

CEP Project, Satellite Image Processing

**Members:**

1. Syed Muhammad Ashhar Shah, 2020478
2. Khizar Ali Shah, 2020196
3. Arbaz Khan, 2020
4. Isaam Ul Haq, 2020

*This paper deals with the process of analyzing a multispectral Landsat image captured by NASA, obtained from a satellite. In today’s time we use satellite images to perform various tasks such as tracking the measuring different types of activities performed by humans or natural sources.*

*Furthermore, we also study the affects of principle component analysis on an image which can be used for a huge range of things such as image compression, image reconstruction, recognition algorithms or efficient data storage.*

Keywords: principal component analysis, image reconstruction, landstat, satellite image, image processing.

# Introduction

In this study we investigate the relationship on how variance in the number of principle components of an image or a dataset affects the overall quality of the image. In addition to PCA we also analyze the overall bands of an image that is captured by a satellite.

We also perform various operations on multispectral images such as scaling, and concatenation of the individual bands obtained to form a high-quality visual satellite image.

Furthermore, we also perform error analysis on the use of principle component analysis and observe the relationship of the amount of error against the number of selected principal components of an image due to which we can find out how many principal components are required to display an image without disrupting its originality to a huge extent.

# Ease of Use

## Planet Activity Analysis

The landstat images obtained from a satellite are high quality, multi spectral images (contain multiple bands) of the earth’s surface. Hence using these images, we can perform analysis on the different activities happening anywhere on the planet such as tracking temperature changes or natural disasters.

## Disaster Predictions and Prevention

Using these satellite images, we can also use powerful machine learning algorithms to predict any natural or man-made disaster that may occur and hence make the necessary preparation in order to avoid such a disaster to cause high intensity catastrophic damage to life or property.

## Data Interpretation and Simplicity

We also use the technique of principle component analysis to interpret the data to search for any important patterns in a dataset. This technique also helps to make sense of the data we receive from the multispectral landstat images.

# Additional Data and information

## Abbreviations and Acronyms

* PCA: Principal Component Analysis
* CEP: Complex Engineering Problem
* NASA: National Aeronautics and Space Administration

## Units

* We use “pixels” to measure the resolution of an image as the data on which we perform image analysis operations are stored in a two-dimensional array which represent the (x, y) location of a pixel and hence the overall resolution of an image.

## Equations

*S = D / 255*   
 *X = X - µ* 2  
 *T = (T. C) + µ* 2

The equation (1) represents the standardization equation as the data we have has different pixel values that range from 0-255 which represent the color density of a pixel. We use this equation (1) to scale all the values from 0-1 so that no component is dominant over the other when we apply PCA.

The equation (2) here is used for application of PCA as we must subtract the mean of the data from each entry of the data so that we can calculate the eigenvalues and eigenvectors.

The equation (3) here represents transformation of a PCA applied dataset back to the original dataset format by performing a dot product of the PCA applied data against the selected eigen vectors and adding back the subtracted mean.

## Some Common Mistakes

* The word “data” is plural, not singular. When we use the word data, we mean a collection of data of an image.
* Here “transformation” means obtain the original dimension of the image so that we can merge them together to form a dimensionally reduced image. Don’t confuse it with the data we get by simply fitting PCA.
* When we mention “Standardization” we mean that we can’t directly apply PCA to any data. We must observe the data first to make sure that the data is linearly distributed so that we can recover the original data from the transformed data.
* When we say “reconstructing” we also mean reducing the dimensions of the image data while still maintaining the principal properties of the original image hence reducing the image size without actual data loss. Don’t confuse it with lossy reconstruction.

# methodology

Satellite image analysis and dimensional reduction using PCA, a growing area of interest among individuals nowadays due to its prediction properties and lossless compression of data for efficient storage. A satellite image is basically a multispectral image that contains different bands of an original image captured by a NASA satellite. We also use dimensional reduction using one of the popular techniques known as PCA.

As we know we satellite images are high quality and are highly expensive to compute and store in huge volumes hence we can use the technique of PCA as a way to detect the different patterns in the large and complex data obtained from the satellite images. Hence, majority of data is transmitted into a handful of principle components which convey meaning to the large and complex data, doing this also reduced the number of dimensions of the original image without discarding any of the dimensions hence no actual data loss takes place and at the same time the overall data is reduced.

All the data that we use in this research is collected from NASA’s official servers which are open to the public. We only use a single landstat image of Pakistan as our dataset on which various image operations along with PCA is applied for dimension reduction.

The tool of our choice was **Jupyter Lab** and **Python** as is a powerful programming language and has a wide range of libraries both of mathematical computation along with image analysis.

## Landstat Image Used

Above is the landstat image that we have used. This is an image consisting of majority of Punjab, Pakistan. The Scene ID corresponding to this image is “”. In the project we have applied various operations of band concatenation and principal component analysis on this image.

## Individual bands of an image.

Here is a list of all the bands that we shall use for concatenation in an attempt to form the original-colored image and perform basic image related operation on it. Above is a list of all the bands that we have collected from NASA’s website and below is a list of the bands which we shall use for concatenation. We have ignored some of the bands due to the difference in dimensional values of individual bands. The selected bands all have dimensions of 30 whereas the others differ from it.

## Principle Component Analysis Function

Above we have the python function that we have coded for performing the task of PCA. In the function we have used the scikit-learn SVD and the NumPy library to compute and find the eigen vectors and the eigen values along with performing various matrix related operation to receive a dimensionally reduced image.

## Mean Squared Error Function

This is an illustration of the mean squared error function that we have developed for our project that takes two images in the form of lists/arrays and then uses the NumPy library to calculate the mean squared error between the two image where one of the images is the original image that is untouched, PCA not applied, while the other is the one on which we have applied PCA. Furthermore, we will compare the channels with each other rather than the RGB images to get an accurate approximation of the error between the two.

## Channels of the image.

Here you can see the three channels, Red, Blue and Green, of the colored image on which we shall apply PCA for our selected value of components. We will first apply PCA to these individual channels/bands and then concatenate them together to form an dimensionally reduced image similar to the original image but with the number of principle components according to our choice.

# Results

Satellite image analysis and dimensional reduction using PCA, a growing area of interest among individuals nowadays due to its prediction properties and lossless compression of data for efficient storage. A satellite image is basically a multispectral image that contains different bands of an original image captured by a NASA satellite. We also use dimensional reduction using one of the popular techniques known as PCA.

As we know we satellite images are high quality and are highly expensive to compute and store in huge volumes hence we can use the technique of PCA as a way to detect the different patterns in the large and complex data obtained from the satellite images. Hence, majority of data is transmitted into a handful of principle components which convey meaning to the large and complex data, doing this also reduced the number of dimensions of the original image without discarding any of the dimensions hence no actual data loss takes place and at the same time the overall data is reduced.

##### Acknowledgment *(Heading 5)*

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g”. Avoid the stilted expression “one of us (R. B. G.) thanks ...”. Instead, try “R. B. G. thanks...”. Put sponsor acknowledgments in the unnumbered footnote on the first page.

##### References

The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use “Ref. [3]” or “reference [3]” except at the beginning of a sentence: “Reference [3] was the first ...”

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the abstract or reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors’ names; do not use “et al.”. Papers that have not been published, even if they have been submitted for publication, should be cited as “unpublished” [4]. Papers that have been accepted for publication should be cited as “in press” [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

1. G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*
2. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
3. I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
4. K. Elissa, “Title of paper if known,” unpublished.
5. R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.
6. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
7. M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.

**IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove template text from your paper may result in your paper not being published.**