A COMPARISON OF TWO METHODS FOR AUTOMATED MINERAL DETECTION IN VISIBLE/NEAR-INFRARED SPECTRA. R. Castaño[1], M. Gilmore[2], B. Bornstein[1], S. Hojnacki[1], and J. Greenwood[2], [1]Jet Propulsion Laboratory, Pasadena, CA 91109, Rebecca.Castano@jpl.nasa.gov, [2]Dept. of Earth & Environmental Sciences, Wesleyan University, Middletown, CT 06459.

Introduction: The analysis of the huge volumes of multidimensional data collected by hyperspectral systems to identify materials within the spectra can be a costly and time consuming task. On Earth, it may necessitate field checking of targets in a given scene and knowledge of scene-specific parameters *e.g.*, [1, 2 and references therein, 3]. For planetary missions, where global datasets are collected, these data analysis costs restrict rapid classification to only a subset of an entire mission dataset and in-depth classification to a very limited set of minerals.

We are developing a series of algorithms to automate the processing and analysis of hyperspectral images to identify constituent spectra. We have focused on the identification of specific important mineral compositions within pixels in the data. The primary goal of this effort is to design automated supervised classifiers that can rapidly analyze hyperspectral data to determine key regions of focus for science teams and individual investigators. Specific targeting is a frequent application of hyperspectral imaging, *e.g*., minerals associated with acid mine drainage, hydrocarbons, or mineral exploration. For Mars, high priority targets include minerals associated with the presence of water (*e.g.*, carbonates, sulfates, evaporates).

We demonstrate our methods on an Earth scene collected by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) and a Mars scene collected by the Observatoire pour la Minéralogie, l'Eau, les Glaces et l'Activité (OMEGA). Here we focus on calcite ($CaCO_3$) in the AVIRIS scene and gypsum ($CaSO_4 \cdot 2H_2O$) in the OMEGA scene.

The AVIRIS scene is a well-studied image of Cuprite, Nevada, which has served as a training site for other classification algorithms (*e.g.*, [1,2]). The AVIRIS instrument has a 17 m/pixel footprint with 224 spectral bands over 400-2500 nm [4].

The OMEGA visible and near-infrared imaging spectrometer has been operating since January 2004 and maps 352 wavelengths from 0.35 to 5.1 mm [5,6]. The OMEGA scene that we used is an atmospherically-corrected [7] image of Iani Chaos.

Method: We use supervised classifiers to develop a mapping function from input spectral data to mineral classes. Learning the mapping function requires a set of training examples with known minerals for training. Acquiring an adequate number of labeled training examples for robust supervised classification can be costly. Manual labeling is labor intensive and can result in noisy labels when the true underlying class is not unambiguous from visual inspection of spectra. To address this issue, we have developed a stochastic function that can be used to generate large volumes of training data. We compare classification performance for classifiers trained on synthetically generated data based on mineral libraries to the performance of classifiers trained on hand-labeled data.

Generative Model. We inexpensively obtain spectral data that represents the characteristics of field data, through a generative model for rock spectra. Using this model, spectra for *virtual rocks* are generated and used to train a classifier. Each virtual rock is formed from a set of minerals described in terms of constituent percentage ranges. Constituent mineral spectra are sampled from a database populated by two well-known spectral libraries: JPL's Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) library [8] and the U.S. Geological Survey's (USGS) speclib04 [9], and combined to form the rock spectra using a linear mixing model. We then add noise and resample from the wavelengths in the database to those of the actual data collection instrument.

Manual Labeling. For the AVIRIS scene, 776 pixels were labeled as containing calcite (positive). The remaining unlabeled pixels were assumed to not have calcite (negative). While 776 positive pixels were identified, the labeling was not exhaustive, i.e. there may be pixels in the data that have calcite present but are not labeled as positive examples.

In the OMEGA scene, 307 pixels including those identified as containing the mineral gypsum by [10] were labeled by a domain expert. All unlabeled pixels (53965) were assumed to not contain gypsum. However, as with the AVIRIS data, the labeling may not be complete.

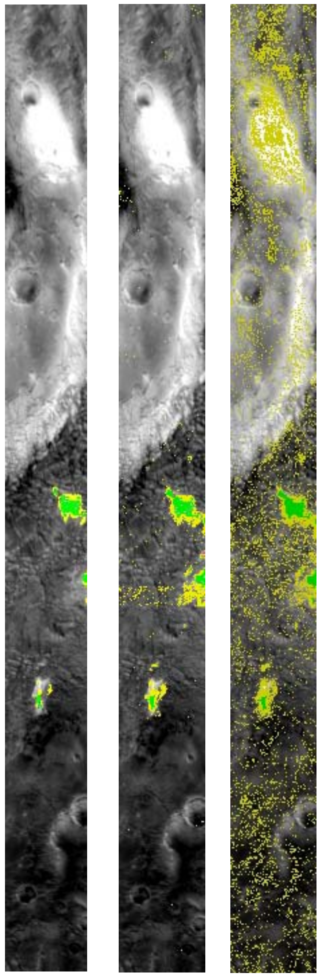
Classifiers. Support Vector Machine (SVM) classifiers were trained for the detection of calcite and gypsum. Labeled pixel (LP) and generated model (GM) detectors were developed for calcite and tested on the AVIRIS scene. The calcite detector used the 2000-2400nm wavelength range. Our second experiment compared the LP and GM detectors for gypsum on the OMEGA data. The gypsum detector input was limited to the 1200-2600 nm wavelength range. These ranges encompass the unique spectral features of each mineral, reducing computation time.

Results and Discussion: Preliminary results are shown for the comparison of the gypsum detectors in Figure 1. Both the LP and GM classifiers were good at correctly identifying the data labeled as gypsum. The LP classifier had considerably higher precision than the GM classifier. False positives that are clustered in regions near the true positives (see Figure 1) are indicative of the detector being more sensitive to the presence of gypsum than the label set. However, the false positives scattered across the scene as in the generative model results are more likely to be true false positives. Results for the detectors on the AVIRIS data are similar where the GM classifier had greater precision (98%) than the LP (95%).

For both data sets, the labeled classifier is more conservative than the generative model classifier, due to the labeled pixels inherently being a better representation of the target data containing scene-specific noise and the nature of the labeling process in which only very high confidence points were labeled. However, since the labeled pixels are modeling the specific scene noise, the we believe the generative model approach will be able to more effectively classify a wider variety of hyperspectral scenes. This method is also more cost-effective for training the classifiers, as it is tedious and can be error-prone to manually label large volumes of image data. In addition, when searching for rare minerals, there may be no known cases or insufficient cases for training a robust classifier. Both classifiers would be useful for the purpose of focusing attention.

Acknowledgements: We thank Jack Mustard and Shannon Pelkey for providing the atmospherically-corrected OMEGA image cubes as well as valuable conversations. Discussions with Scott Murchie are also appreciated. This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology under a NASA contract. Additional funding was provided by the Interplanetary Network Technology Program.

References: [1] Swayze, G. A. et al. (1992) *JPL Airborne Geosciences Workshop*, 92-14, 127-143. [2] Clark, R. N. et al. (2003) *JGR*, 108 (E12), 5131. [3] Dalton, J. B. et al. (2004) *Remote Sens. Environ*. 89, 455-466. [4] Vane, G. et al. (1993) *Remote Sens. Environ*, 44:2-3, 127-143. [5] Bibring, J.-P. et al. (2004) In: *Mars Express: The Scientific Payload,* ESA SP-1240, 37-49. [6] Bibring, J.-P. et al. (2005) *Science*, 307:5715, 1576–1581. [7] Bibring, J.-P., et al. (1989) Nature 341, 591. [8] Hooke, S. (2000) *The JPL ASTER Spectral Library v1.1*. [9] Clark, R. N. et al. (1993) *U.S. Geological Survey Open File Report*, 93-592. [10] Gendrin, A. (2005) *Science*, 307:5715, 1587-1591.



(a) (b) (c)

Detector	TP/Truth	Percent
A	**249/277**	**90%**
B	**266/277**	**96%**
C	**303/307**	**99%**

Figure 1: Gypsum detector results. (a) Labeled-pixel trained, reflectance data. (b) Labeled-pixel trained, continuum removed data. (c) Generative-model-trained. Green pixels are those that were labeled and detected as having gypsum (true positives). Yellow pixels were classified as having gypsum but were not labeled (false positives). Maroon pixels were labeled as gypsum but not detected.