LANDMARK CLASSIFICATION AND CONTENT-BASED SEARCH FOR MARS ORBITAL IMAGERY.

Kiri L. Wagstaff[1], Gary B. Doran[1], Ravi Kiran[1], Lukas Mandrake[1], Norbert Schorghofer[2], and Alice Stanboli[1]; [1]Jet Propulsion Lab., California Inst. of Tech., 4800 Oak Grove Drive, Pasadena, CA 91109 (kiri.wagstaff@jpl.nasa.gov), [2]Univ. of Hawaii, 2680 Woodlawn Drive, Honolulu, HI 96822 (norbert@hawaii.edu).

Introduction: Mars orbital mission archives continue to grow. For example, the HiRISE instrument on the Mars Reconnaissance Orbiter has returned >75,000 multi-band images at very high resolution (~0.3 m/pixel). While it is possible to search for images based on image parameters such as target location, season, or illumination angle, there is also a science-driven need to search for images that contain particular features of interest, such as craters or dark slope streaks. Currently this is done by time-consuming manual review of all possible relevant images.

We aim to enable *automated content-based search* through large volumes of orbital image data. Using machine vision and machine learning techniques, we have constructed a system that leverages a small initial investment of time in providing hand-labeled examples of surface features of interest to enable the automated classification of features in new and unseen images. We have integrated these landmarks into the Planetary Data System (PDS) web search interface to allow open access to content-based searching.

Approach: We used a salience-based detector to identify candidate surface features ("landmarks") within Mars orbital images, then manually labeled them by type. We used the labeled data set to train a machine learning classifier that could then predict the type of new landmarks in previously unseen images. We saved the detected and classified landmarks to a database that is now used by the PDS to provide content-based search in HiRISE images.

Landmark detection. The salience-based landmark detector improves on an existing contour-based salience detector [1] in three major ways. First, it uses a genetic algorithm to identify the optimal salience calculation as a combination of Canny edge detector and pixel-based salience. Second, it extracts a bounding box around the area of interest, which allows for the incorporation of nearby context when classifying the landmark. Finally, it employs an expanded list of descriptive attributes for each landmark that includes information about pixel intensities, the distribution of intensities, the dimension of the bounding box, 128 dense SIFT (Scale-Invariant Feature Transform) attributes [2], and 20 context attributes that capture the spatial distribution of bright and dark pixels within the landmark.

Landmark classification. The machine learning classifier is a multi-class Gaussian Naïve Bayes classifier. This model was chosen because it performed best compared to decision tree, random forest, linear SVM, and RBF SVM models. The classifier also provides the posterior probability (confidence) of its predictions. Since the classes of interest that we identified do not encompass all possible landmarks, we allow the classifier to *abstain* from generating a prediction if its confidence does not exceed a specified threshold.

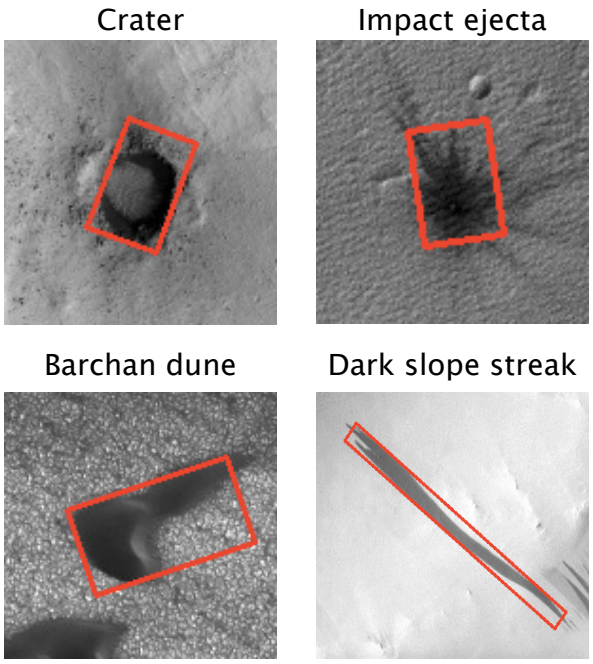

Figure 1. Examples of automatically detected and manually labeled landmarks.

Training Data Set: We assembled a data set containing regions from 65 full-resolution HiRISE images. These regions were chosen to provide good coverage of the landmark classes of interest (craters, impact ejecta, dunes, and dark slope streaks). The landmark detection system identifed 1014 landmarks within the 65 images. We developed a custom graphical user interface (GUI) to facilitate manual labeling of the landmarks. Not all of the detected landmarks qualified as one of the classes of interest. We obtained a total of 126 labeled landmarks (12 craters, 14 ejecta, 43 dunes, and 57 dark slope streaks). Figure 1 shows labeled examples from each class. We augmented this data set with 487 of the unlabeled landmarks (from other classes) assigned to a fifth class we called "None." The total data set contains 613 landmarks.

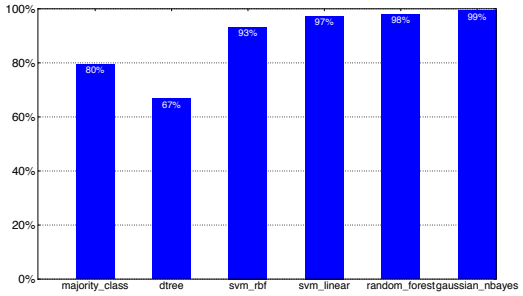

Figure 2. Classifier performance on 613 training examples, using 25% most confident predictions.

We also downloaded browse images for all 38,243 HiRISE images taken with the "RED" channel and applied the trained classifier to identify landmarks in previously unseen images. The results were stored in a PostgreSQL database and used by the PDS to augment the existing web-based Imaging Atlas search interface (more details below).

Results: We evaluated the performance of the landmark classifier using cross-validation on the labeled training examples. We applied a confidence threshold to restrict the classifier's output to the 25% most confident predictions. The Gaussian Naïve Bayes classifier was the best-performing classifier, with 99% accuracy. It also strongly out-performed a simple baseline that classifies any landmark using the majority landmark class observed in the training set (see Figure 2).

This assessment provides the most realistic estimate of how the classifier performs operationally. An abstaining classifier is vital to the full-scale deployment of the system, because within the full set of all HiRISE images, many landmarks are found that fall into none of the currently identified categories.

We applied the trained classifier to the full set of HiRISE browse images, and it identified several new matching landmarks. Figure 3 shows examples of new landmarks found by the classifier for each class. The crater, dune, and dark slope streak landmarks are accurately classified. The impact ejecta example may instead be a polar "spider" feature caused by a gas jet depositing dark material on top of frosted terrain. A closer look at the full image and its imaging conditions would be required to differentiate the two. Nevertheless, the classifier can point searchers to relevant images of interest.

Public Deployment: We stored all HiRISE landmarks detected by the system in a PostgreSQL database. The PDS Planetary Image Atlas[1] added a new search facet (filter) that allows image searches to be restricted to those images containing a particular landmark type at least 90% confidence. Note that the lack of an indicated landmark does not mean that the image cannot contain the landmark, but that it was not detected with sufficiently high confidence. Therefore, positive search results for a given landmark type provide high reliability that the landmark is present, but negative results do not preclude the landmark's presence.

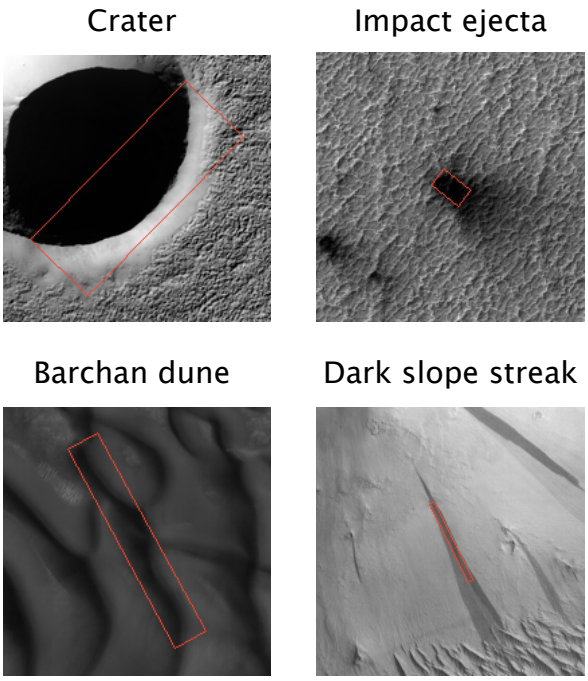

Figure 3. Examples of new landmarks found in HiRISE images by the trained classifier.

Acknowledgements: This work was supported by the NASA Advanced Multi-Mission Operations System Technology Program. We thank the Planetary Data System (PDS) for providing the image data and support for integration of the landmark database. This work was carried out in part at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

References: [1] Wagstaff, K.L., et al. (2012) *ACM TIST 3,* Article 49, *90.* [2] Vedaldi, A. and Fulkerson, B. (2010) *Int'l Conference on Multimedia*. [3] Wagstaff, K.L., et al. (2014) *8th Int'l Conference on Mars*.

[1] http://pds-imaging.jpl.nasa.gov/search/