EVOLVING AUTOMATED FEATURE EXTRACTION ALGORITHMS FOR PLANETARY SCIENCE.

S. P. Brumby[1], C. S. Plesko[2], E. Asphaug[3], [1]Los Alamos National Laboratory, brumby@lanl.gov, [2]University of California, Santa Cruz, cplesko@es.ucsc.edu, [3]University of California, Santa Cruz, asphaug@es.ucsc.edu.

Introduction: Planetary exploration missions have returned a wealth of imagery data over the last 40 years. The problem is how to make best use of it all. Thoroughly analyzing such large datasets manually is impractical, but developing handwritten feature extraction software is difficult and expensive. The current project explores the use of machine learning techniques to automate the development of feature extraction algorithms for the Mars Orbiter Camera (MOC) narrow angle dataset using Los Alamos National Laboratory's GENIE machine learning software. GENIE uses a genetic algorithm to assemble feature extraction algorithms from low-level spatial and spectral image processing steps. Each algorithm is evaluated against user-provided training data, and the most accurate ones are allowed to "reproduce" to build new solutions. The result is automated feature extraction algorithms customized to the dataset at hand and the current feature of interest. A graphical user interface is used to provide training data, allowing map-makers without programming experience the ability to generate new feature extraction algorithms.

Mars Global Surveyor (MGS) [1] has been studying Mars since 1997. The narrow angle dataset produced by the MOC [2] provides imagery with a spatial resolution of approximately 3 meters/pixel in a broad visible/near-infrared spectral range (0.50 □ 0.90 □m). Since its arrival, MOC has taken over 112,000 images, which have been used to study various planetary processes. Craters were selected as our feature of interest because they are a easily recognizable feature that can be used to derive important information about a surface [3-4].

GENIE: GENIE [5-8] uses techniques from genetic algorithms (GA) [9-11] and genetic programming (GP) [12] to construct spatio-spectral feature extraction algorithms for multi-spectral remotely sensed imagery. Both the algorithm structure and the parameters of the individual image processing steps are learned by the system. GENIE has been described at length elsewhere [5-8], so we will only present a brief description here. In particular, the present work explores applying GENIE to panchromatic imagery [13-14].

GENIE begins by randomly generating a population of candidate image-processing algorithms from a collection of spectral and textural image processing operators, including local neighborhood statistics, texture measures, spectral band-math operations (e.g. ratios of bands), and gray-scale morphological filters with various shapes of structuring elements. Each candidate algorithm consists of a number of these image-processing operators, which together generate a vector of processed images in an intermediate, non-linear feature space. These are combined using a Fisher linear discriminant to produce a single gray-scale result image in which bright pixels indicate the presence of the feature of interest. This gray-scale result is converted to a Boolean classification using an optimal threshold [15]. The parameters of the Fisher discriminant and threshold are based on training data provided by the human user via GENIE's graphical interface.

Our fitness metric for evaluating candidate image-processing algorithms measures the total error rate (false positives and false negatives) calculated from the training data. After a fitness value has been assigned to every candidate algorithm less fit members of the population are discarded. A new population is generated by allowing the most fit members of the old population to reproduce with modification via the evolutionary operators of mutation and crossover. To ensure a monotonic increase in fitness the most fit individual in the current population is kept without modification (principle of elitism). This process of fitness evaluation and reproduction with modification is iterated until the population converges, or some desired level of classification performance is attained, or some user-specified limit on computational effort is reached (e.g., a limit on the number of candidate algorithms evaluated). This boolean threshold on the best image processing algorithm returned by GENIE may be adjusted by the user to re-adjust the emphasis of detection rate (true positives) over false alarms and missed detections.

The genetic algorithm used by GENIE is implemented in object-oriented Perl. This language provides a convenient environment for the string manipulations required by the evolutionary operations, and for accessing the underlying operating system. Evaluation of the candidate image-processing algorithms is the computationally intensive part of the evolutionary process. GENIE's genetic algorithm writes code implementing candidate algorithms in the IDL image analysis language (IDL is a product of Research Systems, Inc.), and sends them to an IDL session for evaluation. IDL does not provide all the image processing operators we need, so we have implemented additional operators in a library of C code that can be called from within the IDL environment. Individual image processing steps correspond to primitive image operators, which are coded as IDL procedures. A candidate algorithm is a sequence of calls to primitive operators, and is implemented as a sequence of lines of code in an IDL batch executable. Fitness evaluation is an inherently parallelizable process, and GENIE is able to exploit networks of workstations to speed up its evolution of algorithms. Our graphical interface for supplying training data and examining results (called 'Aladdin') is written in Java. GENIE was developed on and for Linux/Intel platforms, and has since been ported to Sun's Solaris operating system.

Results: We selected a training image, MOC image M0803054 [16], near Louros Valles (8.5S, 82.0W), to present a reasonably homogeneous terrain marked by a number of bowl shaped craters. GENIE was trained on the first 930 rows of pixels in the image (the image is 830 pixels wide), based on a manual analysis (Fig. 1) in which some of the bowl shaped craters and some of the non-crater surface features are marked. The next 970 rows of this scene were also analyzed, to serve as a test scene (Fig. 2).

GENIE's genetic algorithm is a stochastic learning process, so individual results are likely to be highly variable in structure. GENIE was run 6 times, each time with a new population of 30 algorithms per generation, each run lasting for 50 generations. Running on standard Intel/Linux workstations, each run required 1 hour of wall-clock time. The best individual crater finding algorithm achieved a detection rate of 99% and a false alarm rate of 3%. On the test data, the performance of this algorithm dropped slightly.

The results of the individual runs were combined using three different voting schemes (Table 1) [17]. Classifier "Vote 1" uses a majority vote with contributions from all 6 classifiers. As expected, the false alarm rate reported by the voting set is substantially lower than the FAR reported by individual algorithms. This result raises the question, is the voting dominated by good individual algorithms? Classifier "Vote 2" is a majority vote with all classifiers except the strongest individual algorithm. Classifier "Vote 3" shows a "unanimous" voting decision rule for all 6 classifiers. For applications requiring a low FAR, this is a better classifier than individual results.

Future Work: This research presently focuses on Mars cratering as a testbed, but has obvious broader applications to other questions of planetary science, including analysis of cratering rates on satellites of Jupiter and Saturn, analysis of other landforms such as fractures in Europa's lithosphere, detection of changes over time on active planets such as Io, and location of landforms such as dune fields and mineral outcrops.

	Training Scene		Test Scene	
Class.	D.R.	F.A.R.	D.R.	F.A.R.
Best ind.	98.94	3.00	94.76	2.74
Vote 1	97.31	1.24	94.33	2.34
Vote 2	97.48	2.97	94.77	7.59
Vote 3	84.83	0.15	84.38	0.52

Table 1. Results of best individual algorithm and voting on training and test data. D.R. is percent detection rate, and F.A.R. is percent false alarm rate.

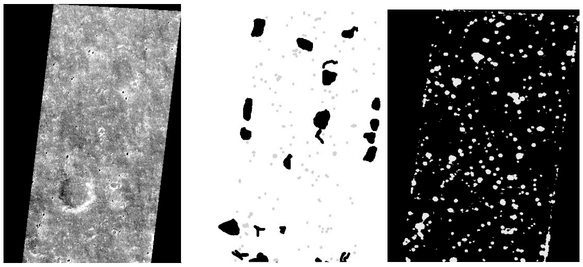

Figure 1. MOC Image M0803054: Training Scene, User-generated Training Data, result of first majority vote.

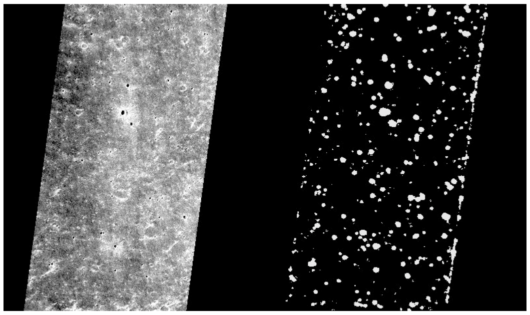

Figure 2. MOC Image M0803054: Test Scene 1, result of first majority vote.

Conclusion: This study investigated the evolution of a voting set of crater finding algorithms for application to the Mars Orbiter Camera narrow angle dataset. We described the results on training and test images. The algorithms are successful at detecting craters within the images, and generalize well to an image that they have not seen before. We find these results to be encouraging for the application of GENIE to the MOC panchromatic dataset.

References: [1] Albee, A.L. et al. (1998) Science, 279, 13. [2] Malin, M.C., et al. (1992) JGR, 97(E5), 7699-7718. [3] Hartmann, W.K. et al. (1999) Nature, 397, 586-589. [4] Hartmann, W.K. et al. (2000) JGR, 105(E6), 15011-25. [5] Brumby, S.P. et al.(1999) Proc. SPIE, 3812, 24 - 31, (http://www.daps.lanl.gov/genie). [6] Theiler, J. et al. (1999) Proc. SPIE, 3753, 416-425. [7] Harvey, N.R. et al. (2000) Proc. SPIE, 4132, 72-82. [8] Perkins, S. et al. (2000) Proc. SPIE, 4120, 52-62. [9] Holland, J.H. (1975) Adaptation in Natural and Artificial Systems, U. Mich. [10] Rechenberg, I. (1973) Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution, Fromman-Holzboog. [11] Fogel, L. et al. (1966) Artificial Intelligence through Simulated Evolution, Wiley. [12] Koza, J.R. (1992) Genetic Programming: On the Programming of Computers by Natural Selection, MIT. [13] Plesko, C.S. et al. (2002) Proc. SPIE, 4480, 139-146. [14] Plesko, C.S. et al. (in press) Proc. SPIE, 4790. [15] For example, see Bishop, C.M (1995) Neural Networks for Pattern Recognition, Oxford Univ., 105 –112. [16] Malin, M.C. et al. (1999) http://photojournal.jpl.nasa.gov/, Catalog number M0803054, [17] R. O. Duda, R.O. et al. (2001) Pattern Classification, 2nd ed., Wiley-Interscience.