International Journal of
Financial and Economic Studies

Academixs Open Access Publishers

Research Article

Dimensionality Reduction for Image Analysis

Rohan J. Dalpatadu[1*], Dwaipayan Mukhopadhyay[2], Lakshmi P. Gewali[3] and Ashok K. Singh[4]

[1]*Department of Mathematical Sciences, University of Nevada Las Vegas, USA*

[2]*Department of Epidemiology & Biostatistics, University of Nevada Las Vegas, USA*

[3]*Department of Computer Science, University of Nevada Las Vegas, USA*

[4]*Resorts, Gaming & Golf Management, William F. Harrah College of Hospitality, University of Nevada Las Vegas, USA*

Corresponding Author:
Rohan J. Dalpatadu. Department of Mathematical Sciences, University of Nevada Las Vegas, USA.

Received Date: 17.10.2025 **Accepted Date:** 17.10.2025 **Published Date:** 03.11.2025

Abstract
In this article, we consider dimensionality reduction methods for image processing. Various dimensionality reduction methods exist in the literature, but we only consider the unsupervised statistical method of Principal Component Analysis (PCA). We show how the dimensionality of an image can be reduced via PCA without losing resolution of the image. For this purpose, a jpeg image of a hummingbird is downloaded and read in R. PCA is next performed of the large image dataset and the mean square error (mse) method is used for selection of number of components retained. This article shows that a clear image of the hummingbird can be obtained by just using first three or four PC-scores.

Key Words: Dimensionality Reduction, Intensity, Principle Component Analysis, Variation

Introduction

Amount of literature on Principal Component Analysis is very large. A vast amount of literature also exists for the discipline of image processing [1-7]. In this article, we demonstrate via a simple example how PCA can be used to reduce the dimensionality of a large image fille.

Description of Principal Component Analysis

Principal Component Analysis (PCA) is a statistical technique primarily used for dimensionality reduction of a dataset by transforming a large set of variables into a smaller set of linear combinations of the original variables (principal components). The principal components are orthogonal to each other and the first PC (PC1) explains the largest variation in the entire dataset; the second PC (PC2) explains the second largest variation in the entire dataset, with all p components cumulatively explaining 100% of variance of the dataset, where p is the total number of variables in the original dataset. The method of scree-plot is used to find the minimum number of principal components which explain a majority of total variance. The PC scores are the values of the PCs for each row in the dataset, and the principal components are the transformed (new) variables. One particular method of performing PCA is eigen analysis of the covariance matrix or the correlation matrix of the given dataset [8].

Data and Results

A picture of a hummingbird, shown in Fig. 1, is downloaded from the web and the dimensionality of the image dataset is 2000 x 3000 x 3, with 2000 rows and 3000 columns or variables; 3 represent the RGB (RED-GREEN-BLUE) values. Each pixel of the image represented by three (3) pixels instead of one (1). In a black & white image, each pixel would be represented by [pixel], In RGB image each pixel would be represented by [pixel(R), pixel(G), pixel(B)]. Each pixel of the image has 3 RGB values. These range between 0 and 255 and represent the intensity of Red, Green, and Blue. A lower value stands for higher intensity and a higher value for lower intensity; the downloaded file has the third column norrmalized to (0,1). For instance, one pixel can be represented as a list of these three values [78, 136, 60], with [0, 0, 0] representing Black.

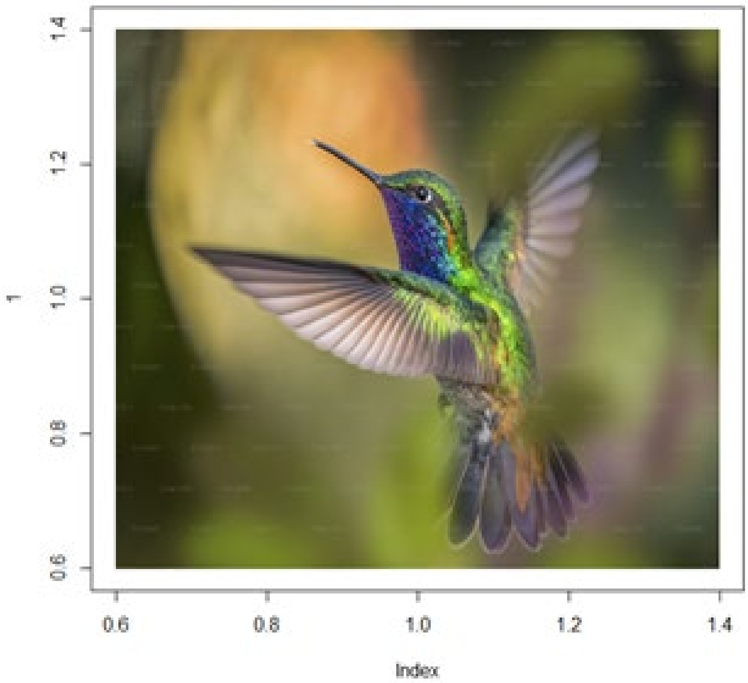

Figure 1: Original image of a hummingbird

In the above color image, there are three matrices pixel by pixel corresponding to the three components of RGB color. Any color image can easily be converted to grayscale by adding the RGB shades and then dividing the total by maximum value to convert to (0,1) scale (see Figure 2). All computations and visualizations are done in the software environment R (2024). The following R-packages were used to perform PCA and image analysis: jpeg, factoextra, gridExtra, ggplot2, magick, imgpalr, Metrics.

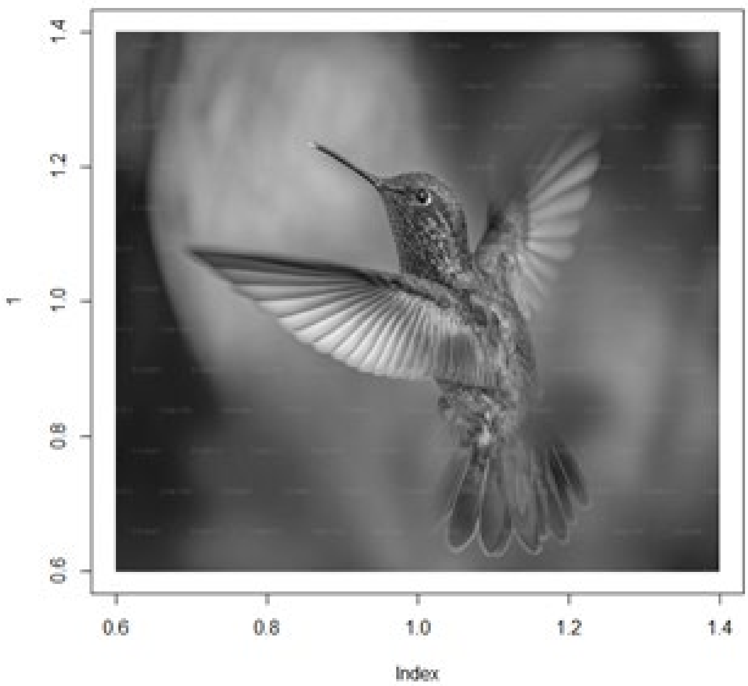

Figure 2: Gray-scaled image of the humming bird

Figure 3, obtained from the R-package factoextra shows the increase in percentage explained for each color (R, G, B) as the number of PC's are increased. It can be seen from Fig. 3 that we can just use first three or four components out of the total of 1000 components.

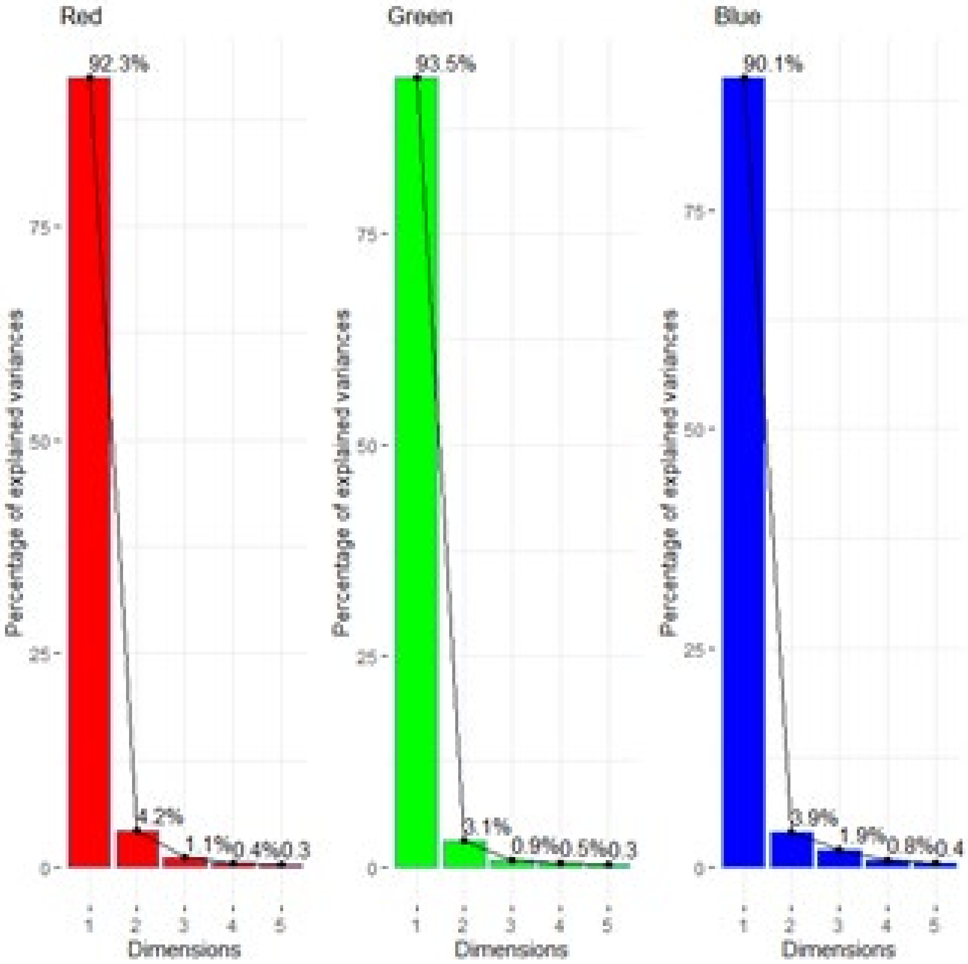

Figure 3: Percentage explained by dimensions used

Table 1 shows that the image reduction obtained when the number of principal components is increased in plotting the hummingbird picture. It is clear from Figure 5 and Table 1 that three components yield a clear picture of the hummingbird and also provides 55.1% compression of the image.

Figure 4 shows the image obtained by just using the first PC for each of the three colors. Figure 5 (next page) shows images obtained by increasing the number of components.

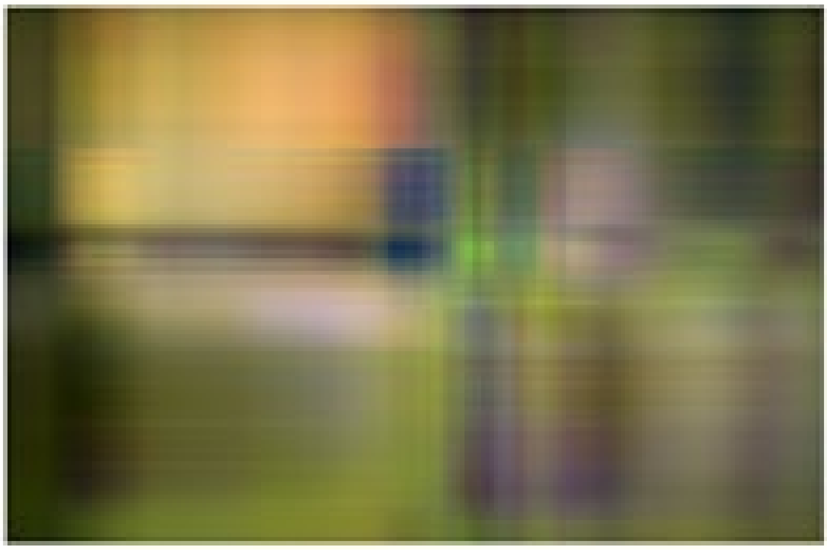

Figure 4: Hummingbird picture just using the first PC

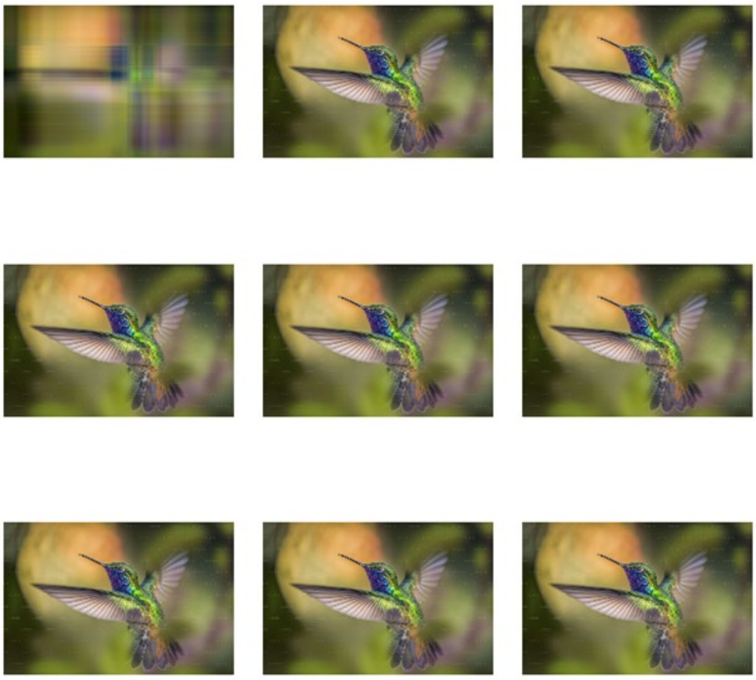

Figure 5: Hummingbird pictures using increasing number of PC's used

	Number of PC	Photo size	Compression ratio	MSE
1	3	166974	0.551	0.0050036
2	253	267193	0.882	0.0000598
3	502	282824	0.933	0.0000414
4	752	296283	0.978	0.0000371
5	1002	301832	0.996	0.0000352
6	1251	302987	1	0.0000338
7	1501	303002	1	0.0000335
8	1750	303002	1	0.0000335
9	2000	303002	1	0.0000335

Table 1: Image compression ratio as a function of number of principal components used

Conclusion

It is clear from figures 4-5 that just using one principal component yields a very blurry image in which the hummingbird cannot be seen; it can also be seen that using three or four principal components are sufficient to produce a clear image of the hummingbird.

References

1. Garzon, M., Yang, C. C., & Deng, L. Y. (2022). What Is Data Science (DS)? In Dimensionality Reduction in Data Science (pp. 1-28). Cham: Springer International Publishing.
2. Kassambara, A. (2017). Practical guide to principal component methods in R: PCA. M (CA), FAMD, MFA, HCPC, factoextra, 2, 12-50.
3. J. A. Rad, S. Chakraverty, and K. Parand; editors (2025). Dimensionality Reduction in Machine Learning, Morgan Kauffman, Cambridge, MAR Core Team (2024). _R: A Language
4. Sorzano, C. O. S., Vargas, J., & Montano, A. P. (2014). A survey of dimensionality reduction techniques. arXiv preprint arXiv:1403.2877.
5. Awcock, G. J., & Thomas, R. (1995). Applied image processing.
6. Prats-Montalbán, J. M., de Juan, A., & Ferrer, A. (2011). Multivariate image analysis: A review with applications. Chemometrics and intelligent laboratory systems, 107(1), 1-23.
7. Teh, C. H., & Chin, R. T. (2002). On image analysis by the methods of moments. IEEE Transactions on pattern analysis and machine intelligence, 10(4), 496-513.
8. Jolliffe, I. (2011). Principal component analysis. In International encyclopedia of statistical science (pp. 1094-1096). Springer, Berlin, Heidelberg.

Citation: Rohan J. Dalpatadu,. Dwaipayan Mukhopadhyay,. Lakshmi P. Gewali,. Ashok K. Singh,. et al. (2025). Dimensionality Reduction for Image Analysis. *Int. J. Financ. Econ. Stud. 1*(1), 1-4.