensionality of this data set. Principal Component Analysis is a technique which uses sophisticated underlying mathematical principles to transform a number of possibly correlated variables into a smaller number of variables called principal components. It is one of the most important results from applied linear algebra. The advantage of PCA is finding the patterns in the data and compressing data by reducing the number of dimensions without loss of information. The Principal Component Analysis (PCA) is one of the most successful techniques that have been used in image recognition and compression. PCA is a statistical method under the broad title of factor analysis. The purpose of PCA is to reduce the large dimensionality of the data space (observed variables) to the smaller intrinsic dimensionality of feature space (independent variables), which are needed to describe the data economically. This is the case when there is a strong correlation between observed variables. The mathematical concepts that are used for PCA for dimensional reduction are Standard Deviation, Variance, Co–variance and Eigenvectors. By using this information, we have found a new set of variables, smaller than the original set of variables, retaining most of the sample's information and useful for the compression and classification of data. In summary, PCA is a specific technique for dimensionality reduction that works by creating new variables that capture the most important information in the data, and it is one of several methods available for reducing the number of features in a dataset.

Keywords: Principal Component Analysis (PCA), Dimension Reduction, Matrix, Standard Deviation, Co–variance.

I. INTRODUCTION

An image processing task consists of acquiring the image, pre-processing, segmentation, representation and description and finally recognition and interpretation. There are four types of digital images, binary, grey scale, true colour or RGB and indexed. The objective of image compression is to reduce irrelevance and redundancy of the image data to an acceptable level in order to be store or transmit data in an efficient form [1]. Image compression is reducing the size in bytes of a graphics file with no degradation in the quality of the image to an undesirable level [2]. We could use statistical analysis for our paper but Statistical analysis [3]. The dimensionality reduction problem is directly related to image compression. Principal component analysis (PCA) also known as Karhunen Loeve expansion, is one of the classical dimensionality reduction methods used for feature extraction which has been widely used in variety of areas such as signal processing, pattern recognition and data mining. PCA has been widely applied in the area of image compression in various forms, i.e., as a standalone image compression technique as well as a pre-processing or post-processing step in combination with other techniques [4]. The image transformation from colour to the gray-level (intensity) image 'I' belongs to the most common algorithms. Its implementation is usually based on the weighted sum of three-colour components R, G, B according to relation I = $w_1R + w_2G + w_3B$. The R, G and B matrices contain image colour components, the weights w_i were determined with regards to the possibilities of human perception [5]. Mathematical projection techniques can be employed in image matrix reduction to transform the image data into a lower-dimensional space while preserving relevant information. By projecting the image onto a lower-dimensional subspace, the dimensionality of the image matrix can be reduced, leading to a compressed representation. Using mathematical projection, the original data set, which may have involved many variables, can often be interpreted in just a few variables (principal components) [6]. The number of components that hold the majority of the information is called the intrinsic dimensionality and each data image may have a different intrinsic dimensionality. PCA condenses all the information of an "N" band original data set Original Image N PCs into a smaller number than "N" of new bands (or principal components) in such a way that maximizes the covariance and reduces redundancy in order to achieve lower dimensionality [8]. Analysis of high-dimensional imagery data is an expensive task in terms of memory utilization, computational power, and execution time. To effectively process high-dimensional imagery data, in minimum possible time and without sacrificing the quality of the process, the researchers proposed, over the years, many image analysis methods that can be categorized into six classes i.e., morphology-based methods, knowledge-based methods, wavelet-based methods, neural network-based methods, fuzzy logic-based methods, and dimension reduction-based methods [9]. Direct PCA provides the best classification accuracy but hierarchical PCA also performs considerably well with high computational results obtained by applying each of the techniques, and then advantage.

In summary, these first results obtained with the hierarchical PCA method are very encouraging, and a larger number of test data will be evaluated [10]. With minimal effort PCA provides a roadmap for how to reduce a complex data set to a lower dimension to reveal the sometimes hidden, simplified structures that often underlie it [12]. Fusion algorithms are input dependent. In order to carry out the quality analysis of the proposed fusion algorithm, different fusion algorithms using pyramids, PCA and wavelets are implemented [13]. PCA for dimensional reduction uses Standard Deviation, Variance, Co–variance, and Eigenvectors. By using this information, we have found an additional set of variables, smaller than the original set of variables. These variables retain most of the sample's information and are useful when it comes to compressing and classifying data. We have used PCA to reduce dimension of any sample image by finding a new variable and by maintaining quality of that image.

II. PRINCIPAL COMPONENT ANALYSIS (PCA)

Principal Component Analysis (PCA) and it's working method: Principal Component Analysis (PCA) is a statistical technique used to reduce the dimensionality of a large dataset. It is a commonly used method in machine learning, data science, and other fields that deal with large datasets. PCA works by identifying patterns in the data and then creating new variables that capture as much of the variation in the data as possible. These new variables, known as principal components, are linear combinations of the original variables in the dataset. The first principal component captures the most variation in the data, the second captures the second most, and so on. The number of principal components created is equal to the number of original variables in the dataset. PCA can be used for a variety of purposes, including data visualization, feature selection, and data compression. In data visualization, PCA can be used to plot high-dimensional data in two or three dimensions, making it easier to interpret. In feature selection, PCA can be used to identify the most important variables in a dataset. In data compression, PCA can be used to reduce the size of a dataset without losing important information. PCA is a powerful tool for data analysis and can help to simplify complex datasets, making them easier to understand and work with. Principal Component Analysis (PCA) is to reduce the dimensionality of a data set by finding a new set of variables, smaller than the original set of variables, retains most of the sample's information and useful for the compression and classification of data. Principal Component Analysis is a technique which uses sophisticated underlying mathematical principles to transform a number of possibly correlated variables into a smaller number of variables called principal components. It is one of the most important results from applied linear algebra. The advantage of PCA is finding the patterns in the data and compressing data by reducing the number of dimensions without loss of information. The mathematical concepts that are used for PCA are Standard Deviation, Variance, Co–variance and Eigenvectors. The database images belonging to same category may differ in lighting conditions, noise etc., but are not completely random and in spite of their differences there may present some patterns. Such patterns could be referred as principal components. PCA is a mathematical tool used to extract principal components of original image data. These principal components may also be referred as Eigen images. An important feature of PCA is that any original image from the image database can be reconstructed by combining the eigen images.

PCA computation : The method to calculate Principal Components of any given matrix is as follows:

a) Represent the image as one-dimensional vector of size N x N. Suppose we have M vectors of size N (= rows of image × columns of image) representing a set of sampled images.
b) The Mean value of the pixels intensities in each image is calculated and subtracted from the corresponding image. The process is continued for all images in the database.
c) The covariance matrix which is of the order N2 x N2 is calculated as given by C = $A.A^T$.
d) Find the Eigen values of the covariance matrix C by solving the equation (Cλ - I) = 0. To find the eigenvector X repeat the procedure where Xi indicates corresponding Eigen values.
e) The Eigen vectors are sorted according to the corresponding Eigen values in descending order.
f) Choose the First 'K' Eigen vectors and Eigen Values.

III. FINDING A NEW VARIABLE

Principal components are then used to create a new variable, which represents the original data in a lower-dimensional space. We first need to compute the covariance matrix of the image matrix. This matrix tells us how the different pixels in the image are related to each other. Next, we need to find the eigenvectors and eigenvalues of the covariance matrix. The eigenvectors are the directions in which the data varies the most, and the eigenvalues represent the amount of variance in the data that is explained by each eigenvector. We can then sort the eigenvectors in descending order of their corresponding eigenvalues. The eigenvector with the highest eigenvalue captures the most important features of the image, and we can use this eigenvector as our new variable.

IV. IMAGE PROCESSING BY MATRIX REDUCTION

Image processing is the technique of manipulating digital images using computer algorithms. The images are usually represented in the form of matrices or arrays of pixel values. The process of reducing the size of these matrices, also known as matrix reduction,

IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

6th National Conference on Science, Technology and Communication Skills – NCSTCS 2K23

Narula Institute of Technology, Agarpara, Kolkata, India

Vol. 10, Special Issue 3, September 2023

plays an important role in image processing. In this article, we will explore the concept of matrix reduction in image processing and its various applications. Matrix reduction refers to the process of reducing the size of a matrix by removing certain elements from it. In the context of image processing, this technique is used to reduce the size of an image without compromising on its quality. There are several methods of matrix reduction, and each method has its own advantages and disadvantages. Some of the commonly used methods are discussed below:

a) Downsampling: Downsampling is a technique of reducing the size of an image by removing some of its pixels. This is done by taking every nth pixel in a row and column and discarding the others. The value of n depends on the desired size reduction of the image. The resulting image is smaller in size but may have a lower resolution and quality.
b) Average Pooling: Average pooling is a technique of reducing the size of an image by taking the average value of a group of pixels. This is done by dividing the image into non-overlapping blocks and taking the average value of each block. The resulting image has a lower resolution but retains the overall structure and features of the original image.
c) Max Pooling: Max pooling is a technique of reducing the size of an image by taking the maximum value of a group of pixels. This is done by dividing the image into non-overlapping blocks and taking the maximum value of each block. The resulting image has a lower resolution but retains the important features of the original image.

Matrix reduction has several applications in image processing, some of which are discussed below:

a) Compression: Matrix reduction is often used to compress images for storage or transmission purposes. By reducing the size of an image, it can be stored or transmitted more efficiently. The compressed image can then be decompressed to its original size without significant loss of quality.
b) Object Recognition: Matrix reduction is often used in object recognition tasks where the size of the image needs to be reduced to a manageable size. This allows the recognition algorithm to process the image more quickly and accurately.
c) Machine Learning: Matrix reduction is often used in machine learning algorithms that use images as input. By reducing the size of the images, the training process can be sped up without significant loss of accuracy.

V. METHOD OF IMAGE COMPRESSION BY USING PCA ALGORITHM

PCA is a powerful algorithm for image compression and can achieve high compression ratios with minimal loss of image quality. However, it is computationally intensive and may not be suitable for real-time compression applications. The idea behind using PCA for image compression is to reduce the dimensionality of the image data by finding the principal components that capture the most variance in the image data. These principal components can then be used to represent the image data in a more compact form, thereby reducing the storage requirements for the image. These are the following steps involved in using PCA for image compression:

a) Convert the image to grayscale: Since PCA works on numerical data, we first need to convert the image to grayscale.
b) Divide the image into blocks: The image is divided into blocks of equal size, usually 8x8 or 16x16 pixels.
c) Subtract the mean: For each block, we subtract the mean value of the pixels in that block from each pixel value. This helps to center the data around zero, which is important for PCA.
d) Compute the covariance matrix: The covariance matrix is computed for the pixel values in each block. The covariance matrix describes the relationships between the pixel values and is used to find the principal components.
e) Compute the principal components: The principal components are computed from the covariance matrix using an eigenvalue decomposition. The principal components capture the most variance in the data and are used to represent the image data in a more compact form.
f) Quantize the principal components: The principal components are quantized to reduce the number of bits needed to represent them. This helps to further reduce the storage requirements for the image.
g) Reconstruct the image: To reconstruct the compressed image, we reverse the process by multiplying the quantized principal components by the corresponding eigenvectors and adding back the mean value.
h) Encode the compressed image: Finally, the compressed image is encoded using a lossless compression algorithm such as Huffman encoding or arithmetic coding.

VI. DISCUSSION

To continue the discussion with this topic, we have taken an image as shown in fig. 1.1. This image is black and white as per our eyes. This reduction is desired to save storage space, improve transmission efficiency. We have performed pca operations on this image and got a new modified image, which is shown in fig. 1.2.

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology
6th National Conference on Science, Technology and Communication Skills – NCSTCS 2K23
Narula Institute of Technology, Agarpara, Kolkata, India
Vol. 10, Special Issue 3, September 2023

Fig. 1.1 Input Image (Black & White) (Before Reduction)

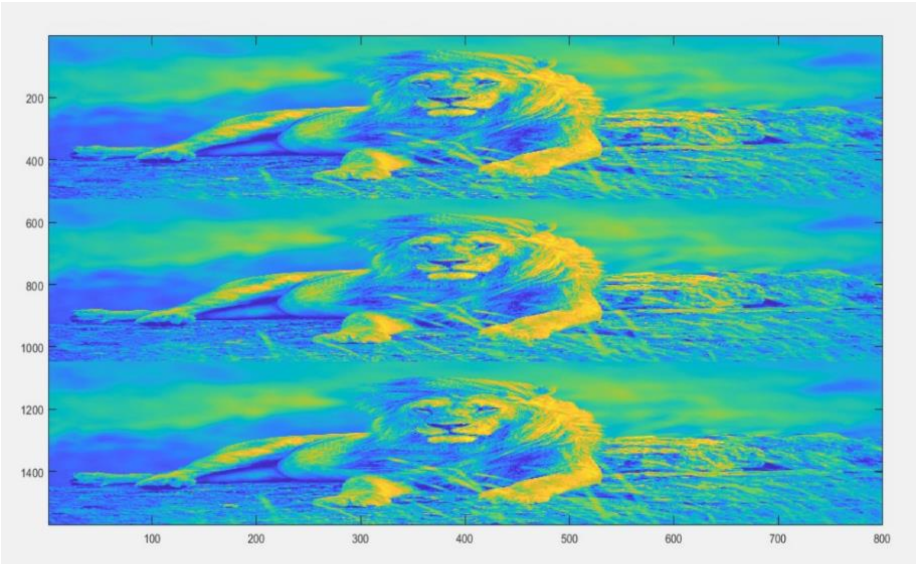

Fig. 1.2 Output Image (After Reduction)

The image (fig 1.2) is not the required image. This image is split into three layers of RGB image format. As we want a single reduced image, so we are working further with this topic regarding turning back the reduced matrix into an original image, but for that we need better resolution and all information intact. For that we need higher knowledge of the Euclidean algorithm to continue with this topic.

The Euclidean algorithm is an ancient and fundamental algorithm used to find the greatest common divisor (GCD) of two integers. The GCD of two integers is the largest positive integer that divides both numbers without leaving a remainder.

The Euclidean algorithm provides an efficient method for finding this GCD. Using Euclidean algorithm of GCD, we are trying to find out the image matrices of highest value which will be then converted to image using programming languages. But this requires higher level of studies, so we are aiming to do it on a later period of time. For that we are trying our best to try out new things and explore further.

REFERENCES

[1]. Xiao Z. (2013, September). An Image Compression Algorithm Based on PCA and JL. Journal of Image and Graphics, 1(3).
[2]. Navaneethakrishnan R. (2012, July). Study of Image Compression Techniques. International Journal of Scientific Engineering Research, 3(7).
[3]. Pandey P. K., Singh Y. and Tripathi S. (2011, February). Image Processing using Principal Component Analysis. International Journal of Computer Applications, 15.

[4]. Stolevski S. (2010). Hybrid PCA Algorithm for Image Compression. 18th Telecommunications forum TELFOR.
[5]. Mudrova M. and Prochazka A. (2005). Principal Component Analysis in Image Processing. Institute of Chemical Technology, Prague Department of Computing and Control Engineering.
[6]. Kaur P. (2012). Implementation of Principle Component Analysis with Fuzzy Annotation for CAD Jewellery Images. International Journal of Emerging Trends & Technology in Computer Science, 1(4).
[7]. Mehrdoust F. (2012, November). A hybrid randomized algorithm for image compression. International Journal of Applied Mathematical Research, 2(1).
[8]. Agarwal A., Ghazawi T. E., Askary H. E. and Moigne J. L. (2007). Efficient Hierarchical-PCA Dimension for Hyperspectral Imagery. IEEE International Symposium on Signal Processing and Information Technology.
[9]. Siddique A., Hamid I., Li W. and Gilani S. M. (2019). Image Representation Using Variants of Principal Component Analysis: A Comparative Study. IEEE International Conference on Communication Software and Networks.
[10]. Bloemendal A., Knowles A., Yau H. T. and Yin J. (2015). On the principal components of sample covariance matrices. Probability Theory and Related Fields, 164, 459–552.
[11]. Naik G. R. and Kumar D. K. (2011). An overview of independent component analysis and its applications. Informatica, vol. 35(1), 63–81.
[12]. Smith L. I. (2002). A tutorial on Principal Components Analysis. in Elementary linear algebra, New York: Wiley.
[13]. Patil U. and Mudengudi U. (2011). Image fusion using Hierarchical PCA in Image