Generalizing an Automated Detection Pipeline for Possible Indicators of Subsurface Water on Mars

Matthew Wu, Kiri Wagstaff, and Brian Bue

Jet Propulsion Laboratory, California Institute of Technology

SURF 2017 Final Report

Abstract

Recurring slope lineae (RSL) are dark streaks on the slopes of Mars that grow or shrink with the Martian seasons. They may be caused by subsurface salty water. Current methods for automatically detecting RSL have difficulty generalizing to new sites of Mars, because different sites may have very different kinds of terrain (for example, crater vs. mountainous). Our approach is to perform terrain classification and identify a subset of the terrain classes that is most likely to have RSL. Then, the performance of an RSL detection pipeline could be improved by filtering out non-RSL-relevant terrain classes.

Using this approach, we conduct preliminary terrain classification experiments of two RSL-rich sites with contrasting terrain. We compute terrain features like slope, local binary patterns, and local standard deviation of slope, and use the k-means method to do the classification. These experiments show that identifying a set of RSL-forming terrain classes that generalizes across multiple sites is possible. There are many directions for future work, such as using more sites and trying out new features.

1 Introduction

Recurring slope lineae (RSL) are dark, narrow (0.5-5m wide) streaks on the surface of Mars that have three characteristics: they appear to lengthen during warm seasons, fade during colder seasons, and recur every Mars year [1, 2]. They have been found in images from HiRISE, a camera on the Mars Reconnaissance Orbiter. An example of RSL is shown in Figure 1.

Little is known about RSL; there are many hypotheses for what causes their formation. The best explanation seems to be that RSLs are caused by water in a briny saturated medium 2-17cm below the surface [1]. RSLs appear dark because water is wicked up to the surface by capillary forces. Other hypotheses do not involve liquid water. For example, one

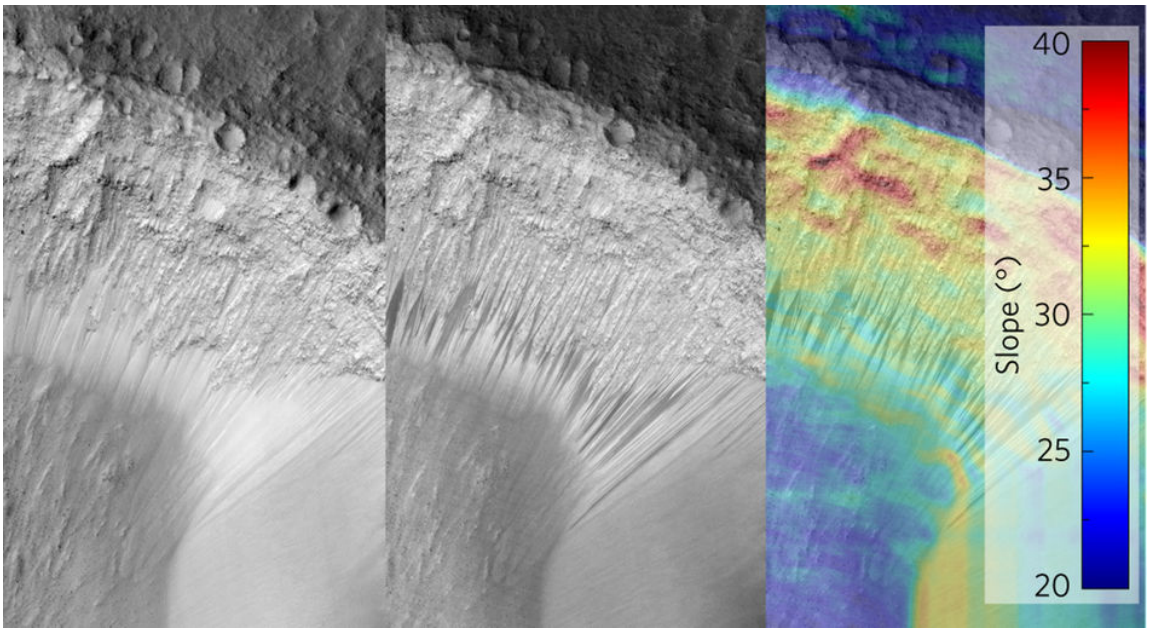

Figure 1: HiRISE images of RSLs in Garni Crater, Mars. (The RSLs are the dark streaks in the middle image that are not present in the left image.) Left image was taken at L_s (solar longitude) = 191.5°; middle was taken at $L_s = 280.8°$. Right is the slope, in degrees, of the area. From [3].

such "dry" hypothesis is that RSL are caused by wasting of a small layer of some surface material [1]. However, these dry hypotheses do not seem able to explain why RSL would recur in the same spot year after year, and why their shrinking and lengthening coincides with the seasons. Because of their possible relationship with water, RSL have drawn interest as one of the most significant pieces of evidence that there may be liquid water on Mars today.

1.1 Automated RSL detection

In order to learn more about RSL, we would like to catalog each RSL streak. This is difficult to do by hand, because there are many HiRISE images, each of which can contain hundreds of RSL. Thus, finding a way to automatically detect RSL would greatly benefit RSL research. Previously, our group developed a system to automatically detect RSL in HiRISE images [4]. A short overview of how it works is:

1. Identify "RSL candidates," which are any streaks in the HiRISE images that are darker than their surroundings. We use information like the elevation, the sun position, and the camera observation position to rule out shadows.

2. Classify these RSL candidates as real or spurious, with machine learning. We integrate data about the underlying terrain, such as elevation, slope, etc. into the classification decision. We use active learning; the computer presents the RSL candidate it is least sure about to a human scientist. The scientist gives an expert classification, which the computer uses to re-train its model and present the next RSL candidate, et cetera. The process stops when the scientist is satisfied with the classifications.

The system worked well on a region with many RSL called Garni Crater. Of the 714 RSL candidates in an image of Garni Crater, an expert only needed to label 31 of them before the algorithm accurately classified all of them [4]. However, the system has difficulty when detecting RSL in new sites, which can have very different terrain. For example, another site with many RSL is Coprates Montes (Figure 2b). Unlike Garni Crater (Figure 2a), Coprates Montes has mountainous terrain and much smaller RSL.

1.2 Automated Mars terrain classfication

Current observations suggest that RSL tend to occur on certain types of terrain, specifically smooth slopes. We believe that automated terrain classification may be helpful to identify these terrain types.

This idea is based on earlier work on automated Mars terrain classification by Bue et. al. (2006) [5]. The study used Mars elevation maps collected by the Mars Orbiter Laser Altimeter on the Mars Global Surveyor. For each pixel of the image, topographic features, such as slope and elevation, were computed. Using unsupervised learning, the pixels are clustered into terrain classes, which are groups of pixels with similar attributes.

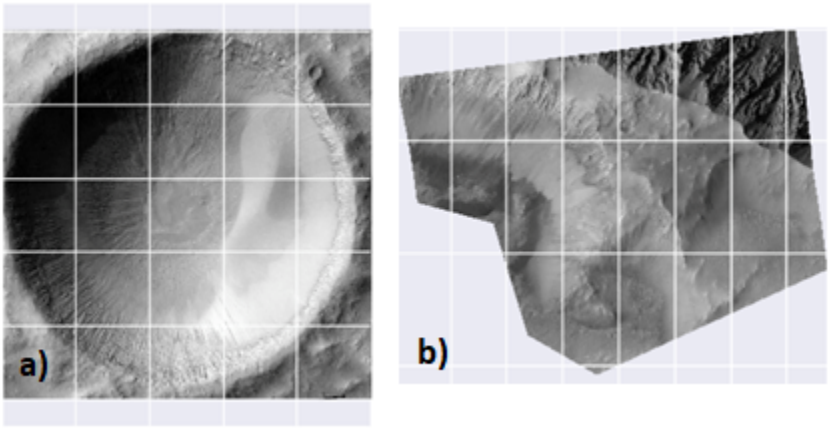

Figure 2: HiRISE grayscale images of a) Garni Crater and b) Coprates Montes.

The features were designed to capture geomorphic properties, so the resulting classes are easily interpretable.

An important part of terrain classification is choosing terrain features. We are interested in topographic features, like those used by Bue et. al. (2006), but we are especially interested in doing terrain classification using textural features, to identify smooth regions.

1.3 Approach

As mentioned, this project's approach is to do automated terrain classification of RSL-rich regions. This approach accomplishes two purposes:

1. Learn about what kinds of terrain have RSL. This could be done by examining the properties of the terrain classes with the most RSL. To our knowledge, there has not been a study that quantifies the characteristics of RSL-forming terrains. So, the terrain characteristics of RSL would be an important discovery in itself and could be valuable for understanding what causes RSL.

2. Improve performance of automated RSL detection. This could be done by, say, filtering out terrain classes with *very few* RSL as an initial step in an automated RSL detection system.

2 Methods

2.1 Terrain classification pipeline

We developed a flexible pipeline that computes terrain classes and determines which classes are likely to form RSL, based on prior data of actual RSL locations. The pipeline also

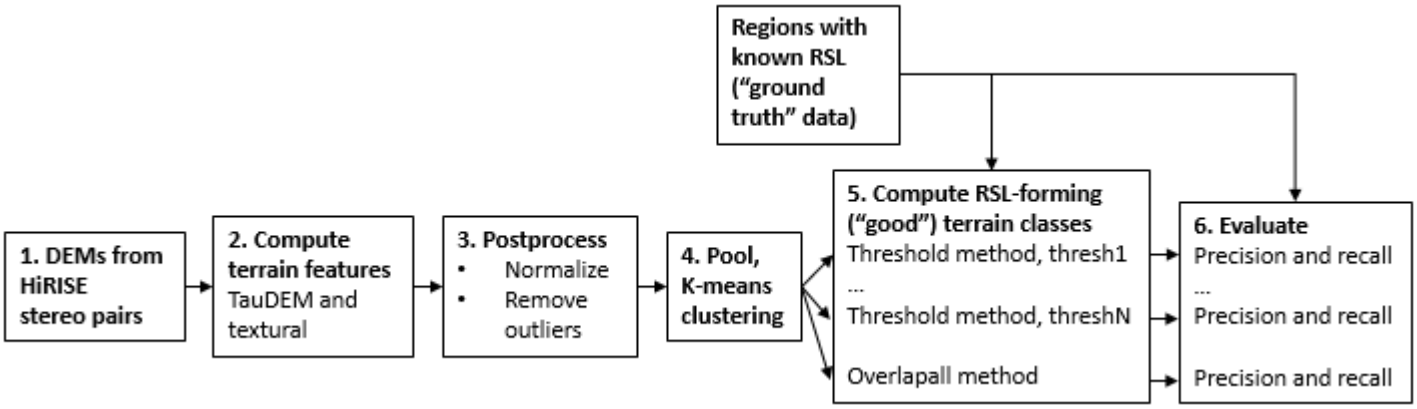

Figure 3: Diagram of terrain classification pipeline.

evaluates how well the resulting RSL-likely/not RSL-likely classification matches up with the actual data. The pipeline has six steps (Figure 3).

Step 1. Digital elevation maps We obtain digital elevation maps (DEMs) of each site from stereoscopy of HiRISE images. For this project, we used DEMs of Garni Crater and Coprates Montes (their grayscale images are in Figure 2). The DEMs are downsampled to 1.01m/pixel, to reduce computation time. Pixels in areas of the DEM with artifacts are discarded.

Step 2. Compute terrain features Then, we compute terrain features on a pixel-wise basis for each DEM. Not all features have to be used.

TauDEM features We use the software package TauDEM [6] to compute some basic topographic features:

- Slope magnitude, which is rise/run in the steepest direction.
- Slope direction, which is the angle counterclockwise from east of the steepest direction.
- Drainage—the number of pixels "sloping down" into a pixel, based on a spanning tree construction.

Textural features Treating the elevation map as a 2D array of intensities, we also compute two kinds of textural features: local binary patterns (LBP) and local standard deviation of slope magnitude.

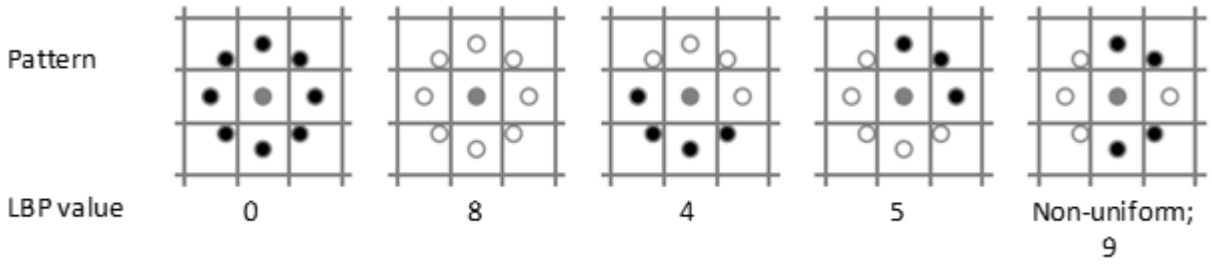

Figure 4: Some example LBP patterns and their given values, with $P = 8$. Based on an image from [9].

Local binary patterns We use the "uniform" variant of LBP, which is rotation-invariant [7]. The feature has parameters P and R. Each pixel is compared with P points equally spaced on a circle of radius R around it to see if they are higher (1) or lower (0) than the central pixel. This is represented as a circular pattern of P 0's and 1's around the pixel (Figure 4). Patterns with no more than two 0-1 transitions are called "uniform."

Based on the pattern, each pixel is given a value between 0 and $P + 1$ inclusive.

- Pixels with a uniform pattern are given a value equal to the number of 1's in the pattern.
- Pixels with a non-uniform pattern are given the value $P + 1$.

It is easy to see that the values assigned by LBP are rotationally invariant.

We used the implementation of LBP in scikit-image [8] with the 'uniform' method to compute this.

Local stdev. slope magnitude Intuitively, this feature is a measure of roughness. It is computed from the slope magnitude map generated by TauDEM, and has one parameter, R, which is the size of the local window.

Specifically, the value of this feature at pixel (r, c) is the standard deviation of pixels in rows $r - R \ldots r + R$ and columns $c - R \ldots c + R$ (both ranges inclusive) of the TauDEM slope magnitude map. This is computed for all pixels whose windows are completely inside the bounds of the DEM.

Step 3. Postprocessing For each feature, we compute the 1st percentile and 99th percentile *across all sites*, and scale that to 0 and 1 respectively. We clip values less than 0 to 0, and values greater than 1 to 1, to remove outliers.

Figure 5: Estimated regions of Garni (left) and Coprates (right) where RSL have been found, which are used in the terrain classification pipeline. Same orientation as Figure 2.

Step 4. Pool and cluster We pool the pixels from all sites, and group them into k terrain classes using k-means clustering. We use the implementation of mini-batch k-means in scikit-learn [10].

Step 5. Compute RSL-forming terrain classes For each site, we have estimated the regions where RSL have actually been found (Figure 5). This represents the "ground truth" data of which pixels are actually RSL-forming.

Using this data, we determine which terrain classes are RSL-forming, or "good." There are many ways to make this decision. We use two different methods, "threshold" and "overlap-all."

- Threshold method: Good clusters are those in which the proportion of pixels that are actually RSL-forming is greater than T, for some threshold T. For each run of the terrain classification pipeline, multiple T can be used.
- Overlap-all method: Good clusters are those which overlap an RSL region by at least one pixel in each site.

Step 6. Evaluate For each set of good clusters resulting from step 5, we compute the precision and recall of the set of good clusters to the actual set of RSL-forming regions. In a way, this is the "training error" of the given assignment of good clusters.

2.2 Experiments

In separate experiments, we ran the terrain classification pipeline with the LBP and local stdev. slope textural features.

Experiment 0 This experiment does not use textural features and consists of two runs. The first uses the DEM elevation itself as a feature, and also uses the TauDEM features [slope mag., slope dir., drainage]. The second is the same, except it does not use elevation.

In both runs, we used $k = 20$ clusters, and computed good clusters using the threshold method with $T = 30\%$.

Experiment 1 This experiment uses the LBP feature. We varied the following (36 runs in total):

- $k = 10, 20, 50$
- TauDEM features used: either [slope mag., slope dir., drainage], [slope mag., slope dir.], or [slope mag.] only
- LBP radius and # points $(R, P) = (2, 16), (3, 24), (5, 40), (10, 80)$. We kept a constant factor of $R = 8 * P$, since that seemed to worked well in the past.

We computed good clusters using the threshold method with $T = 30\%$, and using the overlap-all method.

Experiment 2 This experiment uses the local stdev. of slope feature. In all runs, the only TauDEM feature used was slope magnitude. We varied the following (9 runs in total):

- $k = 10, 20, 50$
- Stdev. slope radius $R = 3, 6, 9$

We computed good clusters using the threshold method with T = 15%, 22.5%, and 30%, and using the overlap-all method.

3 Experimental results

3.1 Experiment 0 (no textural features)

Table 1 summarizes the results for experiment 0. In run 1, since no clusters were above the threshold, we assign a precision and recall of 0.

We found that using elevation as a feature, as in run 1, is problematic. Garni and Coprates have very different absolute elevations (Figure 6). Thus, all Garni pixels appear to be "far" from, or "dissimilar" to all Coprates pixels, so the two sites end up having no terrain clusters in common. Since our goal is to find terrain classes that apply generally to all sites, using absolute elevation as a feature is probably unhelpful.

Run	Features	k	# good clusters	Precision	Recall
1	E, SM, SD, D	20	0	0	0
2	SM, SD, D	20	2	0.32	0.46

Table 1: Parameters and evaluation results for each run of experiment 0. For features, E = elevation, SM = slope magnitude, SD = slope direction, D = drainage.

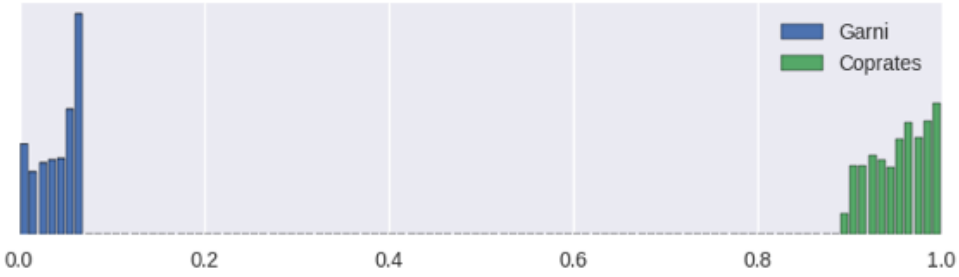

Figure 6: Joint histogram of elevation of Garni and Coprates, after normalization. Their elevation ranges are disjoint.

3.2 Experiment 1 (LBP)

Table 2 and Figure 7 summarize the results for experiment 1. There is a wide range of performance; using the 30% threshold method, many runs have precision and recall $> 30\%$. Many other runs have 0 precision and recall (because there were no terrain clusters above the threshold).

As an example of the outputs of a run, Figure 11 is a map of terrain classes and mask of RSL-forming classes for run 16, the highest-recall run under the 30% threshold method.

From Table 2 and Figure 7b, a striking trend in the overlap-all evaluation method can be seen: all runs that do not use slope direction (0-23) have recall very near 1. This means that for those runs, clusters which have RSL-forming pixels on one site also have RSL-forming pixels in another site. This property is desirable, since our goal is to find terrain classes that apply generally to all sites. Conversely, the property is not true for runs that use slope direction (24-35), which have recall < 0.98. Intuitively, if the slope direction feature is used, then regions that slope in very different directions cannot be in the same terrain class, even if their terrains are otherwise quite similar and both regions are RSL-forming. So the slope direction feature may not be helpful.

Another observed trend is that precision and recall increase as k increases (Figure 8). However, such performance gains could be overfitting, since the precision and recall are just a representation of the "training error." In the limit of (# clusters = # pixels), every pixel would be its own cluster, and precision and recall would be 100%, but that would clearly be overfitting.

Run	TauDEM features	LBP radius	k	Threshold 30%			Overlap-all		
				# good clusters	Precision	Recall	# good clusters	Precision	Recall
0	SM,D	10	10	0	0.00	0.00	9	0.10	1.00
1	SM,D	10	20	0	0.00	0.00	15	0.12	1.00
2	SM,D	10	50	2	0.34	0.37	37	0.12	1.00
3	SM,D	2	10	0	0.00	0.00	10	0.09	1.00
4	SM,D	2	20	0	0.00	0.00	19	0.10	1.00
5	SM,D	2	50	2	0.34	0.37	44	0.11	1.00
6	SM,D	3	10	0	0.00	0.00	10	0.09	1.00
7	SM,D	3	20	0	0.00	0.00	17	0.10	1.00
8	SM,D	3	50	3	0.34	0.36	42	0.11	1.00
9	SM,D	5	10	0	0.00	0.00	9	0.10	1.00
10	SM,D	5	20	0	0.00	0.00	16	0.10	1.00
11	SM,D	5	50	3	0.34	0.38	42	0.11	1.00
12	SM	10	10	0	0.00	0.00	6	0.12	1.00
13	SM	10	20	0	0.00	0.00	14	0.11	1.00
14	SM	10	50	1	0.39	0.33	31	0.12	1.00
15	SM	2	10	0	0.00	0.00	10	0.09	1.00
16	SM	2	20	1	0.30	0.47	17	0.11	1.00
17	SM	2	50	1	0.39	0.27	39	0.12	1.00
18	SM	3	10	0	0.00	0.00	9	0.10	1.00
19	SM	3	20	0	0.00	0.00	17	0.10	1.00
20	SM	3	50	1	0.40	0.29	37	0.11	1.00
21	SM	5	10	0	0.00	0.00	8	0.10	1.00
22	SM	5	20	1	0.33	0.42	14	0.11	1.00
23	SM	5	50	1	0.41	0.28	37	0.11	1.00
24	SM,SD,D	10	10	0	0.00	0.00	8	0.08	0.72
25	SM,SD,D	10	20	0	0.00	0.00	13	0.11	0.87
26	SM,SD,D	10	50	3	0.35	0.42	25	0.14	0.86
27	SM,SD,D	2	10	0	0.00	0.00	9	0.09	0.90
28	SM,SD,D	2	20	1	0.32	0.31	19	0.09	0.94
29	SM,SD,D	2	50	2	0.37	0.26	35	0.11	0.86
30	SM,SD,D	3	10	0	0.00	0.00	9	0.10	0.98
31	SM,SD,D	3	20	0	0.00	0.00	16	0.10	0.87
32	SM,SD,D	3	50	2	0.39	0.27	30	0.13	0.89
33	SM,SD,D	5	10	0	0.00	0.00	8	0.08	0.75
34	SM,SD,D	5	20	1	0.36	0.23	16	0.10	0.88
35	SM,SD,D	5	50	2	0.38	0.32	28	0.12	0.86

Table 2: Parameters and evaluation results for each run of experiment 1. For TauDEM features, SM = slope magnitude, SD = slope direction, D = drainage.

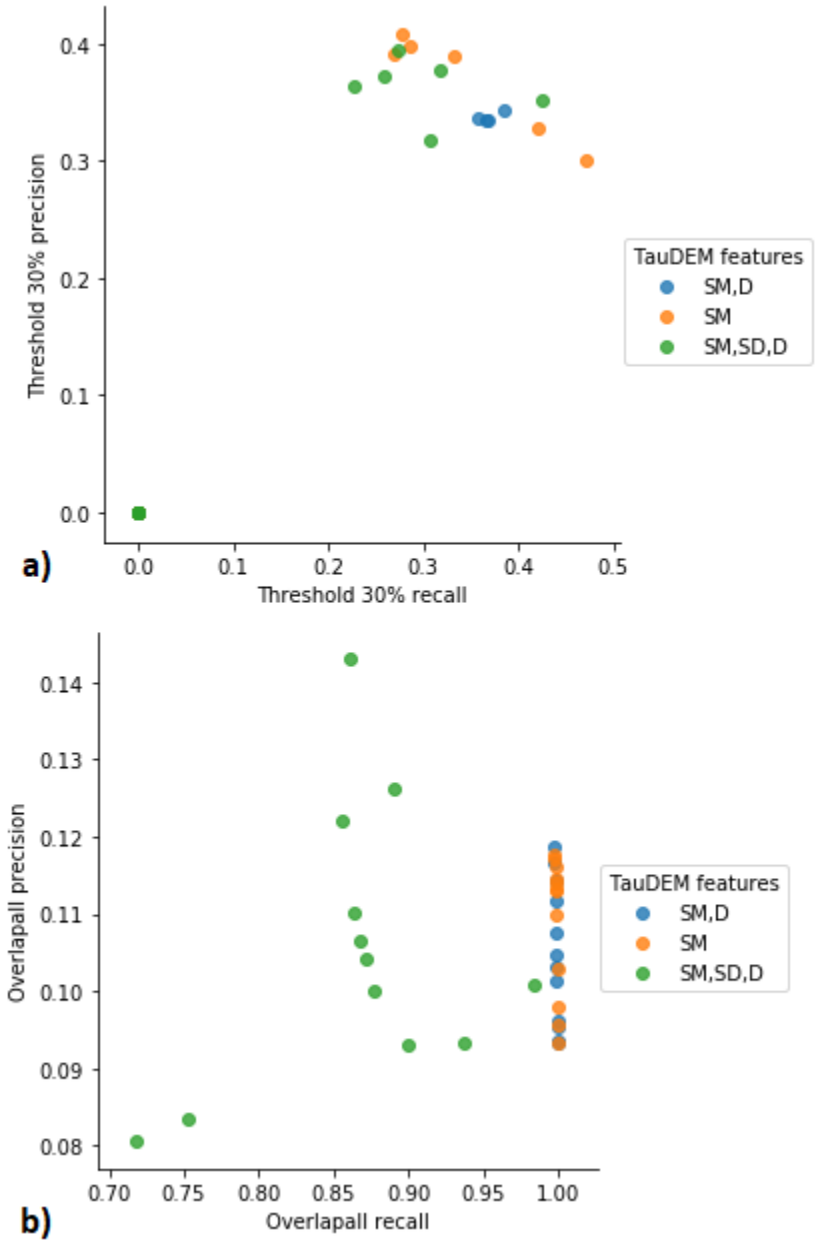

Figure 7: Precision vs. recall for the runs in experiment 1, using a) the threshold method with a 30% threshold and b) the overlap-all method. Points are colored by the set of TauDEM features used.

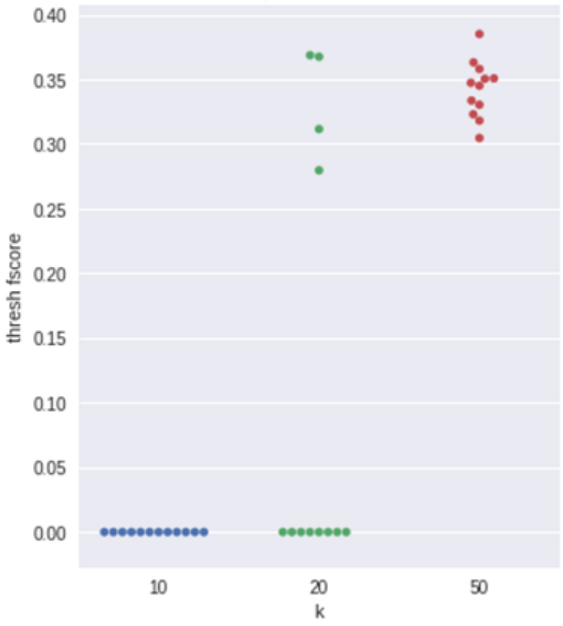

Figure 8: F-score vs. k, using the threshold method with $T = 30\%$, for all runs of Experiment 1.

We can analyze the choice of 30% threshold value using Figure 9. Some runs have clusters that are well over the threshold. In other runs, all clusters are well under the threshold. Changing the threshold, even by a small amount, will change the # good clusters and apparent precision/recall of some runs.

3.3 Experiment 2 (local stdev. slope)

Table 3 summarizes the results of Experiment 2. When using the threshold method, all runs have at least one good cluster and positive precision and recall, unlike experiment 1. Also, for each run, decreasing the threshold causes an increase in the number of RSL-forming clusters, an increase in recall, and a decrease in precision, which is expected.

Additionally, all runs have an overlap-all recall of near 1 (Table 3), which means that the resulting terrain clusters may generalize well.

We have known that there are data artifacts in the elevation map of Garni Crater. However, during this experiment, we found that the artifacts are easily seen in a map of the local stdev. slope feature (Figure 10). The artifacts are due to stereo imaging limitations: some surfaces of Garni are quite smooth, and stereo imaging has difficulty resolving the elevations of very flat surfaces. Whether they have an adverse effect on terrain classification is to be determined.

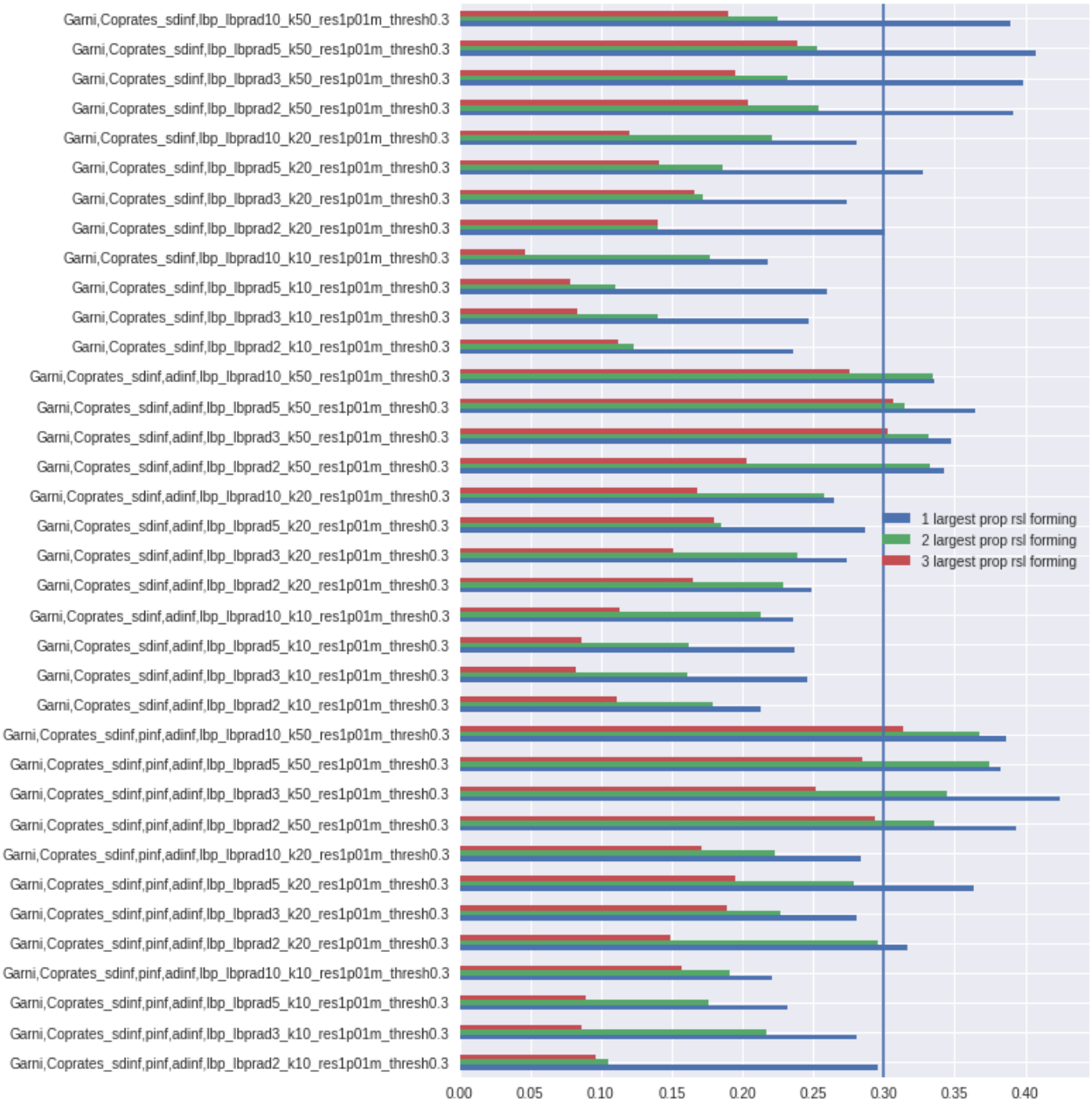

Figure 9: Proportion overlap of cluster with the actual RSL regions, for the 3 clusters with highest such overlap in each run of experiment 1. The vertical line is at 30%, which corresponds to the 30% threshold value.

Run	TauDEM features	Stdev. slope radius	k	Overlapall			Threshold 15%			Threshold 22.5%			Threshold 30%		
				# good clusters	Prec.	Rcl.	# good clusters	Prec.	Rcl.	# good clusters	Prec.	Rcl.	# good clusters	Prec.	Rcl.
0	SM	3	10	9	0.12	1.0	2	0.25	0.65	1	0.31	0.32	1	0.31	0.32
1	SM	3	20	18	0.11	1.0	4	0.23	0.73	1	0.38	0.25	1	0.38	0.25
2	SM	3	50	42	0.12	1.0	8	0.26	0.66	3	0.38	0.36	1	0.52	0.21
3	SM	6	10	9	0.12	1.0	1	0.33	0.41	1	0.33	0.41	1	0.33	0.41
4	SM	6	20	18	0.11	1.0	3	0.27	0.63	2	0.31	0.45	1	0.38	0.25
5	SM	6	50	42	0.13	1.0	7	0.28	0.64	3	0.42	0.39	3	0.42	0.39
6	SM	9	10	9	0.11	1.0	2	0.27	0.65	2	0.27	0.65	1	0.36	0.29
7	SM	9	20	18	0.11	1.0	3	0.27	0.68	3	0.27	0.68	1	0.38	0.25
8	SM	9	50	40	0.12	1.0	7	0.28	0.68	4	0.36	0.49	2	0.49	0.26

Table 3: Parameters and evaluation results for each run of experiment 2. For TauDEM features, SM = slope magnitude.

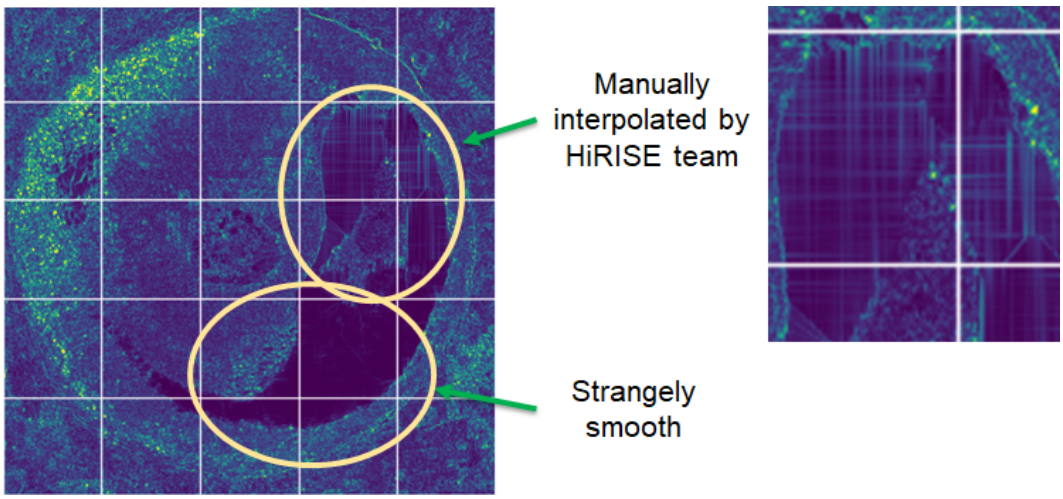

Figure 10: Map of stdev. slope ($R = 3$) of Garni Crater, showing areas with artifacts due to limitations in stereo imaging. An area on the right side of the crater shows cross-hatching, due to manual interpolation of poor-quality areas of the DEM by the HiRISE team. A lower right area is unusually smooth, with a very low stdev. slope.

4 Discussion and conclusion

This project is the first time that terrain classification was done to find RSL-forming terrain classes. The runs achieve a wide range of performance, suggesting varying ability to separate RSL and non-RSL-forming terrain. Both LBP and local stdev. slope features show promise. Using the 30% threshold classification method, more than 30% precision and recall was achieved with these features.

Using the overlap-all classification method, in many runs of experiments 1 and 2, recall is near 1, showing that it may be possible to have terrain classes that apply generally across sites. This seems to support our original assumption that there is correlation between RSL occurrence and the underlying terrain.

But these experiments are really just the beginning, and there are many possibilities for future work. One of the biggest limitations of the experiments is that only two sites were used. It is important that the input data to terrain classification is representative of the different kinds of terrain that have or do not have RSL on Mars. In our experiments, we tried to achieve this goal by using two sites with contrasting terrains (Garni Crater vs. the mountainous Coprates). Fortunately, DEMs are available for at least 5 more RSL-rich sites. So, it may be valuable to do further experiments on a larger, more comprehensive set of sites.

Another direction of future work is about terrain features. We tried TauDEM features, LBP, and local stdev. slope. It remains to perform experiments to determine which features are best and fine-tune their parameters. We can also try new terrain features, such as gray-level co-occurrence matrices [11], another textural feature.

Additionally, further investigation is needed of how to classify clusters as RSL-forming or not. So far, it seems like with the threshold method, changing the threshold can favor precision or recall in a controlled way. We also tried the overlapall method, where a large proportion of the clusters is designated "RSL-forming" and very high recall is achieved. We chose these methods for their simplicity. A next step may be to decide whether they are sufficient, or to devise entirely new methods.

5 Acknowlegments

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration (NASA) with funding provided by the NASA Mars Data Analysis (MDAP) program (solicitation NNH15ZDA001N-MDAP). Government sponsorship acknowledged.

Many thanks to Kiri Wagstaff and Brian Bue for their mentorship, and thanks to David Stillman (Southwest Research Institute) for his expert knowledge of RSL.

References

[1] D. E. Stillman, T. I. Michaels, and R. E. Grimm. "Characteristics of the numerous and widespread recurring slope lineae (RSL) in Valles Marineris, Mars". In: *Icarus* 285 (2017), pp. 195–210.

[2] D. Stillman et al. "Seasonal water budget suggests that a Valles Marineris Recurring Slope Lineae (RSL) site must be recharged by an aquifer". In: *Lunar and Planetary Science Conference*. Vol. 46. 2015, p. 2669.

[3] F. Schmidt et al. "Formation of recurring slope lineae on Mars by rarefied gas-triggered granular flows". In: *Nature Geoscience* 10.4 (2017), pp. 270–273.

[4] B. Bue, D. Stillman, and K. Wagstaff. *Recurring Slope Lineae (RSL) Detection & Characterization in HiRISE Images*. Presentation. 2016.

[5] B. D. Bue and T. F. Stepinski. "Automated classification of landforms on Mars". In: *Computers & Geosciences* 32.5 (2006), pp. 604–614.

[6] D. G. Tarboton. "The analysis of river basins and channel networks using digital terrain data". PhD thesis. Massachusetts Institute of Technology, 1989.

[7] T. Ojala, M. Pietikainen, and T. Maenpaa. "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns". In: *IEEE Transactions on pattern analysis and machine intelligence* 24.7 (2002), pp. 971–987.

[8] S. van der Walt et al. "scikit-image: image processing in Python". In: *PeerJ* 2 (June 2014), e453. ISSN: 2167-8359. DOI: 10.7717/peerj.453. URL: http://dx.doi.org/10.7717/peerj.453.

[9] *Local Binary Pattern for texture classification*. URL: http://scikit-image.org/docs/dev/auto_examples/features_detection/plot_local_binary_pattern.html.

[10] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[11] R. M. Haralick, K. Shanmugam, et al. "Textural features for image classification". In: *IEEE Transactions on systems, man, and cybernetics* 6 (1973), pp. 610–621.

A Example of terrain clusters and masks of RSL-forming clusters

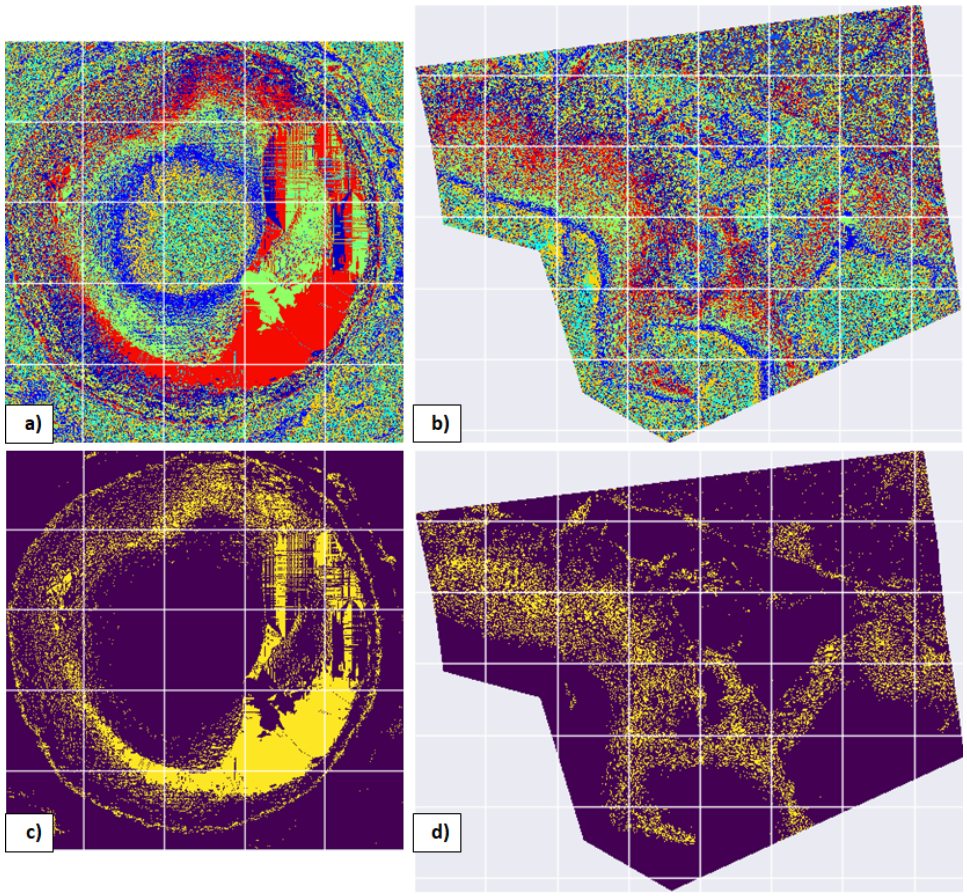

Figure 11: a) and b): Maps of terrain clusters, for a) Garni and b) Coprates, for run 16 of Experiment 1. Each cluster is drawn with a different color, and the clusters are shared across Garni and Coprates. c) and d): Masks of the RSL-forming clusters, using the 30% threshold method, for c) Garni and d) Coprates. Yellow regions are the "good" clusters; blue regions are not.