FINDING SPACE SCIENCE IMAGES IN LARGE DATABASES BY SIMILAR APPEARANCE.

M. C. Deans[1], C. Meyer, P. Lee[1,3,4] and R. Beyer[1,3]. [1]NASA Ames Research Center, Moffett Field, CA. [2]ETH, Zurich, Switzerland. [3]SETI Institute, Mountain View, CA. [4]Mars Institute, Mountain View, CA.

Overview: Planetary missions generate large volumes of data. With the MER rovers still functioning on Mars, PDS contains over 7200 released images from the Microscopic Imagers alone. With MGS ceasing operations after a decade in orbit around Mars, the MOC catalog contains over 240,000 images. Earth observing instruments such as ASTER and MODIS have generated image catalogs into the millions.

In flight operations or on servers such as Planetary Data Systems (PDS) these data products are only searchable by keys such as the Sol, spacecraft clock, or rover motion counter index, with little connection to the semantic content of the images. During mission science operations the science team typically pores over each and every data product returned, which may not require more sophisticated organization and search tools. However, for analyzing existing image databases with thousands or millions of images, manual searching, matching, or classification is intractable.

Approach: We have prototyped a method for matching images based on the visual appearance of images. Initial "proof-of-concept" application to a few different databases of space science data shows the approach to be a promising tool for planetary science.

For every image in a database, a series of filters compute the image response to localized frequencies and orientations. Filter responses are turned into a low dimensional descriptor vector, generating a 37 dimensional fingerprint. For images such as the MER MI, this represents a compression ratio of 99.9965% (the 37 dimensional fingerprint is approximately 0.0035% the size of the original image). At query time, fingerprints are quickly matched to find images with similar appearance. Image databases containing several thousand images are preprocessed offline in a matter of hours. Image matches from the database are found in a matter of seconds[1].

Imagers such as the Mars Orbital Camera—Narrow Angle (MOC-NA) aquire "strips" of image footprint. For consistency in processing images as square tiles, we simply decimate the image strip into a series of square tiles and treat the tiles as individual images for the purposes of image processing, database creation, and search. A small amount of overhead tracks the source image for the tiles to reconcile at search time.

We have demonstrated this image matching technique using three sources of data. The first database consists of 7200 images from the MER Microscopic Imager. The second database consists of 3500 images from the MOC-NA, which were cropped into 1024×1024 sub-images. The third database consists of 7500 scanned archival photos from the Apollo Metric Camera. Example query results from all three data sources are shown below in Figures 1-5.

We have also carried out user tests to evaluate matching performance by hand labeling results. User tests verify approximately 20% false positive rate for the top 14 results for MOC NA and MER MI data. This means typically 10 to 12 results out of 14 match the query image sufficiently. However, the cost of false positive matches should not be considered high. Narrowing down the search from the entire MOC catalog down to even 100 or a few hundred images is an enormous advantage. This represents a powerful search tool for databases of thousands of images where the a priori match probability for an image might be less than 1%.

Qualitatively, correct matches can also be confirmed by verifying MI images taken in the same z-stack, or MOC image tiles taken from the same image strip. False negatives are difficult to quantify as it would mean finding matches in the database of thousands of images that the algorithm did not detect.

Utility: The method provides the capability to search for data products by appearance rather than relying on metadata such as spacecraft ephemeris or other telemetry, target or feature names, acquisition times, etc. This means that features from different parts of Mars, imaged at very different times, and with no naming context to match, can be found by discovering appearance similarities automatically. This is especially critical for post-hoc analysis, where relevant characteristics or similarities in the data are discovered after metadata tagging schemes are in place to track those characteristics manually or automatically.

Many investigations have focused on the detection of specific features[2,3], such as craters, with good success. However, the fingerprints computed here are not specific to any terrain or feature type. The method can discover similarly cratered terrain within the same framework as that which finds dunes, gullies, etc. This same framework can in fact be used for microscopic imager data and orbial remote sensing data without modification, finding similarities between rocks or fines with similar appearance but taken in different places at different times.

The ability to quickly search large databases for similar images represents a tool that can help find haystacks, not needles. However, this can be a powerful tool when more detailed feature analysis has a high

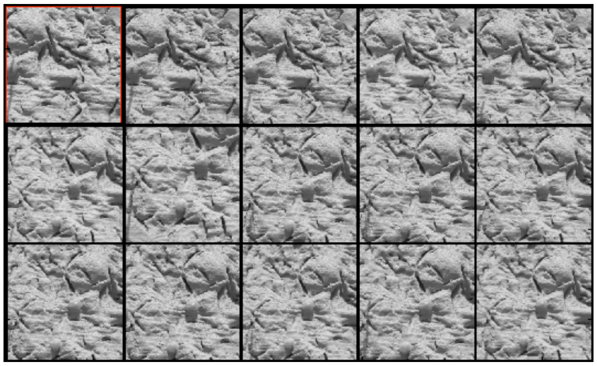

Figure 1. Example MER MI matching Result.

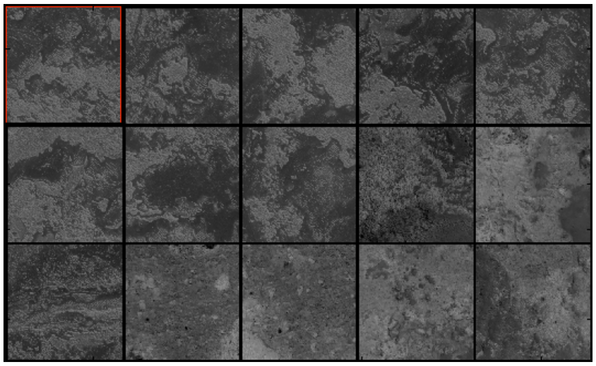

Figure 2. Example MOC-NA matching result.

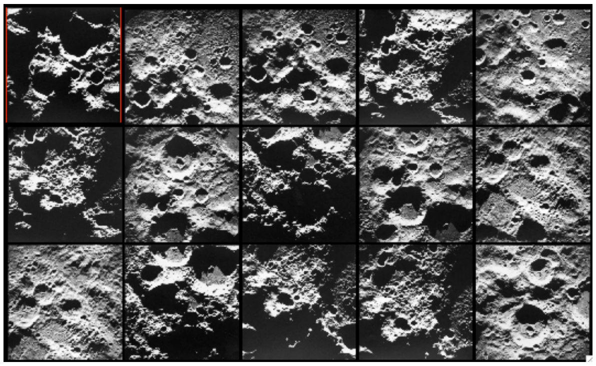

Figure 3. Example Apollo matching result.

computational cost. For example, a broader search for relevant types of images would preclude executing a computationally complex feature extraction or quantification algorithm on images where there are no relevant features.

Future Plans: It is our intention to continue to develop and improve on the underlying image matching algorithms, implement search capabilities with several real world large scale datasets, and quantify search performance and efficacy with statistically significant results. We also intend to investigate matching across instruments, where for example a MOC-NA image is used as a query to find HiRISE data (or vise versa) with the requisite rescaling or other conversions, as well as investigating wether similarities can be automatically discovered between planetary data and terrestrial analogs, for example matching Landsat images to the MOC-NA database.

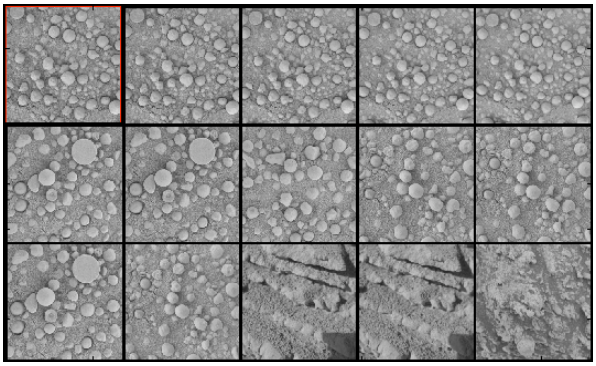

Figure 4. Example MER MI matching result.

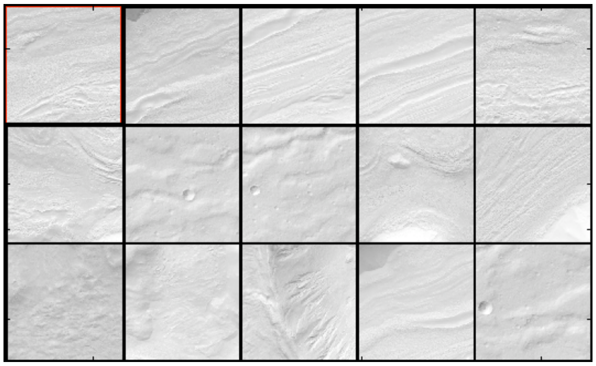

Figure 5. Example MOC-NA matching result.

The underlying method relies on a parameterized family of filters. The filters used in our proof of concept system may be improved upon in several ways, including the use of other parametric filter functions and tuning the parameters used for each family. With a reasonably sized training corpus we can also employ machine learning methods to optimize over the filter classes, number of filters, and the parameters used.

Our proof of concept system has been demonstrated on a restricted set of MER MI and MOC-NA images. With MGS recently ending operations, we can now work to catalog the entire MOC-NA as well as MOC-WA datasets, and begin making powerful content based image search tools available to the broader science community and broader public.

References: [1] Meyer, C. *Master's Thesis, EPFL (2006).* [2] Stepinski, T.F., *9th Mars Crater Consortium, (2006).* [3] B. D. Bue, et al., *LPSC (2005).*

Acknowledgements: This work was supported in part by the NASA Astrobiology Science and Technology for Exploring the Planets (ASTEP) Program.