Multi-scale Features for Detection and Segmentation of Rocks in Mars Images

Heather Dunlop, David R. Thompson and David Wettergreen
The Robotics Institute, Carnegie Mellon University
5000 Forbes Ave. Pittsburgh, PA 15213, USA
heather@dunlop1.net, drt@ri.cmu.edu, dsw@ri.cmu.edu

Abstract

Geologists and planetary scientists will benefit from methods for accurate segmentation of rocks in natural scenes. However, rocks are poorly suited for current visual segmentation techniques — they exhibit diverse morphologies and have no uniform property to distinguish them from background soil. We address this challenge with a novel detection and segmentation method incorporating features from multiple scales. These features include local attributes such as texture, object attributes such as shading and two-dimensional shape, and scene attributes such as the direction of illumination. Our method uses a superpixel segmentation followed by region-merging to search for the most probable groups of superpixels. A learned model of rock appearances identifies whole rocks by scoring candidate superpixel groupings. We evaluate our method's performance on representative images from the Mars Exploration Rover catalog.

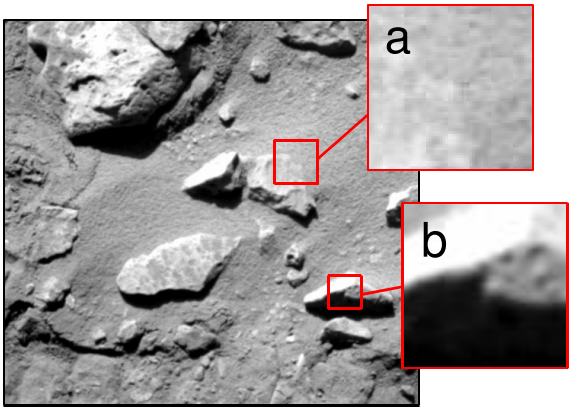

Figure 1. Rocks' diverse textures, variable morphologies, and directional illumination all complicate the segmentation problem. Local cues are insufficient: real borders are often weak (a), while the rocks themselves are non-uniform with false interior borders (b). Image from the Mars Exploration Rover "Spirit," courtesy of NASA/JPL.

1. Introduction

Planetary rovers like those on Mars today can collect more data than can be analyzed in mission-relevant time frames; they have usually moved on before their images of a site are thoroughly analyzed. The Mars Exploration Rovers Spirit and Opportunity have already accumulated a catalog of hundreds of thousands of images. Future rovers will have even greater mobility and lifespan; missions could cover kilometers every day [1] and last for years. These missions traverse vast areas and collect diverse images of the planetary surface. However, improved exploration speed and efficiency will reduce the time window for any data analysis that is pertinent to the ongoing exploration.

The identification of observed rocks is an important task in route planning and geologic analysis. Rock shape, weathering, and dispersion carry important information about environmental characteristics and processes. Currently these attributes are characterized by a person who manually labels the rocks in an image. Comprehensive analyses of this sort are usually too time-consuming to be practical. When the analysis is performed at all, it is too late to inform mission planning decisions. An automated method for identifying rocks in images would allow automatic extraction of geologic information [2] by enabling the real-time calculation of rock size, shape and class distributions [3]. More rapid and accurate rock analysis would support immediate tactical and scientific decision-making.

Previous research has produced several automated algorithms for finding rocks, including template-based approaches [4], stereo geometry [5, 2], finding closed contours with an edge detector [6], and classifying homogeneous regions with a belief network [7]. These algorithms can detect rocks but seldom find their actual boundary contours [4]. This is unfortunate because an accurate segmentation is usually required for estimating rocks' geologic attributes.

Segmenting rocks is difficult because they are non-uniform: their texture, color and albedo varies across their surfaces and from one rock to the next. Often Mars im-

ages exhibit strong directional lighting with cast shadows and highlights that violate the uniformity assumption. Weak boundary edges must be inferred from context (Figure 1).

Recent work in simultaneous segmentation and recognition has tried to reduce reliance on local features [8, 9]. These methods exploit 2D geometric structure to assemble heterogeneous parts into likely objects. Unfortunately, rocks exhibit infinite morphological variation with little common structure. Their appearance can vary greatly with respect to both shape and lighting conditions. Thus, rocks constitute a significant pattern recognition challenge: neither local properties nor global structures alone are sufficient.

This work proposes a new segmentation method that exploits multi-scale cues simultaneously. It aims for a segmentation that is consistent with local features such as texture, object features such as shape and shading, and scene features such as the direction of illumination.

We next describe the algorithm and then demonstrate its performance on a sample dataset from the Mars Exploration Rover catalog.

2. Approach

Our rock identification method fragments the source image using a normalized-cuts strategy [8, 10] and merges the resulting fragments, or superpixels [10], into candidate rock regions (Figure 2). A search process identifies high-probability superpixel groupings where labels are most consistent with the local and global features of our rock model. The search evaluates each possible segmentation of candidate regions using a learned model trained to recognize rocks and soil under directional lighting.

2.1. Creating Superpixels

An initial over-segmentation fragments the image into a set of *superpixels* — areas of homogeneous properties that become the atomic elements of the search [10]. The use of superpixels instead of image pixels reduces the space of possible segmentations while still ensuring that some valid segmentation captures the contours of all rocks. Ideally, we desire to identify the borders of all rocks using as few superpixels as possible.

We evaluated several approaches for producing superpixels, including mean shift [11], Blobworld [12] and graph-based [13] methods. The method of Mori used in [8] and [14] achieved the best qualitative results, accurately capturing the boundaries of rocks with the fewest superpixels. This method uses normalized cuts [15] with a boundary detector [16] to provide a contour grouping cue. In order to capture many sizes of rocks, we compute superpixel fragments at four different scales (Figure 2).

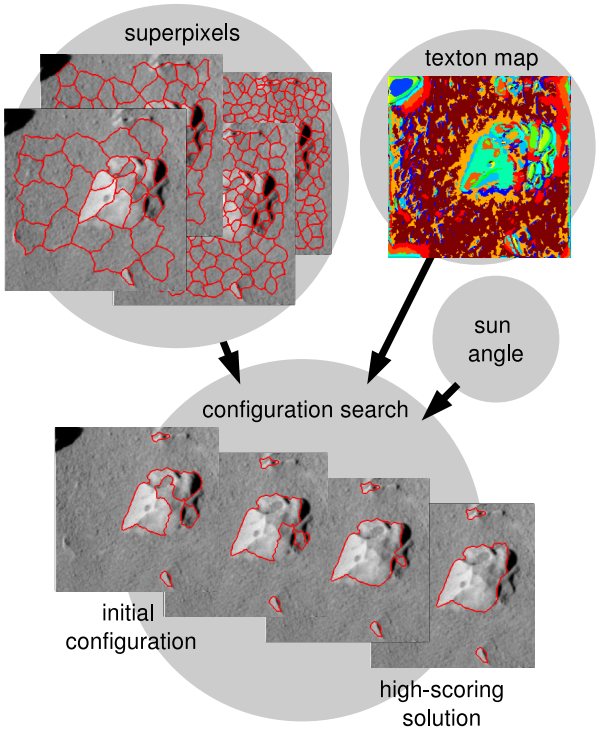

Figure 2. Rock segmentation by merging superpixels. Texture cues provide an initial guess of rocks' locations. Then a search process evaluates candidate segmentations using texture, shape and shading features.

2.2. Modeling Rocks

Our method assembles rocks from one or more superpixels that form a contiguous region in the image. In order to identify quality segmentations, we learn a mapping from an image region's features to the probability that the region is actually a rock. Many of the features we consider for this application are similar to those used in region merging by Mori [8]. A brief description of the feature set follows.

2.2.1 Shape

Rocks have very irregular shapes; no two exhibit exactly the same morphology. However, their shapes are recognizably different from a randomly generated region. They tend to be more ellipsoidal and convex than random groupings of superpixels. Dunlop provides a comprehensive overview of these shape measures and their utility in describing rocks [3].

We characterize each region with several shape attributes. First, we calculate the residual error between the rock region's contour and its corresponding best-fit ellipse [2]. We use two measures of convexity based on the convex hull of the region's pixels: the ratio of the convex hull

perimeter to the actual perimeter of the rock, and the ratio of the convex hull area to the rock's true area [17]. Finally, we consider the region's circularity, defined as the square of its perimeter divided by its area [18].

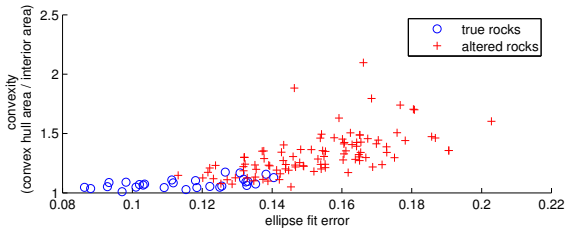

Figure 3. Two shape features distinguish rock regions and altered rock regions in a training image. The plot shows ellipse fit error and convex hull (area) features.

Figure 3 illustrates the ability of two shape features to discriminate between rock and non-rock regions in a training set. We generate non-rock regions by altering a true rock's border through random addition and subtraction of $1-3$ superpixels. To ensure that the change was significant, we restrict the training set to those regions whose area differs from their base rocks by at least 50%. The figure suggests that random perturbations tend to increase both the convexity and ellipse-fit error of a rock region.

2.2.2 Texture

Our texture measure uses the method of Varma and Zisserman [19]. We convolve all images in a training set with the Maximum Response 8 filter bank. This results in an 8-dimensional response vector for each pixel. We aggregate the responses for several training images and cluster them using k-means to form a set of 16 universal textons. We use these textons to compute a texton map for each novel image by assigning pixels to their Euclidean-nearest texton.

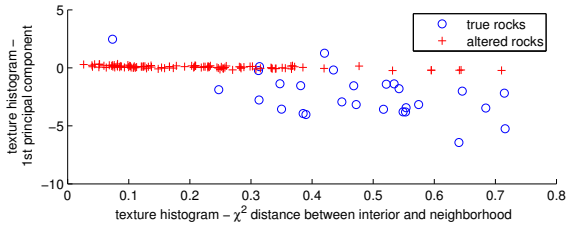

Figure 4. Separability of the true rocks and altered rocks using texture features.

Two texture measures appear in the final feature vector. The first is an area-normalized histogram giving the frequency of occurrence for each of the 16 textons. The second is a χ^2 distance measure between the texton histogram of the interior region and the texton histogram of its local neighborhood. We define the neighborhood of a region to be its bounding box plus a small margin, excluding the region itself. These attributes provide both a measure of the region's texture and the degree to which its texture differs from that of neighboring regions.

Texture attributes for the true rocks and altered rock regions appear in Figure 4. We use Principal Component Analysis, projecting the 16-dimensional texton histogram onto its first principal component for visualization. Rocks are characterized by large texture differences with their local neighborhood; altering their boundaries drives them quickly toward the typical texton distribution.

2.2.3 Shading

Our shading measure exploits several unique features of the Mars problem domain. First, an accurate estimate of sun direction is available for each image in the catalog — the illumination direction can be observed directly using an onboard sensor or calculated from rover pose and ephemeris data. Additionally, rocks on the Martian surface are generally Lambertian. This makes them amenable to the classic shape-from-shading reflectance formula [20]. However, shape-from-shading techniques generally perform poorly on non-synthetic data [20]. Mars imagery, in particular, contains plentiful cast shadows and a range of rock morphologies that make it difficult to constrain the space of possible surface reconstructions.

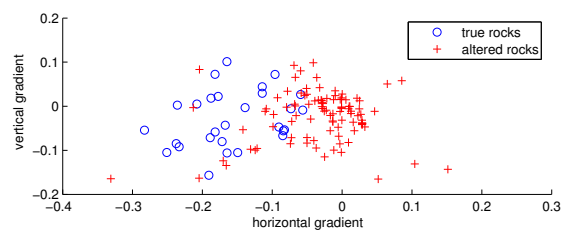

Figure 5. Intensity gradients of the rock and altered rock regions.

Fortunately, scoring a candidate region does not require a full 3D reconstruction. The decision boundary separating rock shading patterns from non-rock shading patterns is far simpler than a general model of 3D rock shapes. Rocks protrude above the soil surface; for a given illumination, their surfaces show a common shading gradient with highlights on one side and shadows on the other. We characterize this phenomenon with a four-dimensional shading measure: the azimuth and elevation of the sun in the camera frame, and the 2D intensity gradient on the rock's surface.

We find the intensity gradient using linear least-squares regression applied to rock pixel intensities. Figure 5 shows intensity gradients for a single image, where the true rocks exhibit a high-strength gradient in the direction of illumi-

nation. Conversely, altered rocks trend toward a zero mean intensity gradient.

2.3. Searching Segmentations

Each candidate segmentation defines a set of rock regions in the image. For our purposes, every pixel not belonging to some rock is considered to be soil. Thus, each segmentation determines a set of rock regions $X = \{x_1, x_2 \ldots, x_n\}$ and soil pixels $Y = \{y_1, y_2, \ldots, y_n\}$. A segmentation's *score* depends on how well these match the learned appearances of actual rocks and soil.

We generate class probabilities using two Support Vector Machine (SVM) classifiers with radial basis kernel functions [21]. The rock SVM uses the complete 26-dimensional feature vector describing both local and global properties of the region. It learns the mapping from this feature vector to the probability $P_r(x_i)$ that a region is a rock.

The soil SVM produces an estimate $P_s(y_i)$ that each pixel is soil. It generates these probabilities independently for each superpixel using local texture histograms alone. Averaging soil probabilities across different scales of superpixels yields a value $P_n(y_i)$ for each image pixel.

In addition to its role in scoring candidate segmentations, the soil classifier acts as a heuristic for initializing the search procedure. The space of possible rock regions is too large for an exhaustive search. Under these conditions, data-driven heuristics that favor more probable superpixel groupings can dramatically improve performance [9]. We group superpixels with a low probability of being soil into contiguous regions that become rocks for the first round of search (Figure 2). This initial guess based on texture dramatically improves performance.

The search procedure seeks higher-scoring segmentations by modifying rock regions in one of four ways: (1) *Growing* the rock by adding a superpixel at some scale; (2) *Shrinking* the rock by subtracting a superpixel at some scale; (3) *Splitting* the rock by removing a superpixel to make a new rock region; and (4) *Merging* two contiguous rock regions into a larger rock region. This search procedure generates regions that could not have been constructed using superpixels from a single scale exclusively. The search seeks a segmentation that maximizes the following objective function:

$$f(X, Y) = \sum_{x_i \in X} A(x_i) P_r(x_i) + \sum_{y_i \in Y} P_n(y_i) \qquad (1)$$

where $A(x)$ is the pixel area of a rock region. Rock and soil areas are zero-sum so the two models compete to explain the scene. During the search we recompute the objective function following each modification.

For the datasets in this work, a hill-climbing search provides a fast, reasonable solution. This search strategy evaluates all possible modifications to a rock, chooses the highest-scoring alternative at each iteration, and terminates after a small fixed number of iterations. At the end of the search procedure we erase rock regions whose probabilities do not exceed some fixed threshold; and any remaining rocks constitute the final segmentation. We employ this greedy method exclusively in the experiments that follow.

3. Experiments

We have conducted experimental analysis of our multi-scale segmentation method using images of the Martian planetary surface. Here we determine the detection accuracy, compare the value of individual features, and measure the error in boundary localization.

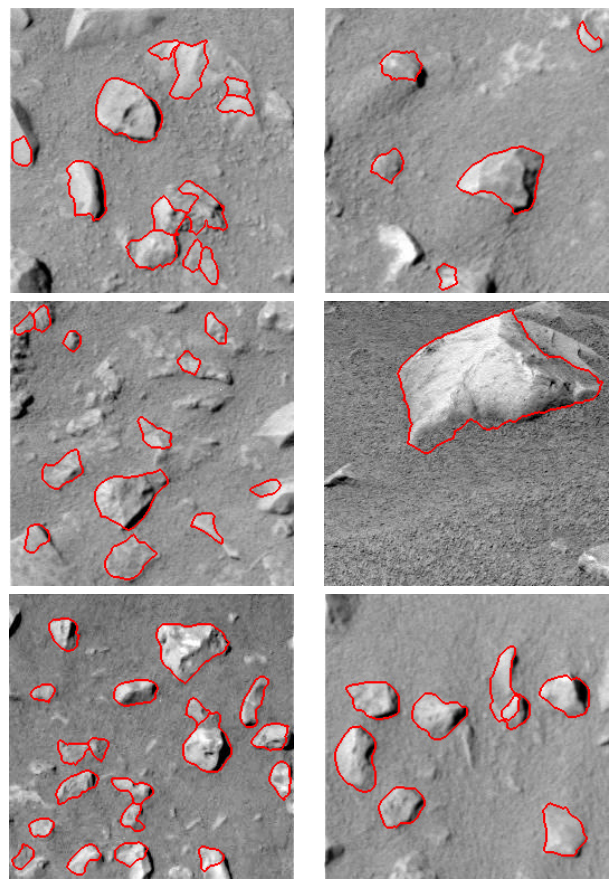

Figure 6. Detected rocks.

3.1. Procedure

The experiments detailed in this section involved images drawn from the Spirit Mars Rover Panoramic Camera image catalog at various locations along its traverse path in Gusev Crater. This study is restricted to near-field images of the terrain directly in front of the rover. We favored these

downward-looking images for two reasons. First, for uniform distributions of rocks, near-field scenes appeared less cluttered due to a lack of foreshortening. Second, near-field rocks contained a larger number of pixels for computing texture and shading statistics. Further studies of Mars images will consider other scene types.

Our training set consisted of 150 hand-labeled images, each 256×256 pixels, that together contained over 500 appropriately-sized Mars rocks under different illuminations. A human created the training set by manually outlining all rocks with an area greater than 100 pixels (the approximate area of the smallest superpixels in our segmentation). We produce negative examples by randomly altering the borders of the hand-labeled rocks as described in section 2.2.1. We analyzed performance using "leave one out" cross validation on the image set.

Creating multi-scale superpixel segmentations for each image was the most computationally-intensive portion of the algorithm, taking several days to compute for the whole data set. The search process for each image required several minutes of processing time on a single-core 3.2 GHz desktop computer. Figure 6 shows some examples of rocks detected in this data set. The following sections provide a quantitative performance evaluation.

3.2. Performance Results

Here we investigate a variety of system performance measures. First we consider the influence of individual features on region labeling accuracy. Then we perform a more formal evaluation on segmentation performance for a sample set of Mars images.

3.2.1 Feature Comparison

The Venn diagram of Figure 7 shows the influence of shape and shading features on rock-labeling accuracy. We trained the SVM for each set of training features and assessed its accuracy in predicting rock versus non-rock regions. A grid search over training parameter values identified optimal training parameters for each feature subset.

The multi-scale features together achieve 97.3% labeling accuracy; this combination is significantly better than for any one or even any pair of attribute sets. The accuracy percentages of this diagram suggest that all feature vectors contribute to the final performance result.

The impact of global features is also apparent in the segmentation. Figure 8 shows two segmentations of the same image using only texture features (left) and the entire feature set (right). Global features appear to improve boundary localization; the contextual cues like shape become significant when choosing between two very similar alternative segmentations.

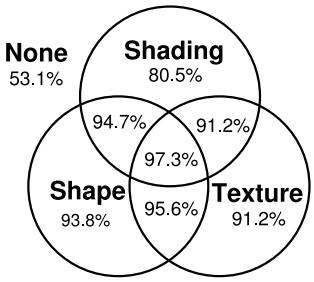

Figure 7. Labeling accuracy for different attribute sets. The *None* feature corresponds to a blind policy that always chooses the most frequent class.

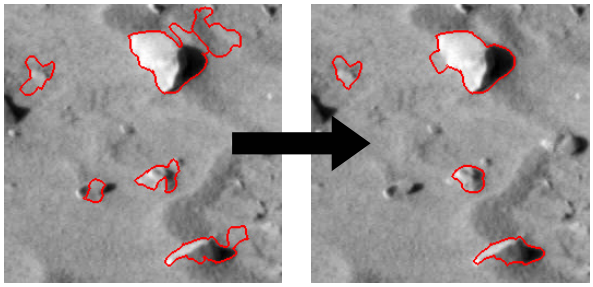

Figure 8. Typical segmentation result using local texture features (left) and the entire feature set (right). Global features like shape provide important cues to localize rock boundaries.

3.2.2 Rock Detection Accuracy

Previous work has used several different performance measures for rock detection and segmentation [7, 4]. We employ an "area of overlap criterion" to reject detections whose borders do not accurately match those of the true rock. We match detections to their most similar true rocks; successful detections are those whose intersection area constitutes 50% or more of their union.

We generate precision and recall scores by varying the confidence threshold for preserving rocks in the final segmentation; this value can be tuned to induce high-precision or high-recall behavior. The *precision* is the percentage of true rocks among candidate segmented rocks, while the *recall* is the percentage of rocks segmented from the total actually present in the scene. The confidence threshold only matters for accurately-segmented rocks; many rocks are never recovered during the segmentation process so no threshold setting ever achieves full recall. We ignore rocks adjacent to the image border to avoid counting partial rocks whose true attributes cannot be determined due to cropping. We also discount any detections contained within these border-adjacent rocks.

Figure 9 shows the precision and recall scores for a test set of 56 images containing over 230 rocks. There are rel-

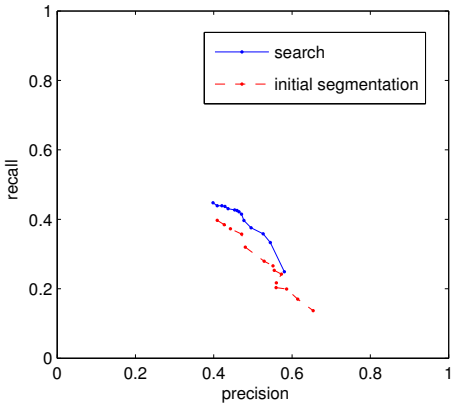

Figure 9. Precision and recall for rock detection using different confidence threshold values. The recall values appearing in the figure are the highest reached under any confidence threshold for this test set.

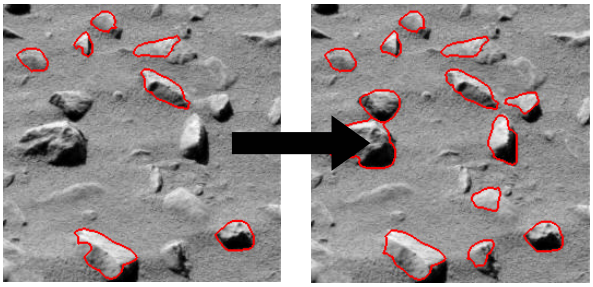

Figure 10. The effect of the configuration search on rock segmentation results. The initial segmentation (left) loses rocks whose probabilities fall below the confidence threshold. The search process recovers their boundaries and improves their score (right).

atively few rocks per image, so performance varies significantly from one image to the next. The standard deviation for search performance generally ranges between 0.2 and 0.3 for both recall and precision. Nevertheless, the test suggests that the segmentation search provides a modest improvement in both precision and recall. This corroborates visual evidence that the configuration search improves the fidelity of the segmentation (Figure 10).

3.2.3 Boundary Localization Accuracy

The Chamfer distance measures how closely the detected rock borders match the actual rock borders [3]. This value is the average distance from a detected region boundary pixel to the closest boundary pixel on the real rock region. The results of this performance measure appear in Figure 11. Only shape matches from successful detections are shown. For this data set the search procedure provides no apparent advantage for boundary localization beyond that which

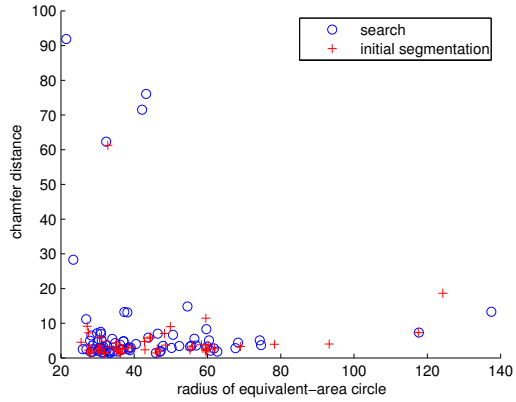

Figure 11. Chamfer distance boundary localization score for a fixed confidence threshold.

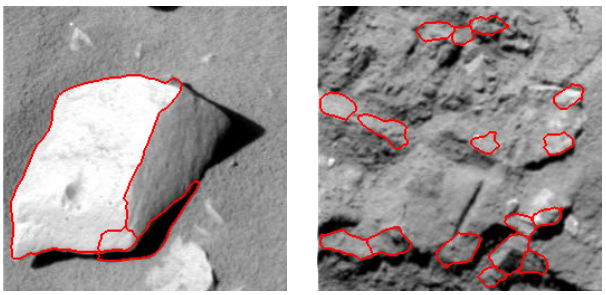

Figure 12. Two failure modes: large rocks (left) and structured soil (right).

is required to detect the rocks in the first place. In other words, the search increases the number of rocks that are discovered but it does not improve their boundary localization accuracy.

4. Discussion

Several scene factors influenced the quality of resulting detections. Figure 12 shows some examples of failure modes. Large rocks are rare in the training set; the method often splits them into pieces or fails to find them altogether. The greedy search cannot find the largest rocks; these require many superpixel additions. Such piecemeal construction involves short-term sacrifice — a non-ellipsoidal shape, for example — that the hill-climbing approach avoids. Structured soil also proved difficult to disambiguate. The rightmost image shows soil that has been compressed by the rover's wheel; there are no rocks in the image but the system returns several false positives due to spurious shadows and patches of unusual texture.

Despite these obstacles, our multi-scale segmentation technique provides promising results for object identification in natural scenes. Our work draws together a wide

range of recent advances in object recognition, including superpixel representations, texton analysis, and simultaneous segmentation and recognition. We demonstrate the applicability of these ideas to the difficult problem of rock detection and segmentation. Rocks in Mars images have the unique challenges of directional lighting, variable object morphology and weak local cues. We have shown how such objects can be segmented by combining local, object-level and scene-level features in a multi-scale technique.

There are several avenues for future development of this work. A long-term goal of automatic geologic analysis is a system that could operate onboard a rover for autonomous science decision making. This algorithm is too computationally-intensive for onboard applications, but faster methods may exist. Texture characterization by texton analysis is currently very common; however, there are many different types of filter banks and analysis methods that could be used and a more detailed examination is desirable. The superpixel segmentation used here was selected after a qualitative comparison with others readily available; a more thorough analysis of different segmentation methods may also be appropriate.

Acknowledgments

Many thanks to Alexei Efros for his counsel. Image datasets were obtained from The NASA Jet Propulsion Laboratory (California Institute of Technology). This research was supported by NASA under grants NNG0-4GB66G and NAG5-12890.

References

[1] D. Wettergreen, N. Cabrol, S. Heys, D. Jonak, D. Pane, M. Smith, J. Teza, P. Tompkins, D. Villa, C. Williams, M. Wagner, A. Waggoner, S. Weinstein, and W. Whittaker, "Second experiments in the robotic investigation of life in the atacama desert of chile," in *ISAIRAS*, vol. 8, 2005.

[2] J. Fox, R. Castaño, and R. C. Anderson, "Onboard autonomous rock shape analysis for mars rovers," in *IEEE Aerospace Conference Proceedings, Big Sky, Montana*, March 2002.

[3] H. Dunlop, *Automatic Rock Detection and Classication in Natural Scenes*, vol. CMU-RI-TR-06-40. August 2006.

[4] D. R. Thompson and R. Castaño, "Automatic detection and classification of geological features of interest," in *IEEE Aerospace Conference Proceedings, Big Sky Montana*, March 2007.

[5] V. Gor, R. Castaño, R. Manduchi, R. C. Anderson, and E. Mjolsness, "Autonomous rock detection for mars terrain," in *Proceedings of AIAA Space 2001*, August 2000.

[6] A. Castaño, R. C. Anderson, R. Castaño, T. Estlin, , and M. Judd, "Intensity-based rock detection for acquiring onboard rover science," in *Lunar and Planetary Science*, no. 35, 2004.

[7] D. R. Thompson, S. Niekum, T. Smith, and D. Wettergreen, "Automatic detection and classification of geological features of interest," in *IEEE Aerospace Conference Proceedings, Big Sky Montana*, March 2005.

[8] G. Mori, X. Ren, A. Efros, and J. Malik, "Recovering human body configurations: Combining segmentation and recognition," in *IEEE International Conference on Computer Vision*, 2004.

[9] Z. Tu and S.-C. Zhu, "Image segmentation by data-driven markov chain monte carlo," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24:5, 2002.

[10] X. Ren and J. Malik, "Learning a classification model for segmentation," in *Proc. 9th Int. Conf. Computer Vision*, vol. 1, pp. 10–17, 2003.

[11] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," in *IEEE Trans. Pattern Anal. Machine Intell*, vol. 24, pp. 603–619, 2002.

[12] C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blobworld: Image segmentation using expectation-maximization and its application to image querying," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24:8, pp. 1026–1038, 2002.

[13] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," in *International Journal of Computer Vision*, vol. 59:2, September 2004.

[14] G. Mori, "Guiding model search using segmentation," in *IEEE International Conference on Computer Vision*, 2005.

[15] J. Shi and J. Malik, "Normalized cuts and image segmentation," in *IEEE Trans. PAMI*, vol. 22:8, pp. 888–905, 2000.

[16] D. R. Martin, C. C. Fowlkes, and J. Malik, "Learning to detect natural image boundaries using local brightness, color, and texture cues," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, January 2004.

[17] M. Peura and J. Iivarinen, "Efficiency of simple shape descriptors," in *Proceedings of the Third International Workshop on Visual Form*, pp. 443–451, May 1997.

[18] V. Mikli, H. Kaerdi, P. Kulu, and M. Besterci, "Characterization of powder particle morphology," in *Proc. Estonian Acad. Sci. Eng.*, vol. 7, pp. 22–34, 2001.

[19] M. Varma and A. Zisserman, "A statistical approach to texture classification from single images," in *International Journal of Computer Vision: Special Issue on Texture Analysis and Synthesis*, vol. 62:1-2, pp. 61–81, 2005.

[20] R. Zhang, P.-S. Tsai, J. Cryer, and M. Shah, "Shape from shading: a survey," in *IEEE transactions on pattern analysis and machine intelligence*, vol. 21:8, pp. 690–706, 1999.

[21] C.-C. Chang and C.-J. Lin, *LIBSVM: A Library for Support Vector Machines*, 2001.