AN AUTOMATED APPROACH FOR ACQUIRING ONBOARD ROVER SCIENCE, R. C. Anderson,[1] R. Castano,[1] D. Decoste, [1] D. Mazzoni[1], and [2]J. Dohm, [1]Jet Propulsion Laboratory, Pasadena, CA 91109, [2]University of Arizona, Robert.C.Anderson@jpl.nasa.gov.

Abstract: Rover traverse distances are increasing at a faster rate than downlink capacity is increasing. As this trend continues, the quantity of data that can be returned to Earth per meter traversed is reduced. The capacity of the rover to collect data, however, remains high. This circumstance leads to an opportunity to increase mission science return by carefully selecting the data with the highest science interest for downlink. We have developed an onboard science analysis technology for increasing science return from missions. Our technology evaluates the geologic data gathered by the rover. This analysis is used to prioritize the data for transmission, so that the data with the highest science value is transmitted to Earth. In addition, the onboard analysis results are used to identify additional science gathering opportunities. A planning and scheduling component of the system enables the rover to take advantage of the identified science opportunity.

Introduction: With the continued expansion of planetary exploration, future missions will cause an increase in the volume of data that can be acquired for transmission to Earth via the Deep Space Network (DSN). Missions will have to make critical decisions regarding the quantity and quality of the downlink data. Although the DSN's transmission capability improves every year, the DSN cannot keep pace with the vast number of missions it must service. New methods must be developed to maximize the science return for the available bandwidth. We are developing an approach to counteract this transmission limitation by improving the quality of useful information conveyed per bit. We have developed onboard analysis methods to autonomously prioritize data for downlink, thus maximizing the science return for the available bandwidth. By carefully selecting key data for transmission, the quality of data returned is increased. Selecting the most interesting data requires encapsulation of characteristics of data with high science value in a form that can be analyzed and evaluated onboard. The spacecraft must have the capability of automatically recognizing data that contains information about targets with high science interest. Since images are among the highest bandwidth data that a rover collects, images represent one of the greatest opportunities for data prioritization.

Methodology: To assess and subsequently prioritize the scientific value of a set of collected images, we must first extract the information found within the images. A geologist in the field would extract the information from a site by identifying geologic features including the albedo, texture, shape, size, color, and arrangement of rocks, and features of the topography such as layers in a cliff face. Our system begins by first evaluating each image and locating the rocks within it. Next, we extract the properties of the rocks, including albedo, texture and shape features. These features are assigned a science value, and the importance of the image is evaluated based on these values as compared to the features extracted from rocks identified in other images.

Rock Classification: Images with interesting features, such as rocks with unusual shapes or textures, should be ranked higher than images without distinctive features. We have developed three different prioritization methods that use the extracted rock features to rank the rocks in terms of scientific importance. Once the rocks are prioritized, the images containing the rocks are ranked based on the rocks contained within them.

Extraction: Our technique for locating rocks is based on finding objects above the ground plane. We begin by determining the ground plane from the stereo range data, which is already calculated for navigation purposes. We then produce a height image, in which the value of each pixel represents the elevation of the point above the ground plane. Level contours in the height image are calculated and then these contours are connected from peaks to the ground plane to identify the rocks [1].

Rock properties including albedo, visual texture and shape, are then extracted from the rocks identified. We measure **albedo**, an indicator of the reflectance properties of a surface, by computing the average gray-scale value of the pixels that comprise the image of the rock. The reflectance properties of a rock provide information about its mineralogical composition. The second rock property extracted is **visual texture**. Visual texture can provide valuable clues to both the mineral composition and geological history of a rock. Visual texture can be described by gray-scale intensity variations at different orientations and spatial frequencies within the image. We measure texture using a bank of Gabor filters [2, 3]. Gabor filters are scale and orientation specific, thus the results of convolving an image with these filters can be successfully used to discriminate between different textures. Another important and geologically useful feature of rocks is their inherent **shape**. For example, a rock that is highly rounded may have undergone fluvial processing and traveled far from its source. Conversely, a rock that is highly angular is likely to be close to its source and to have undergone minimal secondary processing. We begin by fitting an ellipse to the boundary points of the identified rock in the image [4]. Our first shape measure is the eccentricity of this ellipse. Our second measure is the error between the boundary points and the ellipse. The third and final measure is angularity, which is measured as the standard deviation of the angle of the edge at each boundary point.

Prioritization: The features extracted from a group of images are then used to rank the images using the three distinct prioritization algorithms. The first technique recognizes pre-specified *target signatures* that have been identified by the science team as data of high interest. The second technique, *novelty detection*, identifies unusual signatures that do not conform to the statistical norm for the region. The last method, known as *representative sampling*, prioritizes data for downlink by ensuring that representative rocks of the traversed region are returned.

Target Signature: We have implemented a method for enabling scientists to efficiently and easily stipulate the value and importance to assign to each feature. Rocks are then prioritized as a function of the distance of their extracted feature vector from the specified weighted feature vector. Scientists are given two ways to set the target signatures that will determine how the rocks are ranked. In the first method, the scientist may, for example, chose to prioritize rocks based on two aspects of their shape, such as eccentricity and ellipse fit. The second manner in which scientists can specify a target signature is by selecting a rock with interesting properties from the set of already identified rocks. Rocks that resemble this particular rock in the selected properties are given a high priority.

Novelty Detection: We have developed three methods for detecting and prioritizing novel rocks, representing the three dominant flavors of machine learning approaches to novelty detection: distance-based, probability-based (i.e. "generative"), and discriminative. They will have general utility for other novelty detection tasks as well, but they are specifically designed with onboard constraints and large candidate feature spaces in mind. The first novelty detection method is a distance-based k-means clustering approach. Initially, all available rock data is clustered into a specified number (k) of classes. The novelty of any rock is then the distance of the rock feature vector to the nearest center of any of the k clusters. The greater the rock's distance is to the nearest center, the higher the novelty ranking assigned to the rock. The second technique is a probability-based Gaussian mixture model, which attempts to model the probability density over the feature space. In this approach, the novelty of a rock is inversely proportional to the resulting probability of that rock being generated by the model learned on previous rock data. The final method is a discrimination-based kernel one-class classifier approach. Here we treat all previous rock data as the "positive class" and learn the discriminant boundary that encloses all that data in the feature space. We essentially consider the previous rock data as a cloud scatter in some D-dimensional space, where D is the number of features. The algorithm learns the boundary of that cloud, so that future rock data that falls farther outside the cloud boundary is considered more novel.

Representative Target: One of the objectives for rover traverse science is to gain an understanding of the region being traversed. To meet this objective, the downlink back to Earth should include information on rocks that are typical for a region, and not just information on interesting and unusual rocks. A region is likely populated by several types of rocks with each rock type having a different abundance. If uniform sampling is employed for downlink image selection, as opposed to our autonomous onboard selection process, the downlinked set will be biased towards the dominant class of rock present. This situation may result in smaller classes not being represented at all in the downlinked data.

To provide an understanding of the typical characteristics of a region, rocks are first clustered into groups with similar properties. The data is then prioritized to ensure that representative rocks from each class are sampled. The rocks are clustered into groups based on the features extracted from the image data for each rock. To determine the classes, the property values are concatenated together to form a feature vector, and a weight is assigned to the importance of each property. Different weight assignments can be used as a function of the particular properties that are of interest. For example, albedo and texture are typically used to distinguish types of rocks, but rock size may be used if sorting is of interest. Unsupervised clustering is then used to separate the feature vectors into similar classes. We currently employ k-means due to its relatively low computational requirements, although any unsupervised method could be used. For each class of rocks, we find the most representative rock in the class, i.e., the single rock in any image that is closest to the mean of the set. We give a high priority to the image containing this rock. The optimal number of classes can be determined using cross-validation techniques [5]. Since some classes of rock may be more common than others, this prioritization method ensures that all classes of rocks in a set of images are represented. Without prioritizing the rocks by grouping them into classes, a disproportionate number of rocks from the most common classes are more likely to be represented in the data and classes that have fewer members may not be represented at all in the downlinked data.

In the future we will use the spatial location of the rocks in addition to their property values to enable expanded analyses, including characterizing local surface regions and sorting, which requires size and location information. Finally, prioritization can be used for more than just data downlink decisions. It can also be used for opportunistic science. Targets of high science value can be identified for additional instrument measurements. Prioritization that calls for opportunistic science is a wasted capability without a method of re-sequencing the rover, or orbital spacecraft, to obtain the additional scientific observations requested. This ability for real-time opportunistic science requires integrating the prioritization module with the onboard planning and scheduling system. We are in the process of including this integration.

References: [1] Gor V. et al, *AIAA* (2001), [2] Castaño, R. L., *Proc. Applications of Digital Image Processing* XXII, SPIE Vol. 3808, July, (1999). [3] Gilmore, M., et al., *JGR*., Vol. 105, No. E12, Dec. 2000 pp., [4] Fox, J. *IEEE Aerospace Conference*, Big Sky, Montana, Mar. (2002), [5] Smyth, P., *Proc. 2nd Int. Conf. on Knowledge Discovery and Data Mining,* AAAI Press, (1996), 29223-29237.