Principal Component Analysis in Application to Identifying Composition of Glacial Moraines

Amita Anand, Georgia Institute of Technology

EAS 4380 Term Project Proposal

Background and Motivation

The New Horizons mission to Pluto launched in 2006 has provided the highest resolution images of Pluto ever seen before. These high-quality images provide answers to some questions about the composition of Pluto but also prompt many more. An area of interest on Pluto's surface now more clearly visible is the Tombaugh Regio located north of the equator. This area has 2 lobes with different surface properties. The western lobe, known as the Sputnik Planitia (SP), is made up of icy N_2 glaciers which indicate glacial flow based on low viscosity of N_2 ice at Pluto's surface temperature, spectral evidence of nitrogen, as well topological evidence of flow lines **[1]**. While the composition of the larger glacier features has been able to be determined, images indicate dark lines of materials cutting across the planes with unknown composition. The evidence of N_2 glacier flow suggest these dark lines could be analogous to glacier moraines like those observed on Earth **[2]**. Glacial moraines are formed when glaciers retreat and leave behind rocky material. Rapid climate change and global warming has increased glacial retreat which makes moraines more prominent and easily observable on Earth.

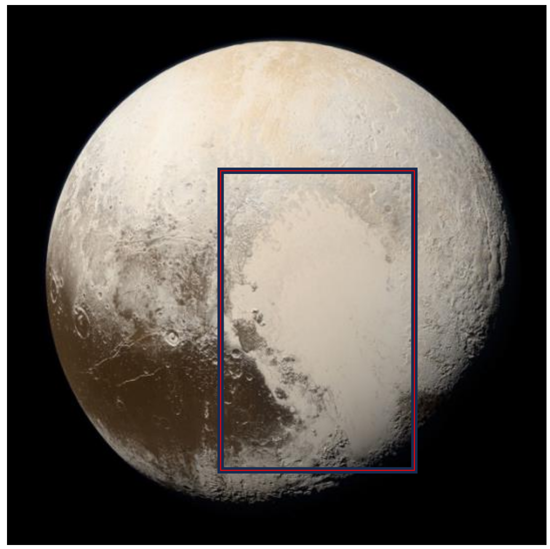


Figure 1 Pluto's Sputnik Planitia NASA/JHAPL

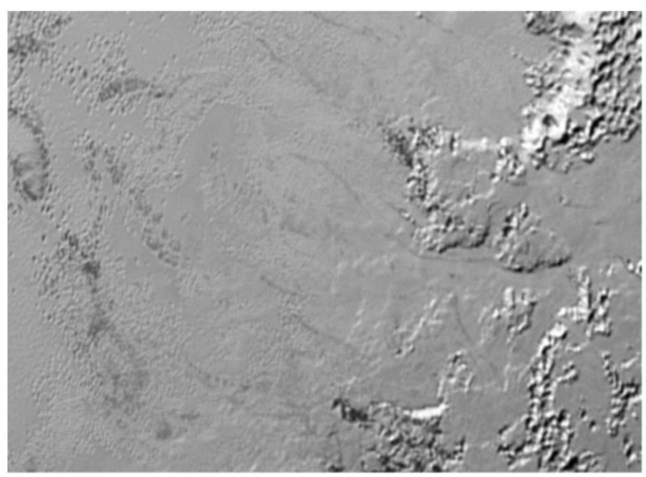


Figure 2 Pluto's Icy plains NASA/JHAPL

Water ice and nitrogen ice behave very differently, furthermore Pluto has a drastically different atmosphere than Earth, so determining what the composition of these dark lines is proves challenging. Additionally, while New Horizons has provided the highest resolution images of Pluto to date, there are still limits as to what can be observed. Specifically, the compositions of deposits on the surface of Pluto have yet to be identified partly because of the image resolution problem **[1]**.

As stated, the N_2 glacial flow on Pluto could cause moraines similar to those observed on Earth. One of the largest glacier systems on Earth is in the Swiss Alps and within this system is Gorner Glacier which has been highly researched. Gorner Glacier displays lateral moraines which are most like what has been observed on Pluto **[1]**. Gorner Glacier is melting quickly due to climate change which makes it an ideal candidate for studying moraines. Furthermore, there have been many field expeditions to the glacier so there is robust ground truth making it a great candidate for remote sensing.

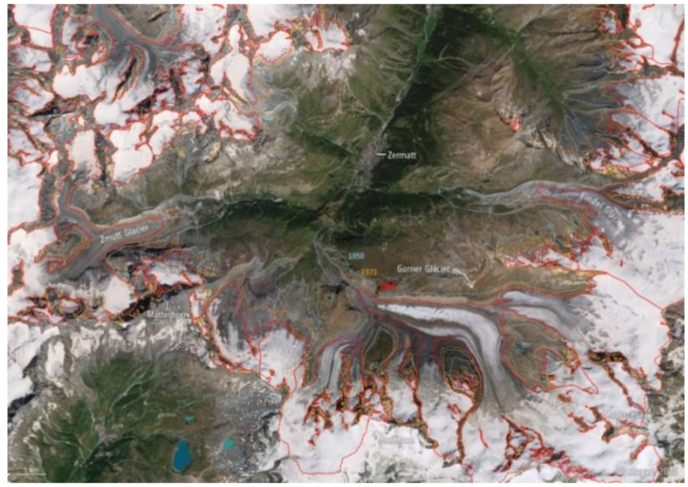


Figure 3 Muller et. al, 1976 and Maisch et. al, 2000 [7]

Scientific Objectives

Determining the composition of the dark materials present on Pluto requires extremely precise and high-resolution images which are not currently available. To determine just how much of the composition can be resolved with the existing data, I propose applying Principal Component Analysis (PCA) to remote sensing data of Gorner Glacier from Landsat-8. With this dimensionality reduced data, I will attempt map the distribution of dolomite by taking band ratios. The success of this experiment would indicate the minimum dimension of data required to resolve surface composition features. Using the dimensionality reduced images as training data, I will build a convolutional neural network (CNN) to identify and classify the materials in images of glacial moraines. Then, the model can be applied to images from New Horizon to identify the composition of the dark materials on Pluto.

Technical Approach and Methodology

PCA is a common dimensionality reduction technique that preserves the variance of a dataset. The data is first standardized by subtracting the mean and then divided by the variance for each band, so the data has mean 0 and variance 1. The covariance matrix for the dataset is then calculated. Each entry of the covariance matrix is the covariance between each band corresponding to that row and column, and along the diagonal the entries are all 1. Next, the eigenvalues and eigenvectors of this matrix are calculated and makeup the principal components. The eigenvalues are listed in descending order of significance, so the largest eigenvalue accounts for the highest percentage of variance. The eigenvector corresponding to this eigenvalue is the direction of

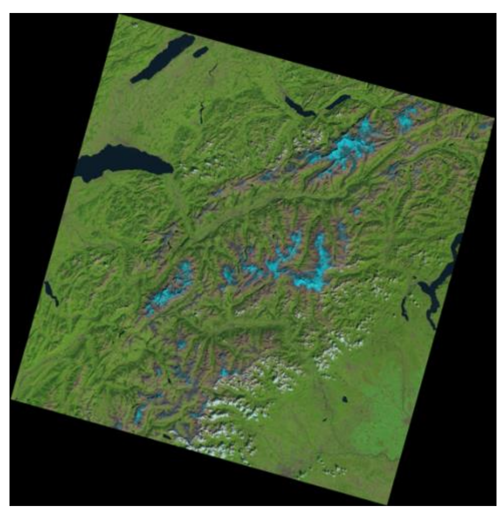


Figure 4 Landsat-8 Scene 08/30/2015

highest variance. Furthermore, since the covariance matrix is symmetric, the eigenvectors are all orthogonal to each other. These eigenvectors are the principal components (PCs). Essentially, PCA linearly transforms the data to an alternative coordinate system where each dimension is a measure of variance. Thus, the first n-components account for a certain percentage of the total variance of the data. After applying PCA, we can specify how many of the PCs we would like to keep. For example, we can select the first 2 PCs. Then, we reconstruct the data using only these PCs by multiplying the matrix containing the first 2 PCs by the standardized input data. We have now reconstructed the data based on only these 2 PCs so the size of the dataset is much smaller however we have maximized the variance still contained in the transformed data.

I used Landsat-8 data from the Operational Land Imager Instrument. Since most glacier flow occurs in summertime, Landsat-8 is optimal for data because of its frequency of data collection. Since the Swiss Alps are primarily made of limestone, dolomite will be the focus of this paper **[3]**. The bands of interest are Band 5 (0.85 – 0.88 um) and Band 7 (2.11 – 2.29um) due to its distinct reflectance in Band 5 and absorption in Band 7.

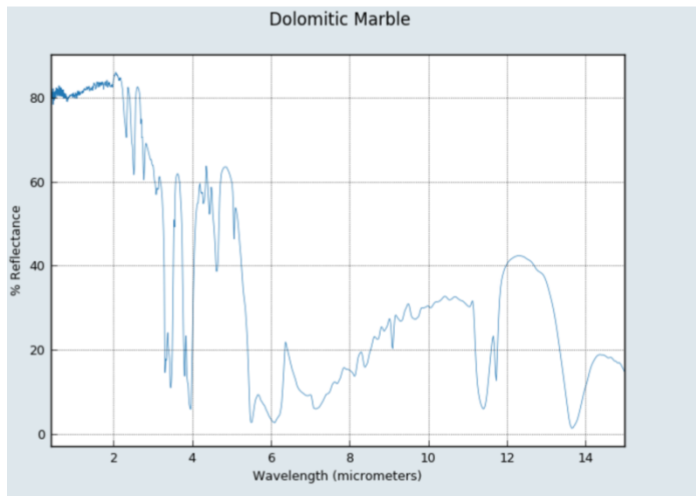


Figure 5 Dolomite Spectrum [6]

First to provide a control, I used band math calculate the band ratio between Band 5 and Band 7 without any data reduction. This provides a good comparison for the different number of principal components used for each application of PCA.

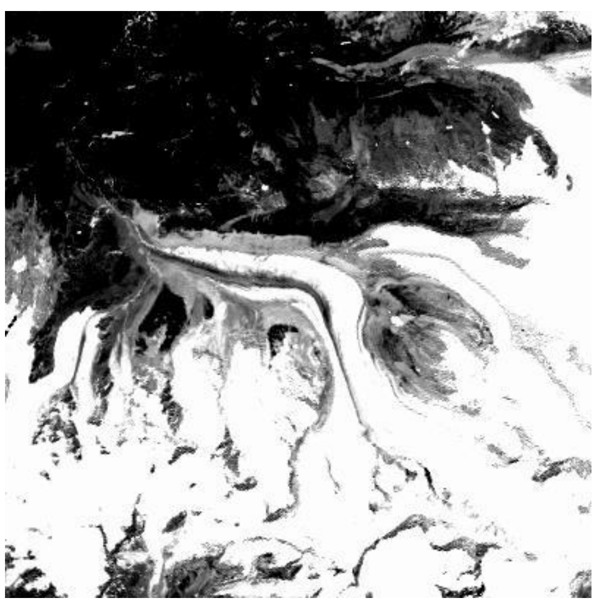
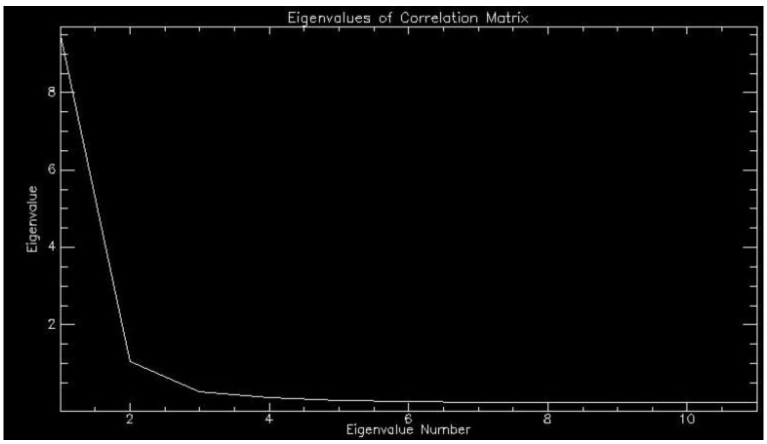


Figure 6 Band 5/7 Ratio No PCA

To preform PCA, I used ENVI's PCA Forward and Inverse transform functions. I used the layer stacking function to combine TIFF files for Bands 1-11 which cover a wavelength range from 0.43 -12.51 um **[4]**. Since USGS offers different levels of preprocessed images, I made sure to select scenes with atmospheric correction and without cloud cover. The scene I used was taken on August 30th 2015 to maximize the amount of moraine that is visible. **[5]**

Then, using the Forward PCA function I generated the statistics of the layered image and calculated the eigenvalues and eigenvectors of the correlation matrix. In EVNI, we can specify if we want to use the covariance matrix which subtracted the mean from each band or use the correlation matrix which subtracts both the mean and divides by variance of each band. I used the correlation matrix because the values in each band differ greatly so by normalizing the data we avoid the algorithm getting thrown off by bands with overall higher values.



When working with PCA, we can select how many principal components (PCs) we want to use. The PCs are listed in descending order of importance in Figure 7. From the plot of the eigenvalues, we see there is a steep drop off around the third eigenvalue.

Figure 7 Eigenvalues of Correlation Matrix

Using the Inverse PCA Rotation function, I reconstructed the images using only 2 and 3 principal components. The band ratios between Band 5 and Band 7 were once again computed for 2 and 3 PCs. The results of the inverse function are shown below.

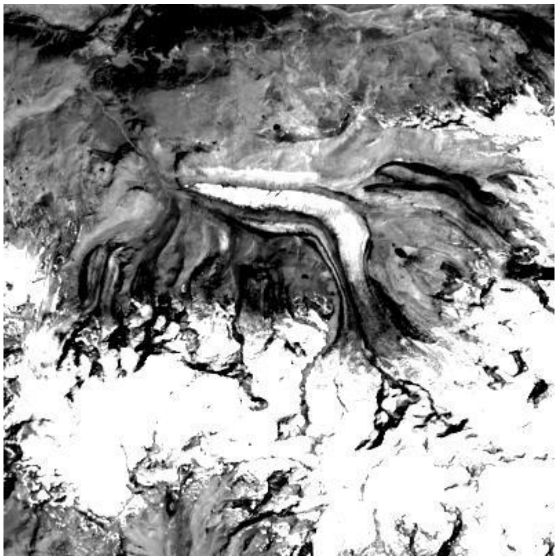


Figure 8 Band5/7 Ratio 2 PCs

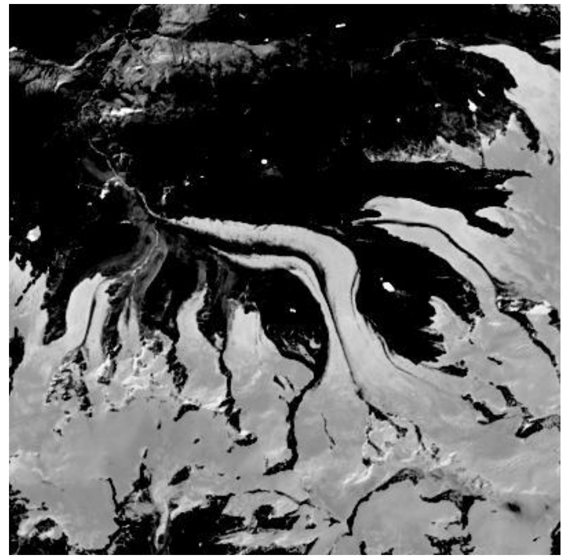


Figure 9 Band 5/7 Ratio 3 PCs

From the results we clearly see the image constructed with the first 3 PCs provides the highest clarity of the scene. The image 2 does provide a bit better clarity than without any PCA. These results are consistent with what was expected. The smaller PCs contain mainly noise which is removed since we only consider the first 2 or 3. Using PCA serves as dimensionality reduction which is useful for several reasons like faster computational time for further image processing like classification. Overall, these results indicate applicability of PCA as a data reduction technique for land remote sensing of glaciers.

Relevance to Course

In completing this project, I used many concepts we learned about in this course. Using Landsat-8 data required an understanding of what wavelengths are relevant to Earth remote sensing. Specifically, the H2O absorption bands from 3-5 um must be considered and this explains why these bands are excluded from Landsat-8 data. Furthermore, using PCA as a data reduction technique is critical to image processing workflows which were also discussed in this course. I also needed to understand how the data files from Landsat-8 are structured which meant understanding the spatial and spectra dimensions contained in the image. We used these concepts on multiple homework assignments in this class.

This project also required understanding of band math and taking band ratios. I had to read the dolomitic marble spectra and identify the key absorption and reflection regions. This also related to using ENVI since we learned how to do band math with this software.

Timeline

In this proposal I have provided a proof of concept for one dimensionality reduced image of Gorner Glacier and demonstrated the feasibility of determining composition from this image. Within a year, I would be able to analyze multiple images from the area in addition to other glacier systems, use images from different times of year, and use other dimensionality reduction techniques like non-linear PCA, autoencoders, and T-SNE to further study these images. The software I used in this paper, ENVI, does not contain functions for these dimensionality reduction methods, however there are many freely available packages in Python and MATLAB with these capabilities. Incorporating all these techniques would take about 1 month, and preforming the band ratios in ENVI would take an additional month because of the large number of images.

Building a convolutional neural network is relatively simple with packages like TensorFlow in Python, however training would take at least 3 months including optimizing the model. After the model is trained and optimized, testing it on New Horizons data will take approximately another 3 months.

By combining the results of these different dimensionality reduction techniques, the neural network will be able to find key spectral features of glacial moraines as observed by remote sensing instruments. The use of dimension reduction indicates that the key spectral features are strong enough to still be detecting in low resolution images. Thus, the findings can then aid in remote sensing of Pluto's surface and will determine the feasibility of detecting minerals with low resolution images.

References

[1] Howard, A. D., Moore,. et al,. (2017). Present and past glaciation on Pluto. *Icarus*, *287*, 287–300. https://doi.org/10.1016/j.icarus.2016.07.006

[2] Keeter, A. B. (2016, January 8). *Probing the mysterious glacial flow on Pluto's frozen 'Heart' – Pluto New Horizons*. https://blogs.nasa.gov/pluto/2016/01/08/probing-the-mysterious-glacial-flow-on-plutos-frozen-heart/

[3] Elkadi, J., Lehmann,. et al,. (2022). Quantification of post-glacier bedrock surface erosion in the European Alps using 10Be and optically stimulated luminescence exposure dating. *Earth Surface Dynamics*, *10*(5), 909–928. https://doi.org/10.5194/esurf-10-909-2022

[4] Landsat NASA. (2021, December 3). *Landsat 8 Bands | Landsat Science*. Landsat Science | a Joint NASA/USGS Earth Observation Program. https://landsat.gsfc.nasa.gov/satellites/landsat-8/landsat-8-bands/

[5] Survey, U.-. U. G. (n.d.). *Full display of LC08_L1TP_195028_20150830_20200908_02_T1 - EarthExplorer*. None. https://earthexplorer.usgs.gov/scene/metadata/full/5e81f14f59432a27/LC81950282015242LGN01/

[6] *Search Library — ECOSTRESS Speclib*. (n.d.). https://speclib.jpl.nasa.gov/library/ecoviewplot

[7] Linsbauer, A., Paul, F., & Haeberli, W. (2012). Modeling glacier thickness distribution and bed topography over entire mountain ranges with GlabTop: Application of a fast and robust approach. *Journal of Geophysical Research*, *117*(F3). https://doi.org/10.1029/2011jf002313