IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Issue-9, Sept 2019 E-ISSN: 2347-2693

Image Compression and Detection Technique Using Principal Component Analysis

Saif Ali[1*], Manish Sharma[2]

[1]Department of Computer Science, Suresh Gyan Vihar University, Jaipur, India
[2]Department of Computer Science, Suresh Gyan Vihar University, Jaipur, India

[*]*Corresponding Author: saifsgvu@gmail.com, Tel.: +91-89479-91340*

DOI: https://doi.org/10.26438/ijcse/v7i9.1316 | Available online at: www.ijcseonline.org

Accepted: 08/Sept/2019, Published: 30/Sept/2019

***Abstract*—**This paper mainly presents face recognition system based on principal component analysis. The goal is to implement the system which is able to distinguish a single face from the larger database. In this research work we are compressing the image using the mathematical tool principal component analysis and then recognize the image from the same data set by the model. First we will describe the basic concepts prevailing with principal component analysis. Then we will see that how principal component can be extracted from a given data set. Then we will go for sampling distribution of Eigen values and Eigen vectors. Then followed by model adequacy test, then we perform our task of image detection. The problem arises when we use high dimensionality space. Because in face or in 3d image, we have different eigen values or vectors and it can't be fixed due to high dimensions as compared to 2d image. Hence, we use Principal Component Analysis (PCA).

***Keywords*—**PCA, Eigen values, Eigen vectors, image compression. Dimension reduction

I. INTRODUCTION

In recent years, with the development of computer technologies and their applications, today's digital world deals with large amount of digital images. Research on image compression has been carried out for a long time, with the aim of reducing the size of images for efficient storage and fast transmission. The visually irrelevant information refers to the information that is not perceived by human eyes [1]. Irrelevancy reduction thus aims at removing certain information from the image that is not perceptually visible to Human Visual System (HVS). The compression is lossless if the redundancy and irrelevancy reduction do not result in any loss of information in the original image. In loss compression, some amount of loss of information is tolerable with an acceptable degradation in the compressed images [2]. The aim of data reduction techniques using PCA is to provide an efficient representation of the data by removing the dependencies in the data. PCA is also called KLT, named after Kari Karhunen and Michel Loeve or Hotelling Transform [2], [5]. KLT, consists a mapping of higher dimensional input space to a lower dimensional representation space by means of linear transformation [6-9]. There are basically two approach to perform PCA: Classical Statistical method and Artificial Neural Network method. Statistical method includes finding of eigenvalues and corresponding eigenvectors of the data set using covariance matrix [10].

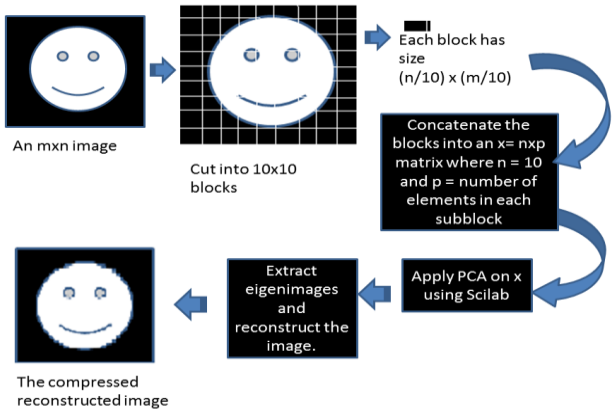

Figure 1. Block Diagram of PCA

The corresponding eigenvalues give an indication of amount of information that the respective principal components represent. The dimension reduction is achieved by using the eigenvectors corresponding to significant eigenvalues [11, 13] as a new orthonormal basis or by using Singular Value Decomposition (SVD).

Fig.1 is showing the block diagram of the image compression and extraction technique. Where image is divided into 10 x 10 blocks and each block size was n/10 x m/10. Then by applying PCA, size of the image is reduced and it is again

recovered by the reconstruction filter. But in this process some information has lost.

Fig. 2 is showing the data has maximum variance in x1 and x2 direction but if we transform the axis in z1 and z2 direction we found the maximum variance in z1 direction and less variation in z2 direction so we can remove the z2 dimension from our data set.

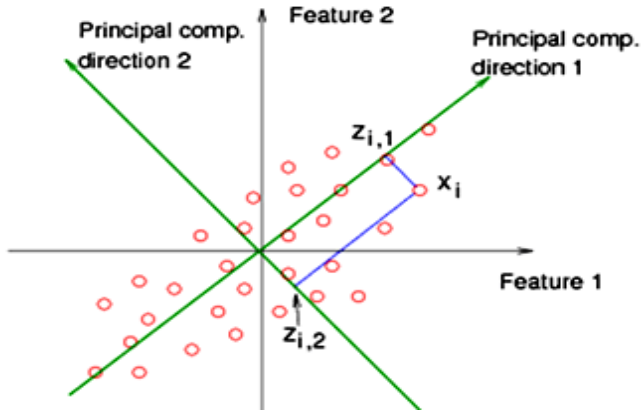

Figure 2. Transformation of dimensions

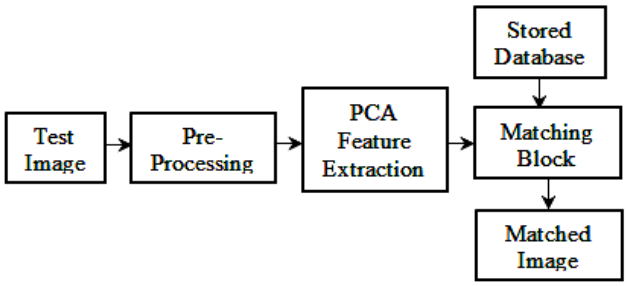

Figure 3. Block diagram of working principal

As a result, what will happen what are the advantages we will be getting from here advantages is, suppose we want to do a prediction model using multiple regression and X variable, these are all x_i independent variables. If x_i are correlate then that ultimately leads to multi co-linearity problem multi co-linearity are there. So, under multi co-linearity condition the regression model what you want to fit that y = f(x) linear model. This model will not be a good one, because under multi co linearity it may not be possible to estimate the parameters.

II. RELATED WORK

A lot of work has been done in spatial domain but image compression is more efficient if the images are processed in transformed domain, that can be achieved via non-redundant and invertible transformations such as Discrete Cosine Transformation (DCT) [3], Karhunen-Loeve Transform (KLT) [4] or Principal Component Analysis (PCA) [5] etc.

III. METHODOLOGY

This paper validates the PCA algorithm on images for the first time where it has not been tested on images before. It also proposes a new learning method for face detection and recognition using the PCA algorithms. The PCA algorithm starts with the creation of a data set (in this paper, it contains training faces and ends with the projection of the data on the eigenspace. A covariance matrix will compute for the data; in addition, the eigenvectors and eigenvalues of the covariance matrix will also obtain. The following working steps are involved here. Collection of images, Feature reduction by PCA, Designing of model, Training of model and Testing of model.

Face is a complex multidimensional structure and needs good computing techniques for recognition. Features extracted from a face are processed and compared with similarly processed faces present in the database. If a face is recognized it is known or the system may demonstrate an alike features accessible in database else it is unknown. In observation organization if an indefinite features appears more than one time then it is stored in database for further recognition. The above steps can be used or image of any criminal identification. Face recognition is the challenge of classifying whose face is in an input image. With face recognition, we need an existing database of faces. Given a new image of a face, we need to report the person's name. The problem arise when we use high dimensionality space. Because in face or in 3D image, we have different eigen values or vectors and it can't be fixed due to high dimensions as compared to IInd image. Hence, we use Principal Component Analysis (PCA). The major advantage of PCA is using it in eigen face approach which helps in reducing the size of the database for recognition of a test images.

IV. RESULTS AND DISCUSSION

To understand the principal of working model first we understand the data set which we have used in our research work. WE have 400 number of samples, each image size is 64 by 64. There are 4096 features as input and40 output classes.

- Import the data set from fetch Olivetti faces from sklearn datasets.
- Slicing data by training, testing and splitting from sklearn model selection library.
- Plotting the gray scale image of mean image andimages with unique Id's.

For image compression and extraction we have use the Principal Component Analysis (PCA) by Eigen value, Eigen vectors Covariance Matrix and K-Nearest neighbor. In the result part first we draw the mean figure which is showing in Fig.4. In this research work we are compressing the image

using the mathematical tool principal component analysis and then recognize the image from the same data set by the model. Here in this section we will discuss about our data set and the outcomes of our model.

```
In [5]: mean_image = np.mean(X_train,axis=0)
        plt.imshow(mean_image.reshape((64,64)), cmap=plt.cm.gray)
Out[5]: <matplotlib.image.AxesImage at 0x26678931390>
```

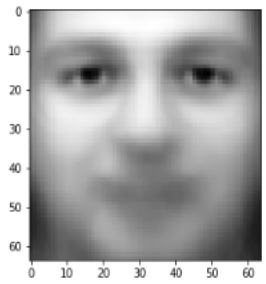

Figure 4. Mean face composed from all the faces in dataset

Figure 5. Original Images from dataset.

Fig 5 is the collection of all images and then by algorithm we have design the mean image showing in the Fig.4. After applying the principal component we have reduce the size of the images by eliminating highly related components. Here we are losing some information and it is approximately 20% information of the actual image.

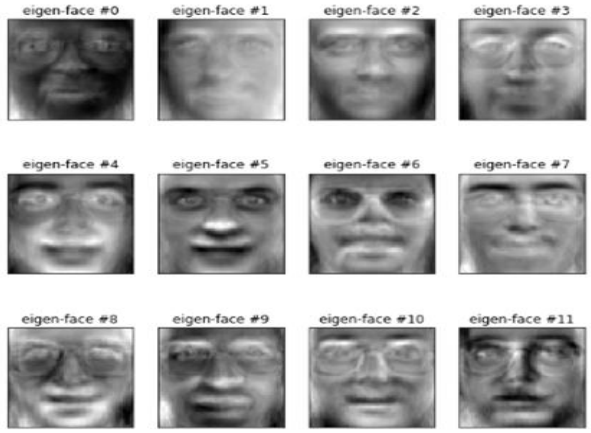

Figure 6. Eigen Faces after applying PCA algorithm

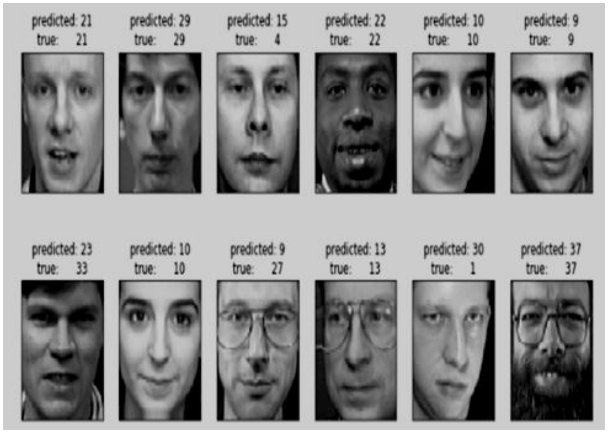

Figure 7. Image detected by the model.

Fig 7 is showing the result of some randomly selected images from data set. By the result we can see that image number 21, 29, 22, 10, 9, 13 and 37 are predicted correctly and image number 4, 33, 27, and 1 are predicted wrongly. The overall accuracy of the system is 57%. By the above experimental results it is great to see that we can save the images in the lower dimension using PCA or SVD algorithms and can reconstruct them with some amount of loss. But overall it's a great technique and used in the machine learning domain where computation time is a big challenge.

V. CONCLUSION AND FUTURE SCOPE

In this research work we are compressing the image using the mathematical tool principal component analysis and then recognize the image from the same data set by the model**.** First we will describe the basic concepts prevailing with principal component analysis. Then we will see that how principal component can be extracted from a given data set. Then we will go for sampling distribution of Eigen values and Eigen vectors.

Accuracy: 0.57

	precision	recall	f1-score	support
0	0.00	0.00	0.00	4
1	0.00	0.00	0.00	2
2	1.00	0.50	0.67	2
3	1.00	0.25	0.40	4
4	0.50	1.00	0.67	3
5	0.75	1.00	0.86	3
6	0.00	0.00	0.00	1
7	1.00	0.29	0.44	7
8	0.33	1.00	0.50	2
9	1.00	1.00	1.00	3
10	1.00	1.00	1.00	3
11	1.00	0.25	0.40	4
12	1.00	1.00	1.00	2
13	1.00	1.00	1.00	1
14	0.25	1.00	0.40	3
15	1.00	0.50	0.67	2
17	0.67	0.67	0.67	3
18	0.67	1.00	0.80	2

Then followed by model adequacy test, then we perform our task of image detection. The problem arises when we use high dimensionality space. Because in face or in 3d image, we have different eigen values or vectors and it can't be fixed due to high dimensions as compared to 2d image. Hence, we use Principal Component Analysis (PCA). After applying the principal component we have reduce the size of the images by eliminating highly related components. Here we are losing approximately 20% information of the image. The accuracy of the system is 57 %

REFERENCES

[1]. L. I. Smith, *"A tutorial on principal components analysis,"* February 2002. 537

[2]. S. F. Ding, Z. Z. Shi, Y. Liang, and F. X. Jin, "*Information feature analysis and improved algorithm of PCA*" International Conference on Machine Learning and Cybernetics. **vol.3**, Aug 2005**, pp. 1756–1761, 2005.**

[3]. A. A. Alorf, "*Comparison of computer-based and optical face recognition paradigms*" **vol.4,** pp **538-545**, 2014.

[4]. P. Kamencay, D. Jelovka, and M. Zachariasova, "*The impact of segmentation on face recognition using the principal component analysis (PCA)*" Signal Processing Algorithms, Architectures, Arrangements, and Applications Conference Proceedings (SPA), **Sept 2011**, pp. **1–4. 2011**

[5]. B. C. Russell, W. T. Freeman, A. A. Efros, J. Sivic, and A. Zisserman, "*Using multiple segmentations to discover objects and their extent in image collections*" in Computer Vision and Pattern Recognition, IEEE Computer Society Conference on, **vol. 2**, pp. **1605–1614**, **2006.**

[6]. S. He, J. Ni, L.Wu, H.Wei, and S. Zhao, "*Image threshold segmentation method with 2-D histogram based on multi-resolution analysis*" in Computer Science Education, 2009. ICCSE. 4th International Conference, pp. **753–757**. July **2009**

[7]. I. Kim, J. H. Shim, and J. Yang, "*Face detection*" Face Detection Project, EE368, Stanford University, **vol. 28**, pp **538, 539, 2003**

[8]. R. Gonzalez, R. Woods, and S. Eddins, "*Digital Image Processing Using MATLAB"*. Gates mark, ISBN: **978-0-9820854-0-0. 538, 540, 543, 2009.**

[9]. M. Turk and A. Pentland, "*Face recognition using eigenfaces*," in Proceedings of Computer Vision and Pattern Recognition, 1991. pp. **586–591**, ISSN: 1063-6919. 538, **1991**.

[10]. J. Cognitive Neuroscience, "*Eigenfaces for recognition*" vol. **3,** pp. **71–86**, Jan. 1991.

[11]. W. Yang, C. Sun, L. Zhang, and K. Ricanek, "*Laplacian bidirectional pca for face recognition,"* Neurocomputing, **vol.** 74, no. 1, pp. **487–493, 2010**.

[12]. M. Al-Amin, "*Towards face recognition using eigenface*," International Journal of Advanced Computer Science and Applications, **vol. 7**, pp **539, 544**, **2016.**

[13]. Q. Fu, "*Research and implementation of pca face recognition algorithm based on matlab," in MATEC Web of Conferences,* **vol. 22**. **539-544.** EDP Sciences, **2015**

Authors Profile

Mr Saif Ali pursued Bachelor of Engineering and Technology in Computer Science and Engineering from Suresh Gyan Vihar University, India in 2017. He is currently pursuing Master of Technology in Computer Science from Suresh Gyan Vihar University. His main research work focuses on Machine Learning, Deep Learning.

Dr. Manish Sharma is Associate professor of Department of Computer Science and Engineering in Suresh Gyan Vihar University, India. He has done Master of Technology in Computer Science and Engineering. He has experience of 13 years in his field of Computer Science.

