THEMATIC MAPS OF MARTIAN TOPOGRAPHY GENERATED BY A CLUSTERING ALGORITHM.

R. Vilalta, *Dept. of Computer Science, University of Houston, 4800 Calhoun Rd., Houston, TX 77204, USA, (vilalta@cs.uh.edu)*, T. F. Stepinski, *Lunar and Planetary Institute, Houston TX 77058-1113, USA, (tom@lpi.usra.edu)*.

Abstract. We have developed a method for autonomous construction of thematic maps of Martian terrain from digital topography. These maps show spatial distribution of topographical features and are generated by a computer algorithm that classifies pixels on the basis of topographical information they carry. We apply our technique to generate a thematic map of Tisia Valles region on Mars.

Introduction. The morphology of Martian terrain is of great interest because it helps to identify physical processes responsible for the observable topography. Traditionally, the descriptive method, applied to imagery data, has been used to study and categorize Martian terrain (1). We are developing a complementary approach, wherein a thematic map of topographical features is automatically constructed by clustering computer algorithm applied to perform an unsupervised classification of pixels in a digital elevation model (DEM) of a given Martian terrain. The input data is a DEM constructed from the Mars Orbiter Laser Altimeter (MOLA) data, and the output is the thematic map, wherein each pixel is assigned a label associated with a particular topographical feature.

To illustrate our technique we have constructed a thematic map of topographical features for Tisia Valles region (see Fig. 1). We use DEM with resolution of 1/128 degree in both latitude and longitude. The terrain shown in Fig. 1 is covered by a DEM with 385 rows (south to north) and 424 columns (west to east).

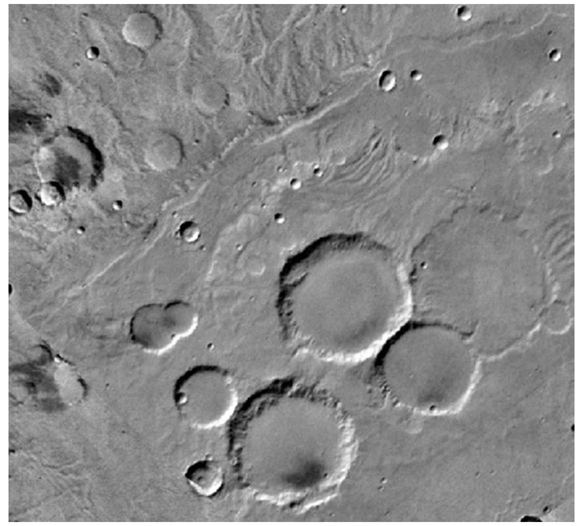
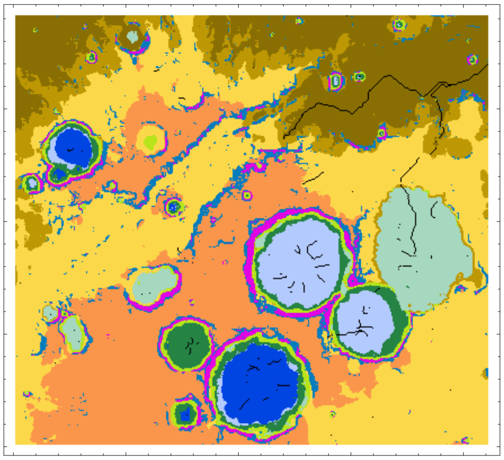


Figure 1: Viking image of Tisia Valles. The center of this image is located at 46.13E, 11.83S. The terrain shown is approximately 215 km west to east and 192 km south to north. This is a typical Noachian Martian surface, heavily cratered and showing presence of channels.

Methods. Digital topography, as represented by a DEM, is a digitized elevation field, where space is discretized into a 2-D grid and each cell (which we call a pixel) carries an elevation value. We modify a DEM, so, in addition to elevation, each pixel carries five other "bands" of information derived from the elevation field. Thus, we construct a "multiband DEM" in which each pixel carries topographical information in a form of a list of six numbers. Our approach is analogous to a concept of multispectral image where each pixel also carries a list of numbers describing intensity at number of wavelengths. The analogy to multispectral images extends to our method of analysis, we classify all pixels in a multiband DEM into a number of classes using clustering algorithm to produce thematic map of topography, whereas similar classification of pixels in multispectral image produces thematic map of composition.

The six bands of topographical information are: 1) elevation, 2) elevation difference between flooded and original terrain, 3) slope in an original terrain, 4) slope in a flooded terrain, 5) contributing area in original terrain, 6) contributing area in flooded terrain. The original elevation field is "flooded" in order to make it fully drainable (2). On Mars the presence of craters and other natural pits makes flooded and original terrain different. The difference between the two terrains is the second band in multiband DEM. Pixels with high values in the second band are likely to be located inside craters. The third and fourth information bands are slopes in original and flooded terrain, respectively. The slope of a given pixel is the slope between its center and the center of neighboring pixel in the direction of steepest descent. Pixels that are flooded have zero slope. Pixels at the boundary, where slope cannot be determined have values of -1. The fifth and sixth information bands are contributing areas in original and flooded terrain, respectively. Contributing area is the number of pixels (including itself) that drains through a given pixel. Pixels with high values of contributing areas are likely locations of streams and rivers. Pixels with contributing area $= 1$ are ridges, nothing drains through them, except themselves. Note that information in bands 2, 5, and 6 is not local, but instead depends on geographical context of a pixel.

Clustering a large number of pixels is computationally expensive. Our approach sampled the data first to generate a smaller dataset as follows. The multiband DEM was divided into 10X10=100 equal sized cells. From each cell we sampled with replacement 400 pixels. The result is a dataset of 40,000 pixels. We cluster this dataset using probabilistic clustering algorithm that follows the Expectation Maximization (EM) technique (3). It groups vectors into classes by modeling each class through a probability density function. Each vector in the dataset has a probability of class membership and is assigned to the class with highest posterior probability. Euclidean metric in space of topography descriptors is used to measure the



CLASS	COLOR	PIXELS	TOPOGRAPHICAL INTERPRETATION	GROUPS
1		32%	Inter-crater plains located at medium-high elevations.	INTER-CRATER PLANES
5		21%	Inter-crater plains located at high elevations.	
8		13.2%	Inter-crater plains located at low elevations.	
11		7.5%	Inter-crater plains located at medium-low elevations.	
3		4.4%	Terrain inside craters located quite deep below rim, but not as deep as class 4.	INSIDE CRATERS
4		2.8%	Terrain inside craters located deepest below crater rim.	
7		2.7%	Terrain inside crater, located just below the crater rim.	
10		3.4%	Terrain inside crater, intermediate between classes 3 and 10.	
12		5.2%	Terrain inside shallow, partially buried craters.	
2		4.9%	Terrain located at valley ridges and outside crater ridges.	RIDGES
6		2.5%	Terrain located near the crater rim, but outside the crater.	
9		0.4%	Part of the landscape that constitutes the major drainage systems - channels.	CHANNELS

Figure 2: Thematic map of topography for Tisia Valles region. The actual map is on the left with different topographical features (classes) indicated by different colors. The summary of classification is given in the table on the right.

"closeness" between pixels. The number of classes is calculated using cross-validation (4). This algorithm classified all 40,000 pixels into 12 separable and exhaustive classes. Remaining 121,626 pixels were classified into those 12 classes using a decision tree learning algorithm. The algorithm used the initial dataset to construct decision tree. The decision tree classifier was used to classify remaining pixels into the same 12 classes representing different topographical features.

Results. Fig. 2 shows generated thematic topographic map of Tisia Valles region. Pixel class membership is indicated by color. The table on the right of Fig. 2 summarizes the result of our classification. The first column gives class number, the second column indicates the color used to denote pixels belonging to a given class on thematic map to the left. The third column gives the percentage of pixels in the class, the fourth column gives our topographical interpretation given to each class. The last column is the name of the larger group. We have organized 12 classes into four larger groups: inter-crater planes, regions inside the craters, ridges, and channels.

Some subtle differences between otherwise similar terrain are picked up by our classification. Five classes represent crater interiors; they discriminate between different crater depths. Four classes represent inter-crater plains, they differ by actual elevation. Two classes represent ridges, they discriminate between different slopes. Finally, a single class represents channels.

Conclusions. We have used Martian digital elevation data to construct a mulitiband DEM that carries enriched topographical information. Some data bands in such a DEM carry information integrated from a region around a pixel. This makes a pixel "aware" of its geographical context and makes possible using a mulitiband DEM to construct a useful thematic map. We have demonstrated that an autonomous, unsupervised clustering algorithm can classify pixels in mulitiband DEM into classes that corresponds to topographic features. Thus, we have developed a method that produces, in automated fashion, thematic map of Martian topography.

Such map contains "smart" information and it may support geomorphic investigations better than a visual rendering of a DEM. Because all information bearing pixels are labeled by their class membership, comparative statistical studies are possible. Distributions of particular variable (such as, for example, a slope) for different topographical features can be compared. A given terrain can be succinctly described by a list of, say, 12 numbers, each number giving a percentage of pixels in corresponding topographical class. For example, Tisia Valles terrain can be succinctly described by the following list: $(32, 4.9, 4.4, 2.8, 21, 2.5, 2.7, 13.2, 0.4, 3.4, 7.5, 5.2)$ This vector lists percentage of pixels in classes 1 to 12 as given in a table in Fig. 2. If pixels in other terrains are classified into the same 12 classes, a analogous vector can be constructed for each terrain. This opens possibility for classification of terrains (not pixels) based on quantitatively expressed similarity between their overall topographies.

References

(1). Scott, D.H., and Carr, M.H. 1977, Geological Map of Mars, *U.S.G.S. Misc. Geol. Inv.*, Map I-1093

(2). Tarboton, D.G., Bras, R.L., and Rodriguez-Iturbe, I. 1989, Technical Report no. 326, Ralf M. Parsons Lab., Massachusetts Institute of Technology, Cambridge, MA.

(3). Witten, I.H. and Frank, E. 2000, Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations. Academic Press, London U.K.

(4). Cheeseman, P. and Sturz 1995, Bayesian Classification (AutoClass) Theory and Practice. Advances in Knowledge Discovery and Data Mining. MIT Press