CASCADING CRATER DETECTION WITH ACTIVE LEARNING W. I. Miller[1], T. F. Stepinski[2], Y. Mu[1], and W. Ding[1] , [1] Department of Computer Science, University of Massachusetts Boston,100 Morrissey Blvd., Boston, MA 02125-3393(wimiller,yangmu,ding@cs.umb.edu), [2] Department of Geography, University of Cincinnati, Cincinnati, OH 45221-0131 (stepintz@uc.edu).

Introduction: Impact craters are among the most studied features found in planetary images because their counts are used to measure relative surface ages [1,2,3]. High resolution images contain very large number of craters making, in principle, possible to date surfaces with a high spatial and temporal resolution. However, this promise remains unfulfilled because of lack of effective algorithm for detecting craters in an image. A detection algorithm needs to identify craters, in broad spectrum of sizes and forms. It needs to do it reliably in large number of images showing terrains of different character. The challenge is to develop an algorithm that detects craters in all images with consistent accuracy and minimum human involvement.

In this paper, we report on our investigation of a promising strategy to address the aforementioned challenge. Our strategy consists of employing a cascading AdaBoost classifier for identification of craters in images, and using the Self-Organized-Map (SOM) as an active learning tool to minimize the number of image examples that need to be labeled by an analyst. The cascading AdaBoost classifier is a methodology that achieves an efficient and accurate recognition of objects in an image. An image (or its small portion referred to as a sub-window) is passed through a cascade of simple but progressively more complex classifiers, each optimized for rejection of non-crater sub-windows. This method has been proved to work very well with a problem of face detection and is expected to work well with crater detection. However, such classifier is only as good as its training set. If we apply such classifier to a new image that contains craters having features not present in the training set the classifier will fail. The active learning (through the SOM) is designed to point out sub-windows in a new image that are not represented in the training set, so an analyst can label them, add to the training set, and, by doing this, maintain accuracy of detection while minimizing a number of sub-windows that need to be viewed.

Methodology: First, we describe detection of craters by the cascading AdaBoost classifier. For this exploratory investigation we employed a following protocol: we have chosen four images of nearby and similar terrain and manually marked all craters on them. We call one image a “training” image and remaining images “test” images. The classifier is constructed using the training image and tested on both the training image and test images. Test images are needed to check the generalization power of the classifier – how well can it work on an image that was not used for its training?

The two images, each having 1,700 by 1,700 pixels, were extracted from the nadir panchromatic, 12.5 m/pixel image (h0905) of Mars taken by the High Resolution Stereo Camera (HRSC) on-board the Mars Express spacecraft in the area of Nanedi Valles. The strategy is to decompose each image into exhaustive set of sub-windows having sizes as small as 24 by 24 pixels and as large as the entire image; there are over 2 million such sub-windows in each image. Each sub-window is by a very large set (45,000) of image texture features that encode the content of the surface seen in this sub-window. We use the training set (sub-windows that are labeled to either contain craters or not) to train a cascading AdaBoost classifier [4, 5]. The cascade (which, in this context means a series of subsequent classifiers) is a way to efficiently handle a large number of features; each successive classifier is trained only on samples which pass through the preceding classifiers. The classifiers at the early stage of the cascade are designed to discard a large amount of non-crater sub-windows, whereas classifiers at the later stage of the cascade concentrate on accurate distinction between craters and non-craters.

Second, we describe our approach to maintaining accuracy through different images. The goal here is to be able to quickly identify sub-windows in the test image that has a character different from anything coded in the training set. In order to do it we construct a SOM from a combination of training set and sub-windows in the new image. Each member of the set is a multi-dimensional vector representing a sub-window. The SOM is a neural-net method for clustering and visualization of sets of multi-dimensional vectors. It consists of an array of “nodes”. The vectors are placed into the nodes on the basis of their similarity, thus a single node contains vectors similar to each other. Moreover nearby nodes are also similar to each other, thus two neighboring nodes contain vectors more similar to each other than to distant nodes. Overall, the SOM provides a visual representation of the topology of multi-dimensional space. In the context of the crater detection the SOM will provide a visual indication of the sub-windows in the test image that are not similar to either “crater” examples or “non-crater” examples in the training set. Once identified these sub-windows needs to be looked at and labeled.

Results: We labeled 168 sub-windows from the training site as craters and 628 sub-windows from the training site as non-craters; some examples of sub-windows labeled as craters and non-craters are given in Fig. 1. The cascading AdaBoost classifier was constructed using this training set and applied to all four images.

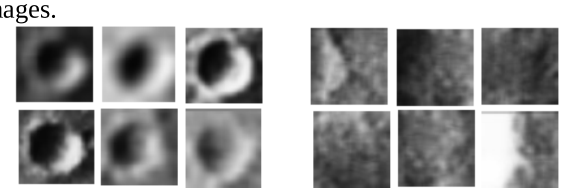

Figure 1. Left: examples of sub-windows containing crater; right: examples of sub-windows that do not contain craters.

To evaluate the performance of the classifier we measured the detection percentage D=100 (TP/(TP+FN)), the branching factor B=FP/TP, and the quality percentage Q=100 (TP/(TP+FP+FN)). Here, TP stands for the number of true positive detections (detected craters that are actual craters), FP stands for the number of false positive detections (detected craters that are not actual craters), and FN stands for the number of false negative "detections" (real craters failed to be indentified). D can be treated as a measure of crater-detection performance, B as a measure of delineation performance, and Q as an overall measure of algorithm performance. Fig.2 shows the visual indications of crater detection, as well as D, B, and Q for the four images we considered.

Conclusions: Automatic detection of small (200 m in diameter and larger) craters in images is a challenging task. As can be seen from Fig. 2 even the sophisticated cascading AdaBoost classifier does not find craters with great accuracy. This is because distinction between small craters and other terrain features are often fussy even for an analyst. Future research needs to establish a size limit of robust detectability of craters. The idea of active learning via SOM is promising but requires more research. Fig. 3 shows the SOM constructed from the training set, sub-windows representing craters and non-craters are visually separated showing potential for active learning.

Acknowledgements: This work was supported by NASA under Grant NNX09AK86G.

References:

[1] Salamuniccar, G. and Loncaric, S. (2008) *Advances in Space Research*, *42*, 6–19. [2] Urbach, E. R. and Stepinski, T. F. (2009) *PSS*, *57*, 880–887. [3] Martins, P., Pina, J. S. Marques, and Silveira, M. (2008) *IEEE Geoscience and Remote Sensing Letters*, *6(1)*, 127-131. [4] Viola P. and Jones M. (2004) *Inter. Journal of Computer Vision*, *57*, 137–154. [5] Ding, W. , Stepinski, T. F., (2010), the 19th ACM Inter. Conf. on Information and Knowledge Management.

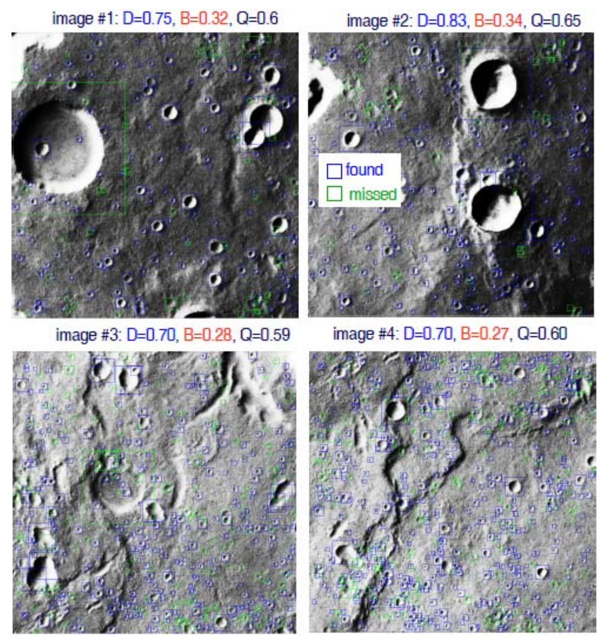

Figure 2. Crater detection results on 4 images considered in our evaluation.

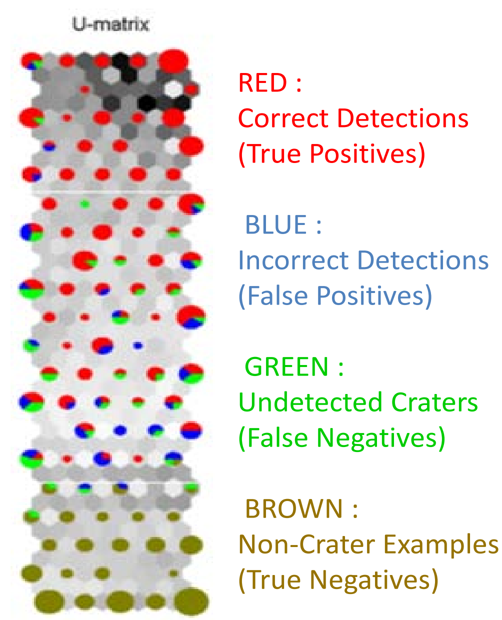

Figure 3. SOM showing the structure of the multi-dimensional training set.