

TOWARDS BUILDING AN AUTOMATED ROCK CLASSIFIER FOR PLANETARY ROVER MISSIONS.

L. A. Valenzuela¹, P. M. Freeman², and V. C. Gulick³, S.T. Ishikawa³, S. M. Bass³, ¹Cal Poly Pomona/SETI Institute ²San Jose State University Research Foundation/NASA Ames ³NASA Ames/SETI Institute, MS 239-20, Moffett Field, CA 94035. virginia.c.gulick@nasa.gov.

Introduction: Building upon our previous work on an automated mineral classifier [1], we are working to extend its functionality towards classification of rocks.

The current work focuses on building an automated classifier using Raman spectral data to identify key minerals contained in igneous rocks. The ability to identify minerals in igneous rock samples provides important information on their composition.

Raman spectroscopy provides unique chemical signatures for minerals and other geological samples. It is a non-destructive technique that does not require sample preparation. Thus it makes an excellent choice to use in conjunction with image data since particular areas of an image (e.g., specific crystals, grains, layers or veins), may correspond to specific Raman signals.

Using Artificial Neural Networks to develop a Rock Classifier

As in our previous study [1], we are using artificial neural networks (ANNs) because of their ability to learn by example. More specifically, we used a Multi-layer Perceptron (MLP), which is an ANN that performs its example based learning using backpropagation, a gradient descent algorithm, which seeks the lowest point of error.

Methods: Using the Raman spectra from rock samples and ANNs, we built a classifier to determine if key minerals were present in rocks. As with the mineral classifier in our previous work [1], this classifier had four stages: The input stage, the preprocessing stage, the classification algorithm stage, and the output stage.

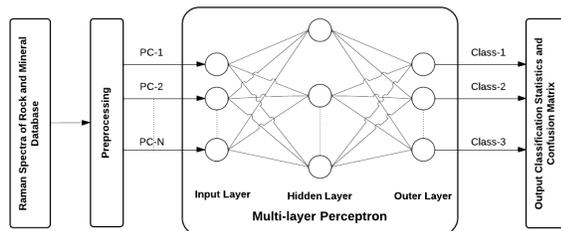


Figure 1. An illustration of the stages of the rock classifier: input, preprocessing, the MLP, and the output (Figure modified from [1]).

Organizing the Input Raman Spectral Data of Rock and Mineral Samples

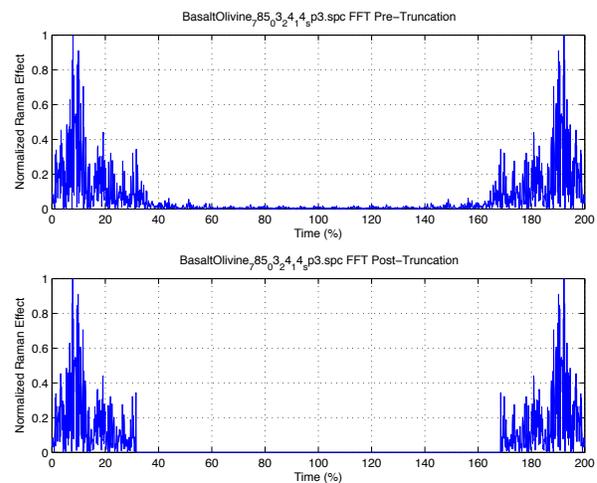
In machine learning, large data sets are needed to build prediction and classification models. To organize the spectral data, we had to define their instances and attributes. The instances are the unique individual

spectra that corresponded to particular samples- each time we recorded spectral data, we obtain a new instance. Attributes are the intensity values that correspond to the 1650 pixels in the spectrometer that sense the scattered photons due to Raman scattering. However, for the classifier to learn from the spectral data, the instances had to be labeled. For each spectrum, we identified by hand the key minerals present. We then designed our classifier to test for the presence of each of these minerals.

Preprocessing the Input Data

Spectral data contain noise due to limitations of the spectrometer and background fluorescence of minerals and rocks. To increase the signal to noise ratio, we used several preprocessing algorithms, including baseline reduction to remove fluorescence and normalization and smoothing to reduce noise in the spectra [1]. In this study, we used a Savitzky-Golay smoothing filter [1] and an FFT truncation filtering method, which reduces noise by transforming the Raman spectrum to the time domain and truncating information that doesn't pertain to the transient Raman effect [2].

Figure 2. Illustrates the time-domain representation of a



Raman spectrum. The important Raman information is within the initial transient. Symmetry is a property of the Fast Fourier Transform.

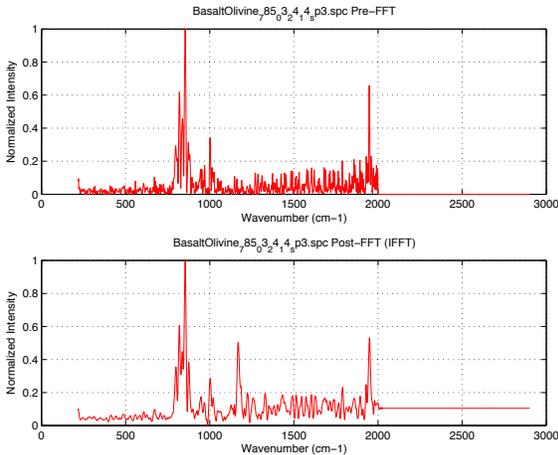


Figure 3. Illustration of spectrum before and after FFT truncation of noise. Notice the reduction of noise in the spectrum.

Furthermore, we used principal components analysis (PCA) in the preprocessing stage to reduce dimensionality of data [1]. The spectral data initially consisted of 1650 dimensions because each attribute (dimension) pertains to a single pixel in the CCD of the spectrometer. PCA essentially captures a desired variance (typically 95%) and clusters data into a new lower-dimensional axis.

Design the Multi-layer Perceptron for Learning

In our previous study [1], the multi-layer perceptron architecture has three layers: input, hidden, and output. The input layer consists of neurons which correspond to the number of principal components (typically less than ten). The hidden layer is quantified by the number of discrete mineral classes being identified in the rock spectra. The output layer- like the hidden layer- is quantified by discrete classes.

Results:

For these tests, we used 156 spectra from 19 igneous rocks: 6 granite, 4 diorite, 4 gabbro, 2 granodiorite, 1 rhyolite, 1 quartz monzonite, and 1 quartz monzodiorite. PCA results using both the Savitzky-Golay and the FFT truncation approach for classifying the individual minerals in igneous rock samples are presented in Figure 4. We used a 10-fold cross-validation scheme to test our classifier and varied the number of principle components. The percentage of minerals correctly classified improves with the number of components and then plateaus after about eight components.

The classifier identified the presence of several important mineral groups in igneous rocks. Both the Savitzky-Golay and FFT truncation techniques performed comparably. However, the results for K-feldspar and plagioclase for mineral ID in rock spectra are only around 60%. Although if we combine both the K-spar and plagioclase into a single class, feldspar, then the results improve considerably (~80% range). However,

clearly, more work needs to be done to improve the classifier especially in the preprocessing stage.

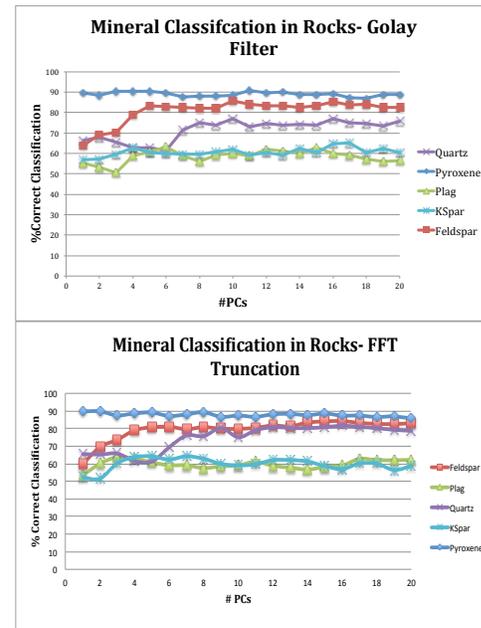


Figure 4. Cross-validation results for determining presence of minerals in rock spectra with known mineral composition as a plot of classification percentage vs. number of principal components.

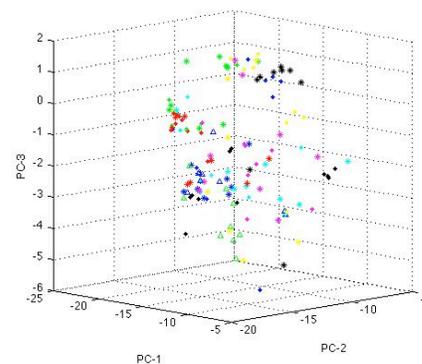


Figure 5. Illustrates the rock data processed after using PCA. Each color represents a unique rock sample. Clearly, the rock data does not cluster as well as the mineral data in [1].

Future Work: As Figure 5 shows classifying key minerals is not sufficient to classify rocks. Therefore, we will be incorporating automated color and texture classifiers from our previous work [3] to provide additional insight into the automated rock classification.

References:

- [1] Ishikawa S. T. and Gulick V. C. (2013) Computers & Geosciences, 54, 259-268, [2] Sobron P. et al. (2008) 62, 364-370. [3] Freeman et al 2014. LPSC, abs. 2739.