

Saliency-Based CBIR System for Exploring Lunar Surface Imagery

Prof. Kien A. Hua, University of Central Florida

Kien A. Hua received the BS degree in computer science and the MS and PhD degrees in electrical engineering, all from the University of Illinois at Urbana-Champaign. He is currently a professor in the School of Electrical Engineering and Computer Science, and is the Director of the Data Systems Lab at the University of Central Florida. His research interests include image and video retrieval, medical imaging, network and wireless communications, sensor computing, location-based services, and intelligent transportation systems. Dr. Hua has published widely, including several papers recognized as best/top papers at various international conferences. He has served as a conference chair, vice-chair, associate chair, demo chair, and program committee member for numerous conferences, and on the editorial board of the IEEE Transactions on Knowledge and Data Engineering, Journal of Multimedia Tools and Applications, and the International Journal of Advanced Information Technology. Dr. Hua is a Fellow of IEEE.

Dr. Gholam Ali Shaykhian, NASA

Ali Shaykhian (Ali.Shaykhian@nasa.gov) is an engineer with National Aeronautics and Space Administration (NASA), Kennedy Space Center (KSC), Information Technology (IT) Directorate. Ali has earned his Ph.D. in Operations Research from Florida Institute of Technology (FIT). He has received a Master of Science (M.S.) degree in Computer Systems from University of Central Florida and a second M.S. degree in Operations Research from the same university. His research interests include knowledge management, data mining, object-oriented methodologies, design patterns, software safety, genetic and optimization algorithms and data mining. Dr. Shaykhian is a professional member of the American Society for Engineering Education (ASEE), serving as the past Program Chair for the Minorities in Engineering Division.

Mr. Robert J Beil, NASA Engineering and Safety Center

Mr Beil currently serves as a Systems Engineer for the NASA Engineering & Safety Centers (NESC) Systems Engineering Office (SEO). Mr. Beil was the requirements manager and systems engineer for the developmental, full scale Max Launch Abort System project. He leads an NASA, agency level data mining and trending working group. He worked for many years as the Orbiter Main Propulsion System (MPS) lead engineer at Kennedy Space Center. He earned a Bachelor of Science in Mechanical Engineering from Old Dominion University (Norfolk, VA) and a Master of Science in Industrial Engineering from the University of Central Florida.

KUTALMIS AKPINAR, School of Electrical Engineering and Computer Science, University of Central Florida

Kutalmis Akpinar is a Ph.D. student at the University of Central Florida, Orlando. He received his B.Sc. degree in Electrical and Electronics Engineering from Bilkent University, Turkey in 2009 and his M.S. in Electrical and Electronics Engineering from the Middle East Technical University, Turkey in 2012. From 2009 to 2012, he was also a researcher in the Informatics and Information Security Research Center, Turkey. His past experience includes human activity recognition features and surveillance systems. He is currently a member of the Database System Group (DSG) at UCF and is working on new techniques for image retrieval in image database systems.

Kyle A. Martin, University of Central Florida

Saliency-Based CBIR System for Exploring Lunar Surface Imagery

Kien A. Hua, Ph.D.¹

Gholam Ali Shaykhian, Ph.D.²

Robert J Beil²

Kutalmis Akpinar¹

Kyle Martin¹

¹Department of Computer Science, University of Central Florida

²National Aeronautics and Space Administration (NASA)

Abstract

Recent NASA missions like the Lunar Reconnaissance Orbiter (LRO) have produced vast archives of surface imagery that must be analyzed to locate landmarks with distinctive visual features, like craters, which provide important information about geologic history and potential mineral resources. Content Based Image Retrieval allows large archives of images to be efficiently queried based on visual content by indexing multidimensional feature vectors extracted from the images. Unlike images of particular objects or scenes traditionally retrieved using CBIR, surface images are not focused on any particular landmark so they must be pre-processed to identify Regions of Interest (ROI) to be indexed for retrieval. Previous work identified ROIs using manual annotation and expensive detection algorithms for specific types of landmarks, such as Crater Detection Algorithms (CDA). In contrast, this work utilizes a general-purpose saliency-based landmark detection algorithm for identifying ROIs which are then indexed for retrieval using feature vectors extracted from the ROI images. We evaluate the retrieval performance of several feature vectors and assess the saliency-based landmark detection performance in comparison to a comprehensive crater database created using manual annotation and a CDA. Experimental results demonstrate the advantages of the general-purpose saliency-based CBIR system for exploring lunar surface imagery.

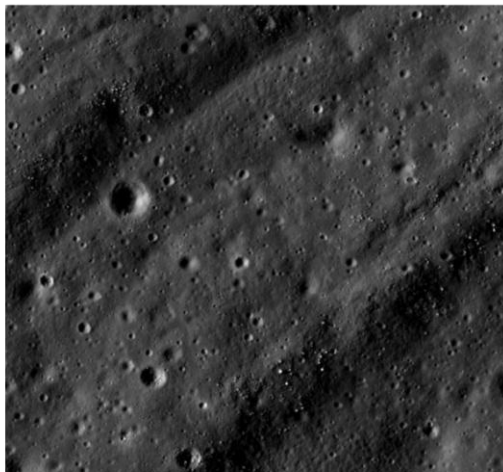
1 Introduction

Outer space contains a vast amount of resources that offer virtually unlimited wealth to the people that can access and use them for commercial purposes. The Moon and asteroid belt have a variety of resources varying from iron-nickel to silicates and possibly subterranean water ice [1]. Planning landing sites for future missions to the Moon and beyond will require balancing a number of competing objectives, such as available water ice, proximity to interesting surface features, hazard to landing craft, rovers, habitats, etc. Image data mining of extraterrestrial data sets will help automate the discovery of geological formations important for missions to the Moon, Mars, and other planetary bodies.

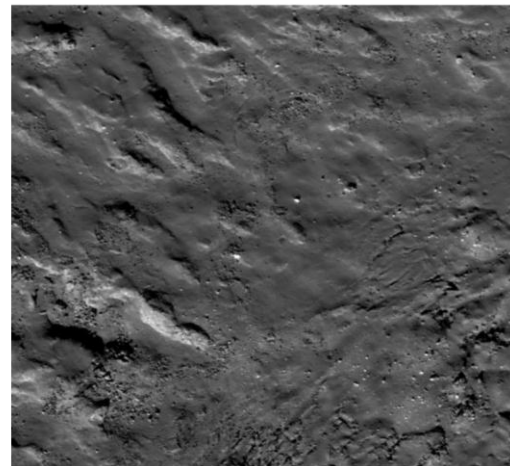
Planets and other bodies in the solar system share a number of visual features that provide important information on their formation and geology. Impact craters are distinct circular

depressions resulting from the hypervelocity impact of an asteroid or comet with a planetary surface [1], [2]. They are among the most studied geomorphological features in the Solar System because they yield information about the past and present geological chronology [3]. There are also a number of other visually distinct surface features, such as basaltic flows and fractures, which yield important contextual information for accurately determining the geologic properties of craters.

NASA is charting a bold new course into the cosmos, a journey that will take people back to the Moon, and eventually to Mars and beyond. In addition to technological challenges on earth, it must fully account for meteoroids and their impacts. To this effort, in June 2009, NASA launched the Lunar Reconnaissance Orbiter (LRO) to the moon. LRO is equipped with seven onboard instruments, which together have delivered more than 192 terabytes of data, images and maps (Figure 1) -- the equivalent of nearly 41,000 typical DVDs. An objective of this study is to explore novel information technology in order to maximize the value of such large data sets.



These step-like features are located on the flanks of a shield volcano! Down-slope to the lower right, LROC NAC image M166406436R, image width is 728 m [NASA/GSFC/Arizona State University (ASU)].



Contact between channelized impact melt and fractured melt on the floor of Thales crater. Crater floor is at lower right, LROC NAC image M170218508L, image width is 700 m [NASA/GSFC/ASU].

Figure 1: Lunar crater images from LROC data set

2 Informatics-based Approach to Planetary Exploration

Data from past missions and many other resources can be used to discover visual features that characterize particular geological formations such as craters. For instance, feature vectors can be extracted from LROC imagery to build a large metadatabase (i.e., a multidimensional feature space). By clustering these feature vectors (i.e., points with coordinates) in the multidimensional feature space, the corresponding images are classified into categories (Figure 2). Known properties of craters and other surface features can then be used to annotate feature clusters and assign geological properties to the categories.

Content Based Image Retrieval (CBIR) utilizes the metadatabase to search for images that are visually similar to a query image, such as an image of a newly discovered crater (Figure 3). It

allows the use of images of lunar craters with known geological properties to retrieve images of visually similar craters in the LROC data set. Annotations in the metadatabase may also be used as training sets for machine learning algorithms mining the large data set for unique features. This mechanism provides a novel means to remotely study lunar craters and plan future science and resource prospecting missions. It may also reveal implicit information, otherwise hidden in the very large LRO data sets.

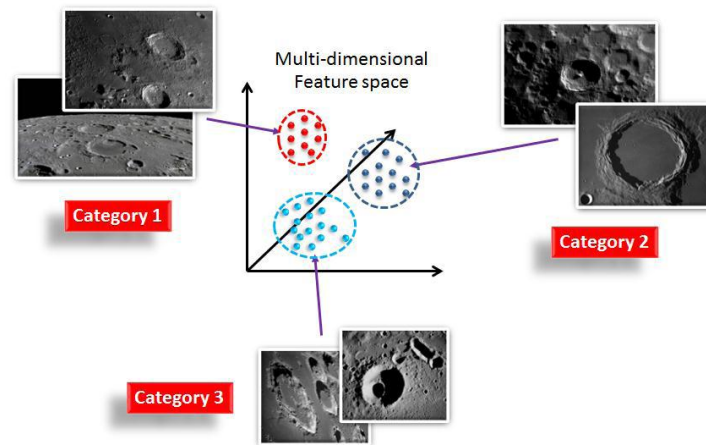


Figure 2: Categorization of lunar craters based on their visual features.

Unlike images of particular objects or scenes traditionally retrieved using CBIR, surface images are not focused on any particular landmark so they must be pre-processed to identify regions of interest (ROI) to be indexed for retrieval. Previous work identified ROIs using manual annotation and expensive detection algorithms for specific types of landmarks, such as Crater Detection Algorithms (CDAs). In contrast, this work utilizes a general-purpose saliency-based landmark detection algorithm for identifying ROIs that are then indexed for retrieval using feature vectors extracted from the ROI images. In this study, a general-purpose definition of landmark is provided and saliency-based object detection techniques, which are popular in computer vision research, are adopted to solve the landmark detection problem. Landmarks are extracted as ROIs and indexed into a CBIR system to provide a fast search capability.

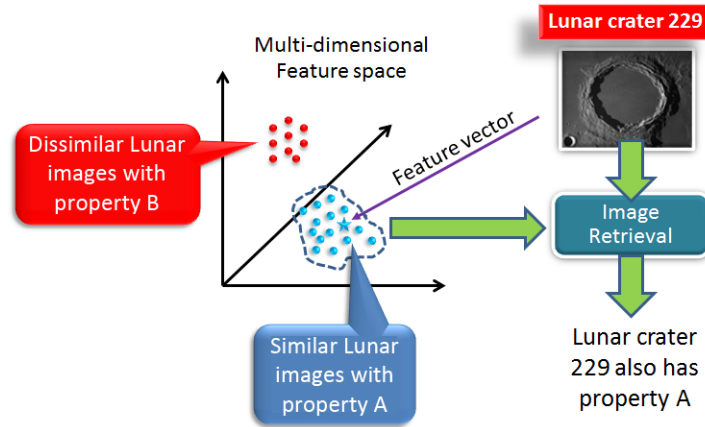


Figure 3: Investigate lunar craters by examining the geological properties of craters with similar appearance

3 Related Work

A number of geo-spatial image exploration methods have been proposed for varying purposes and planet types. As previously noted, impact craters are the most studied geomorphological features, and numerous CDAs have been published. They can be classified into unsupervised [4], [5] methods that use traditional image analysis techniques, and supervised methods [6], [7] that use trained classifiers accompanied with annotations. While early approaches use exhaustive search techniques [4], recent studies focus on faster techniques for detection of sub-kilometer sized craters due to the computational bottleneck to process mass amount of high-resolution surface image [6], [7]. Although sub-kilometer crater detection is not mature and fast enough to create a crater database, there exists some crater databases for larger lunar and Martian craters [8]. From those databases, it can be possible to query crater images through features such as radius and depth; however, crater databases contain just few useful features per crater, vastly limited compared to the corresponding visual features.

There are a small number of Content Based Image Retrieval (CBIR) systems proposed for geo-spatial image exploration. One of the early studies [9] proposes extraction of a feature vector per image segment and comparison of search segments to each other segment in the database. This can be considered as the most basic application of CBIR on planetary images. It can be applicable to retrieval of images in terms of surface texture type. Similarly, Jaroslaw et al. [10], proposes generation of a similarity map when a search region is given. Even though features are stored in database, both of the methods require search of entire set of image segments after the query. Those methods are not efficient enough to be applicable to a complete set of high-resolution lunar images. A recent publication proposes to narrow down search space into ROIs by applying CDAs. However, as in the case of crater databases, this limits the functionality to a specific type of landmark. Saliency based ROI selection proposed in this study has a wider landmark definition that is more appropriate for a variety of emerging applications.

Other than CBIR systems, Wagstaff et al. [11] uses saliency-based landmarking for surface imagery to identify and classify Martian surface features such as craters, dark slope streaks, or dust devil tracks. Wagstaff et al. demonstrates the flexibility of entropy-based saliency detection

on different landmark types. The present study applies a similar landmarking approach to CBIR systems using more challenging imagery from the LROC data set.

4 Saliency-based Landmarking

Compared to many landforms, the usual surface textures which cover wide areas on planetary surfaces are easy to explore because the extents of those regions can be determined by sampling over large distances. However, landforms which occur in small regions of interest (such as many craters) are difficult to explore due to the following reasons: 1) There are large numbers of these features with random variations and sizes distributed over a large area; and 2) Some of them cannot be explored because they occur rarely and hence researchers are unaware of their existence. Despite their rarity and small size, the distribution or existence of these landforms carry important geological information about the surface.

We have developed and implemented a method for extracting regions of interest (ROI) which have more distinctive information content than the surrounding area. Since these types of regions also stand out from the background and attract human attention, we used methods which are well studied in computer vision research for saliency detection. The saliency approach for landmark detection has also been seen in recent literature [11].

Our method for detecting landmarks depends on the information content of a region measured by the entropy of each pixel in a small neighborhood. Thresholding and binary processing are applied in order to separate regions with higher entropy from the background. This procedure is applied at multiple scales in order to cover landmarks of different sizes. For each ROI detected, visual features are extracted for indexing and search. The landmark detection procedure is depicted in Figure 5.

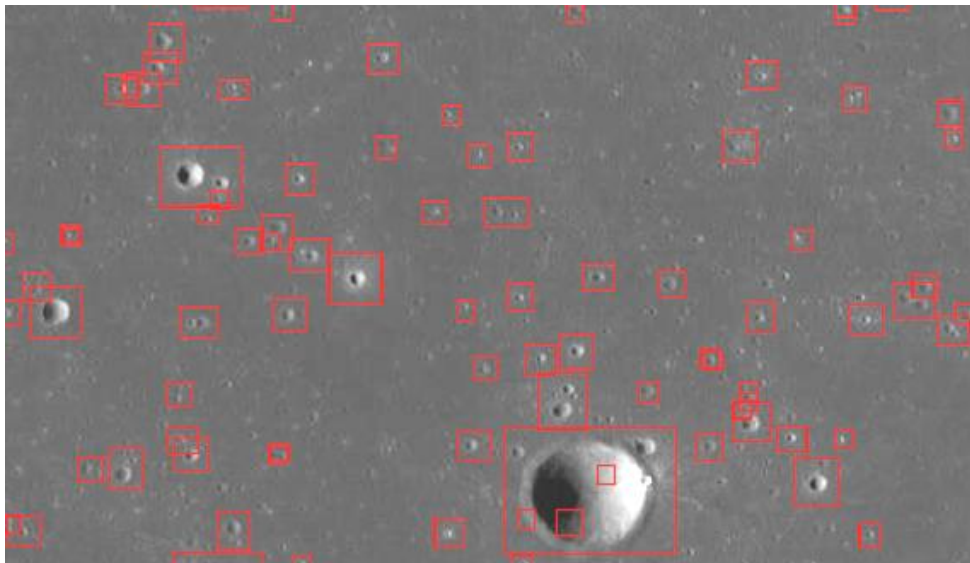


Figure 4: An example of ROIs resulting from saliency-based landform detection

Initial tests of the saliency based landform detection algorithm were performed on 100m resolution lunar surface images using an OpenCV [12] implementation of the technique presented in [11]. This implementation was the basis for the CBIR system described in the next

section. In a single scale, the algorithm was able to detect most of the craters from 5x5 to 30x30 pixels as landforms as shown in Figure 4. Larger landforms were also detected in different scales. In most cases, landforms that were or were not craters were clearly differentiable from the shape of the landform contours (note that only the bounding boxes of contours are shown in the figure). False detections tended to occur in cluttered regions where larger landmarks were searched for smaller landmarks (Figure 6). There were also cases where small landmarks on cluttered regions are skipped since they are combined with background. Future work can investigate techniques to further improve the saliency based landform detection and also explore alternative methods such as the approach based on Digital Elevation Model [13].

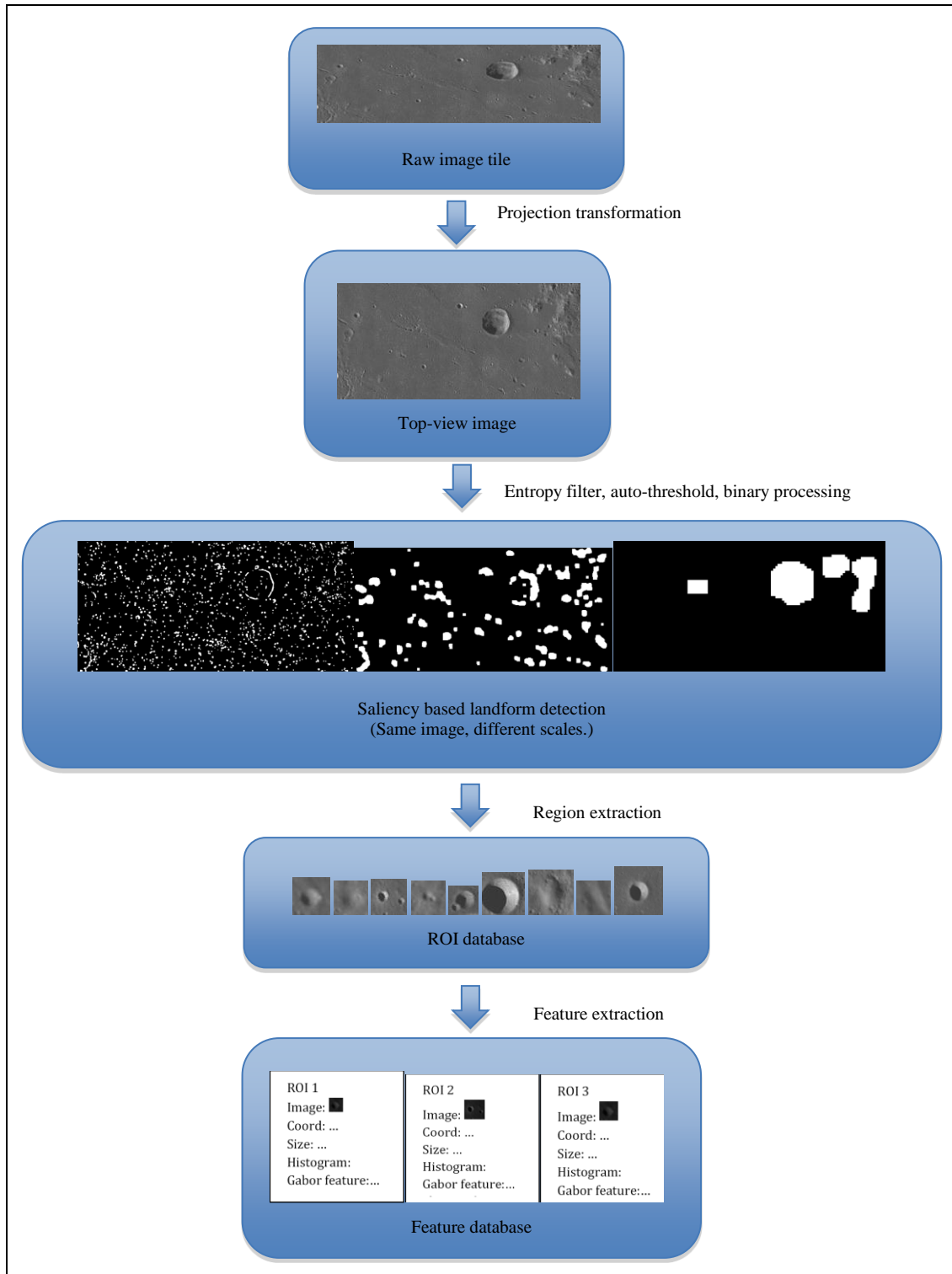


Figure 5: Salient region detection and feature extraction

5 Lunar Image Mining System

The *Lunar Image Mining* (LIM) system has been implemented using several frameworks which were integrated with the OpenCV-based image-processing pipeline to provide a user-friendly interface for studying lunar imagery. The Geospatial Data Abstraction Library (GDAL) [14] is used to efficiently access LRO imagery stored in the PDS image file format. We use the Lucene Image Retrieval (LIRE) library [15], [16] with the search engine to store and index landmark feature vectors for use during similarity search. NASA's World Wind library [17] is used to implement a 3D interface (Figure 7) for viewing and interacting with GIS (Geographic Information System) data sets. This approach is motivated by the fact that NASA scientists are already familiar with the World Wind operating environment. The current prototype allows the user to browse and search imagery to identify interesting locations and display similar locations.

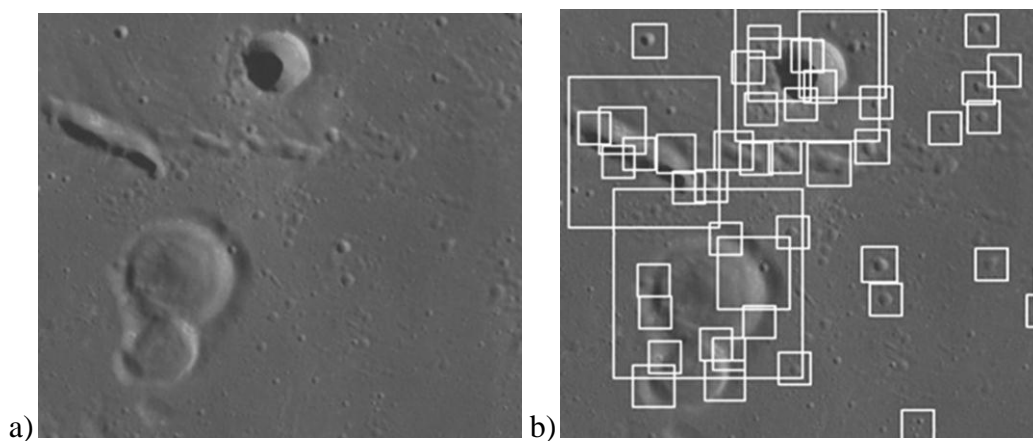


Figure 6: False detections of small landmarks on top of large landmarks. Although three main large landmarks are taken inside bounding boxes, many unnecessary small landmarks also appear inside large landmarks. a) Surface

LRO imagery was retrieved from [18]. The multi-resolution Reduced Data Records (RDR) for the LROC were downloaded as they provide preprocessed projected mosaics at relatively high resolution suitable for our study. An imagery pre-processor was developed, using GDAL and OpenCV, to construct image pyramids from the LRO imagery, suitable for display using World Wind. Each tile in the pyramid is then processed as shown in Figure 5 to detect landmark ROIs and perform feature extraction. We note that the image pyramids will also be useful in the long-term for image processing and data mining at different resolutions, and partitioning the imagery for distributed processing using Hadoop [19] and other Big Data [20] technologies.

The current LIM implementation provides command-line tools for extracting and indexing landmarks and features while similarity search is performed from a user-friendly UI as shown in Figure 7. Image retrieval queries begin with the user selecting a region with known properties such as a previously studied crater. The corresponding feature vector is then extracted from this query region, and compared with feature vectors stored in the Lucene index to find the set of feature vectors nearest to the query image in the multidimensional feature space.

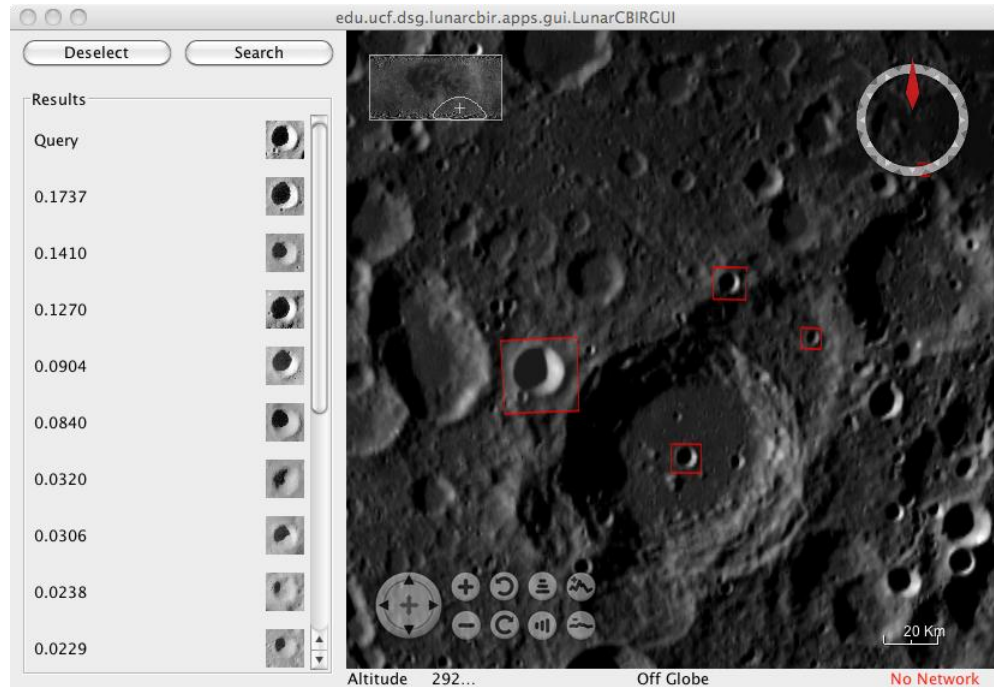
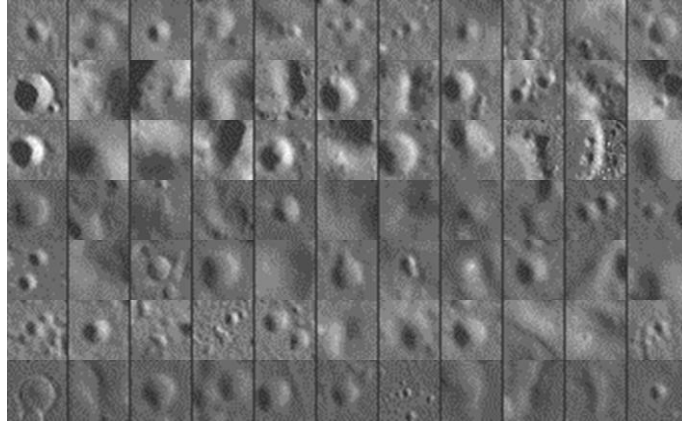


Figure 7 Screenshot of UI prototype showing similarity search results

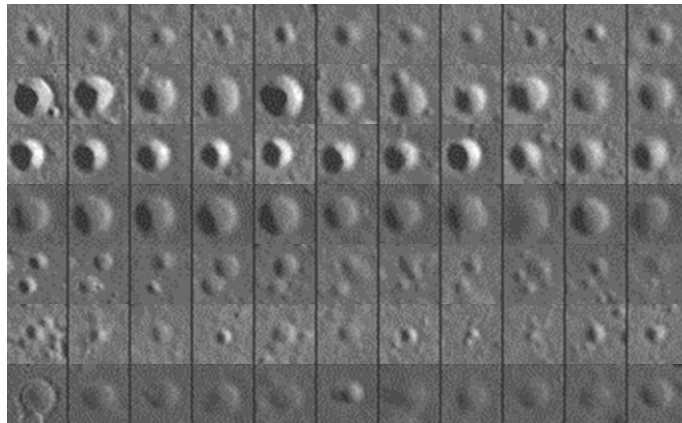
6 Software System Testing and Experiments

Three kinds of initial visual features were used in our study to facilitate software testing, namely Gabor features, image intensity, and Hu invariants. Gabor features count occurrences of special spatial frequency regions in an image. The intensity feature is implemented by normalizing image sizes and creating feature vectors directly from pixel values of the ROI image. Finally, Hu invariants are scale and rotation invariant values which represent shape of a binary image. The shape in this case is the binary mask of the region that is extracted in earlier processes of entropy filtering and thresholding (Figure 5). Although software testing indicates that the LIM prototype is operating correctly and efficiently, it is desirable to investigate additional visual features to improve the image data mining results.

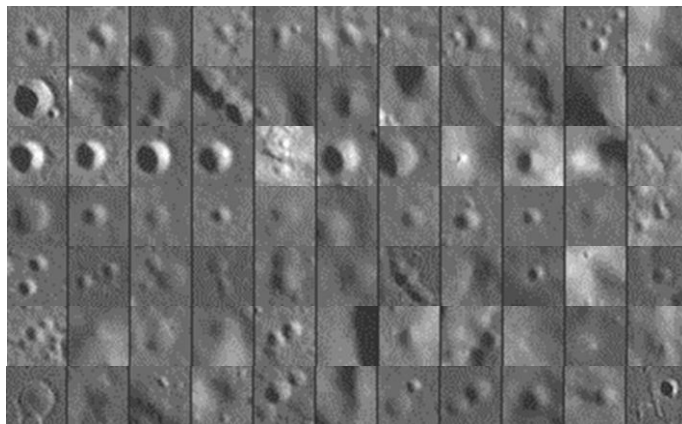
The results of several similarity searches using the three features are shown in Figure 8. These results are based on 12,905 ROI images extracted from 100m/pixel-resolution LRO Wide Angle Camera imagery in region bounded by coordinates 0.0N 0.0E and 60.0N 90.0E [18]. 1/8 of the total image area was processed. As can be seen from the results, searches that depend on Gabor feature are not consistent. This is likely due to ROIs being too small to provide enough texture content for Gabor features to be effective. Intensity feature performs best, however, compared to Hu invariant features, intensity features are sensitive to position and intensity variations and are rotationally variant. These limitations are most noticeable in query images that have direction, such as the 5th, where searching using intensity feature brings results in a single direction, while using Hu invariant features yields results invariant of direction. While the Hu invariant feature is quite successful in representing the shape of the region, it does not consider the content of the region, yielding less accurate results compared to the intensity feature.



(a)



(b)



(c)

Figure 8: Similarity search results. For each row, the leftmost image is the query image and the remaining images are the top 10 most similar. Using feature:
a)Gabor b)Intensity c)Hu invariants

6.1 Landmarking

It is important to gauge the effectiveness of the general-purpose saliency-based landmarking approach for identifying different classes of landforms. As craters are the only known lunar surface features to be extensively catalogued we utilize a recent crater database [8] to serve as a ground truth and measure how well the set of detected landmarks covers the set of known craters. For this purpose a subset of the crater database was extracted yielding 12,212 entries lying within the range of indexed LROC RDR imagery between 60.0S 0.0E and 30.0N 90.0E. Each of these craters was then compared with the 309,827 indexed ROIs to determine whether it significantly overlapped, contained, or was contained by an ROI. Covered ROIs were annotated as such for the following experiments. Landmarking coverage was found to be 40.95%: 5001 crater DB entries covered by 6190 ROIs. A summary of the different types of coverage is shown in Table 1. The large number of ROIs contained by crater DB entries is due to the maximum size of ROIs being much smaller than for the crater DB entries, i.e. the ROIs cover portions of large craters.

	DB Entries	ROI
Overlapped	1201	1356
Containing	791	856
Contained	3676	3978

Table 1: Crater DB Coverage

It is important to note that a significant number of ROIs were of craters that were not covered by crater database entries. This would imply that there is in fact a significant false negative rate and actual coverage would be greater than 40%, however quantifying the false negative rate will be left to future work.

6.2 Precision and Recall

To measure the CBIR performance of the three visual features two cross validation experiments were performed to measure precision and recall based on search queries using previously annotated ROIs covered by craters from a crater database [8] and a set of manually annotated craters. In the first experiment, a 10-fold cross validation was performed using 1000 randomly selected crater DB annotated ROIs and 8100 non-annotated ROIs, for a final precision of 10% using 100 queries per validation. The second experiment was similar, except 880 randomly selected manually annotated crater ROIs and 794 manually annotated non-crater ROIs were used with 792 queries per validation. In both experiments the interpolated precision-recall curve and mean average precision (MAP) were calculated.

The results of the first experiment are shown in Figure 9 and Figure 10. Relevant ROIs were randomly selected from the set of annotated ROIs that were not contained by a crater DB entry, i.e. they overlap or contain the entries (to avoid ROIs of portions of craters). There is a rapid decline in maximum precision for all features, however the Intensity features seems to perform the best, followed by the Hu features which fare slightly better than Gabor features. For all features, the mean precision does not significantly vary from the final precision of the validation data set. This would indicate that on average none of the features are discriminative enough to effectively distinguish between crater and non-crater ROIs, however the quality of the crater DB

annotations is also to blame. Many annotated ROIs were parts of craters, such as crater rims, sloping walls, etc., which would negatively effect the resulting precision. The extremely low mean precision with extremely high max precision for the Intensity feature may also be explained by the low quality of the annotations, since the Intensity feature is much more selective than the Hu and Gabor features.

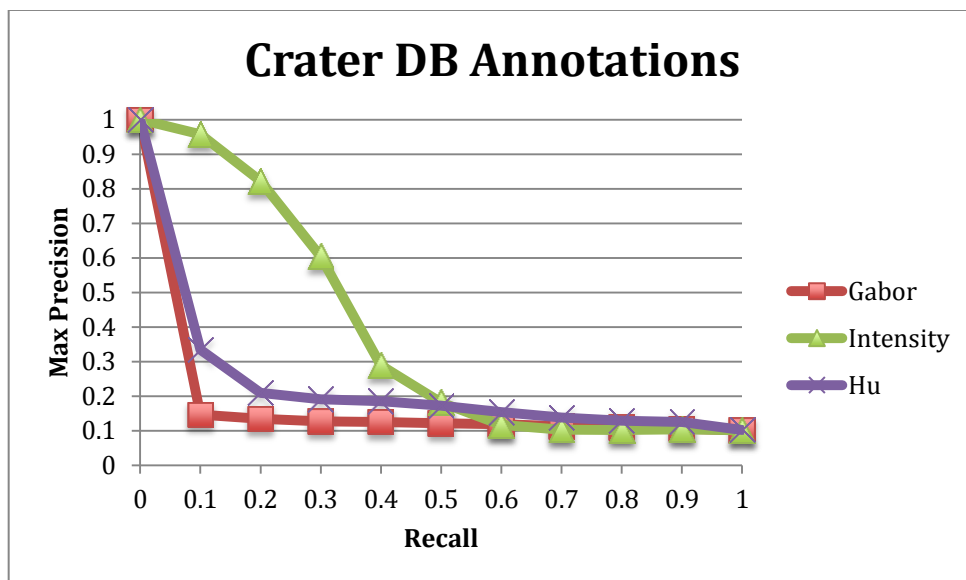


Figure 9: Maximum precision vs. recall using crater DB annotations

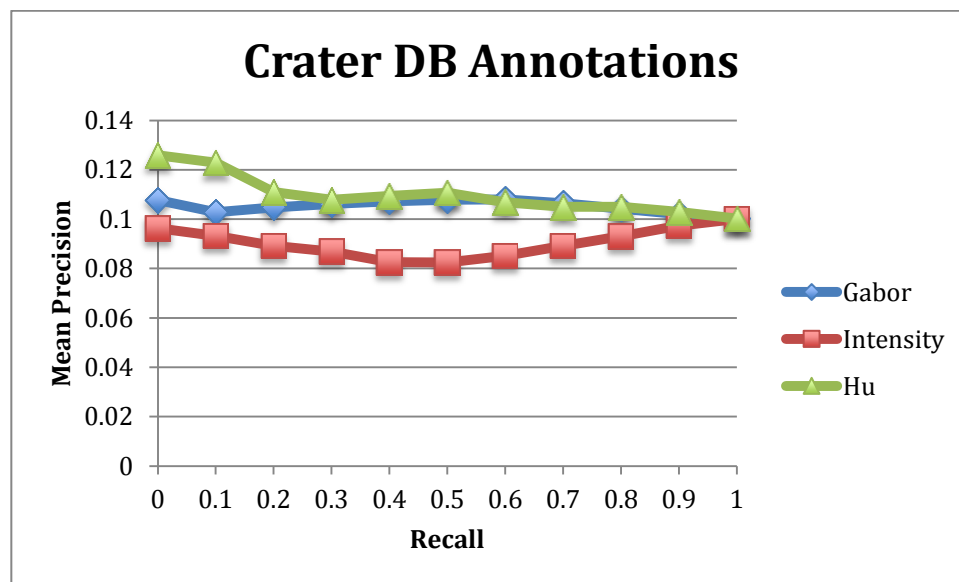


Figure 10: Mean precision vs. recall using crater DB annotations

To better judge the quality of the three visual features a portion of the indexed ROIs were manually annotated to only label those ROIs which contained one or more significant craters as relevant, all others, such as crater rims, slopes, and other landforms were labeled as irrelevant. Note that these annotations were based purely on subjective visual inspection and not on

containment of known craters from the crater database. The results of the second experiment using these manual annotations are shown in Figure 11 and Figure 12. As before, Intensity features perform significantly better than Hu and Gabor features. Compared to using crater DB annotations, using manually annotated crater ROIs results in a slightly slower decline in maximum precision for the Gabor and Hu features, with a noticeably greater loss in precision with Intensity features. Likewise, the mean precision varies much more significantly from the final precision, but still does not achieve greater than about 17%. Clearly more discriminative visual features will need to be found to improve the average search performance.

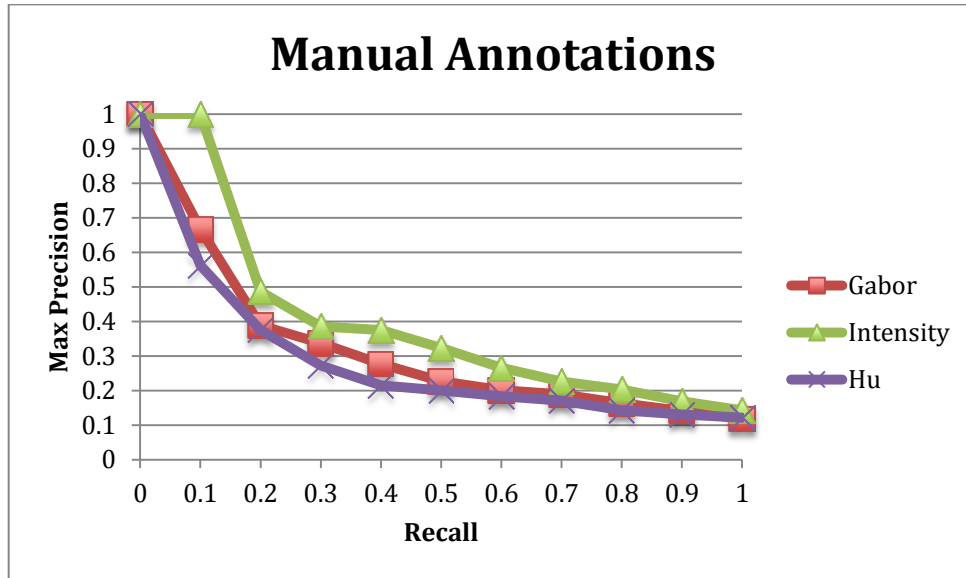


Figure 11: Maximum precision vs. recall using manual annotations

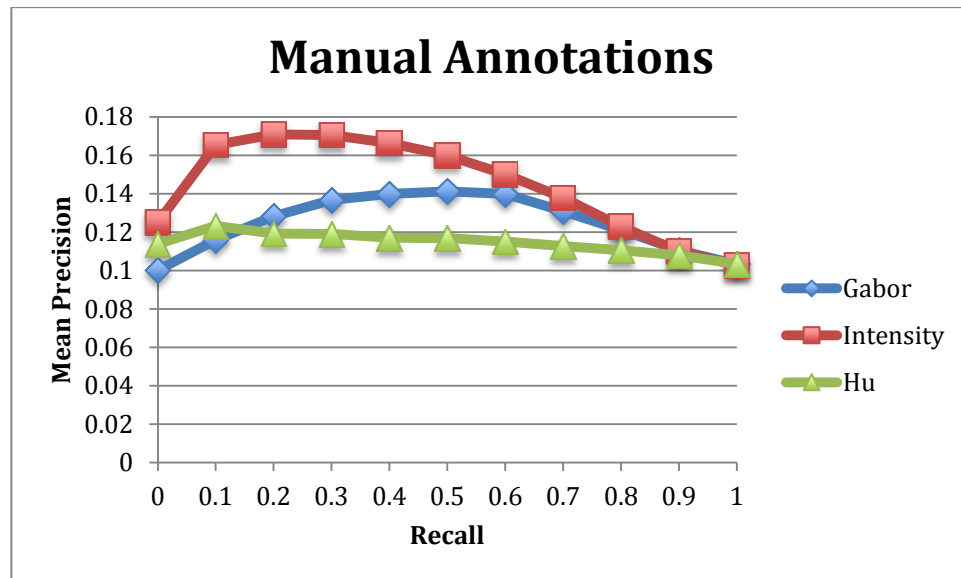


Figure 12: Mean precision vs. recall using manual annotations

The mean average precision for all features in both experiments is summarized in Table 2.

MAP	Gabor	Hu	Intensity
Crater	0.1078	0.1157	0.1168
Manual	0.1381	0.1251	0.1647

Table 2: Mean Average Precision

6.3 Discussion

Based on the previous results it is clear that more discriminative ROI feature vectors are required. However, without an accurate ground truth it will be difficult to evaluate the performance of additional visual features and the crater database used [8] is the most comprehensive small-crater database known. Detailed manual annotation of high-resolution images for a relatively large area will likely be required to best compare the performance of different visual features for retrieving different landmarks. Thankfully the LIM system currently provides an easy to use interface designed for this task. There is also significant room for improvement in the landmark detection algorithm as well. Although approximately 40% of known craters were detected using only saliency-based landmarking without employing specialized CDAs, it is likely that a more refined landmarking algorithm that took into account additional visual features would cover more known and unknown craters as well as other types of landmarks.

Many other feature types, such as Haar-like features or gradient histograms, could possibly be added to the system. However, the usefulness of a feature depends ultimately on the intention of search. If the user is searching for a landmark with a particular shape, a feature like the Hu invariant may be best suited. On the other hand, if a user is seeking landforms with the same geographic orientation, features that are sensitive to directionality (like the intensity feature) will yield better results. Future work can study interactive image retrieval techniques that use machine learning to refine query parameters and determine the relevance of available features to a particular query.

In these initial results, a very large number of ROIs were extracted from a relatively small sample of lunar imagery. As the system is scaled up and higher resolution imagery is processed there will likely be many millions of indexed ROIs. Searching through such a large data set will not be very efficient. To address this issue we have begun investigating clustering techniques, such as k -means clustering [21], [22] and Locality Sensitive Hashing [23] to identify clusters of neighboring ROIs within the multidimensional feature space and assign representative ROIs for each cluster. Searching through this much smaller set of representative ROIs will be much more efficient. These clusters will also form the basis for the image mining and annotation features described earlier.

7 Future Work

The results accomplished in this six-month exploratory research have demonstrated the validity of the informatics-based approach to planetary discovery. Continued research and development is needed to significantly improve the LIM system:

1. Currently, the similarity search is based on a sequential scan of the feature vectors. More effective multidimensional access methods are necessary to support efficient search on much larger data sets.
2. The initial feature vectors used in this study to facilitate software testing have limited performance. The LIM prototype, nevertheless, provides a test bed for exploring additional feature vectors in order to better classify landform types and improve the image data mining results.
3. No single set of features is universally effective for different search requirements. It is desirable to employ machine learning methods to develop intelligent query optimization techniques to select the best subset of features for a given query. For instance, classifiers can be trained to recognize different categories of queries and select the best features for each category.
4. Search techniques based on user relevance feedback can also be considered for the LIM environment. In each search iteration the user marks the relevant images retrieved by the system. These images can then be used to refine the search criteria for the next round of user relevance feedback. This iterative process can be repeated until the user is satisfied with the retrieval result.
5. Image segmentation technique can be developed to uncover and characterize meaningful regions of planets. This requires novel landmark-based spatial clustering techniques.
6. The LIM system can be extended to support elevation data and hyperspectral imagery.

Ultimately, it is highly desirable to develop a more complete IT environment, as illustrated in Figure 13, which will enable NASA scientists to effectively study various very large image data sets using image mining facilities. This system has four major subsystems:

1. **Meta Database Management System:** Scientists will be able to use the automatic feature extraction mechanism to build the multidimensional metadatabase and maintain the image database for the LROC and LOLA data sets.
2. **Image Retrieval System:** Scientists will be able to use an image with known geologic properties to retrieve similar images to study their properties.
3. **Image Mining System:** Scientists will be able to use machine learning and data mining techniques such as image classifiers and mining algorithms to investigate different classes of extraterrestrial objects (i.e., lunar craters).
4. **User Interface:** The user interface provides a unified view of the above systems to ensure user friendliness.

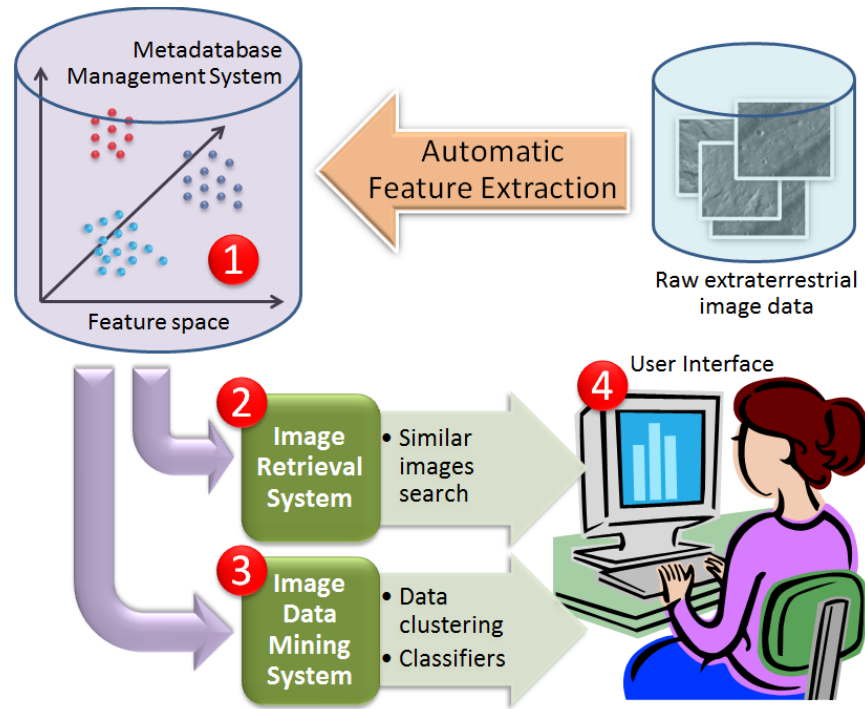


Figure 13: An advanced IT Environment for studying very large extraterrestrial data sets

The implementation of this system should be highly modularized to provide extensibility. Users will be able to install additional feature extractors into the image-processing pipeline. Additional machine learning algorithms may also be installed into the image mining system. The user interface, metadatabase, image-processors, and image mining components will be highly decoupled to allow for replacement or use with similar components. Loose coupling will also allow the system to be easily scaled over a cluster or cloud-computing provider for “big data” operations.

8 Conclusion

There is a clear need for better tools to tackle the overwhelming amount of imagery data collected by NASA missions in the past, present, and future. This study has demonstrated the potential of an image data mining system based on a content-based image retrieval framework. Further work as discussed in this report will provide a powerful image data mining tool that can help NASA and private-sector scientists and engineers analyze surface imagery for mission planning, resource prospecting, and scientific research.

References

- [1] R. P. Mueller and P. J. van Susante, "A Review of Extra-Terrestrial Mining Concepts," Apr. 2012.
- [2] S. Liu, W. Ding, J. P. Cohen, D. Simovici, and T. Stepinski, "Bernoulli trials based feature selection for crater detection," in *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, New York, NY, USA, 2011, pp. 461–464.
- [3] K. L. Tanaka, "The stratigraphy of Mars," *J. Geophys. Res. Solid Earth*, vol. 91, no. B13, pp. E139–E158, 1986.
- [4] Y. Sawabe, T. Matsunaga, and S. Rokugawa, "Automated detection and classification of lunar craters using multiple approaches," *Adv. Space Res.*, vol. 37, no. 1, pp. 21–27, 2006.
- [5] G. Salamuniccar and S. Loncaric, "Method for Crater Detection From Martian Digital Topography Data Using Gradient Value/Orientation, Morphometry, Vote Analysis, Slip Tuning, and Calibration," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 5, pp. 2317–2329, 2010.
- [6] E. R. Urbach and T. F. Stepinski, "Automatic detection of sub-km craters in high resolution planetary images," *Planet. Space Sci.*, vol. 57, no. 7, pp. 880–887, Jun. 2009.
- [7] W. Ding, T. F. Stepinski, L. Bandeira, R. Vilalta, Y. Wu, Z. Lu, and T. Cao, "Automatic Detection of Craters in Planetary Images: An Embedded Framework Using Feature Selection and Boosting," in *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, New York, NY, USA, 2010, pp. 749–758.
- [8] G. Salamuniccar, S. Lončarić, and E. Mazarico, "LU60645GT and MA132843GT catalogues of Lunar and Martian impact craters developed using a Crater Shape-based interpolation crater detection algorithm for topography data," *Planet. Space Sci.*, vol. 60, no. 1, pp. 236–247, Jan. 2012.
- [9] C. Meyer and M. Deans, "Content based retrieval of images for planetary exploration," in *IEEE/RSJ International Conference on Intelligent Robots and Systems, 2007. IROS 2007*, 2007, pp. 1377–1382.
- [10] J. Jasiewicz and T. F. Stepinski, "Example-Based Retrieval of Alike Land-Cover Scenes From NLCD2006 Database," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 1, pp. 155–159, 2013.
- [11] K. L. Wagstaff, J. Panetta, A. Ansar, R. Greeley, M. Pendleton Hoffer, M. Bunte, and N. Schörghofer, "Dynamic Landmarking for Surface Feature Identification and Change Detection," *ACM Trans Intell Syst Technol*, vol. 3, no. 3, pp. 49:1–49:22, May 2012.
- [12] "OpenCV | OpenCV." [Online]. Available: <http://opencv.org/>.
- [13] B. D. Bue and T. F. Stepinski, "Automated classification of landforms on Mars," *Comput. Geosci.*, vol. 32, no. 5, pp. 604–614, Jun. 2006.
- [14] "GDAL: GDAL - Geospatial Data Abstraction Library." [Online]. Available: <http://www.gdal.org/>.
- [15] M. Lux and S. A. Chatzichristofis, "Lire: lucene image retrieval: an extensible java CBIR library," in *Proceedings of the 16th ACM international conference on Multimedia*, New York, NY, USA, 2008, pp. 1085–1088.
- [16] "Lire | SemanticMetadata.net." [Online]. Available: <http://www.semanticmetadata.net/lire/>.
- [17] "World Wind JAVA SDK." [Online]. Available: <http://worldwind.arc.nasa.gov/java/>.
- [18] "Lunar Orbital Data Explorer - Data Set Browser." [Online]. Available: <http://ode.rsl.wustl.edu/moon/indexDatasets.aspx>.
- [19] K. Shvachko, H. Kuang, S. Radia, and R. Chansler, "The Hadoop Distributed File System," in *2010 IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST)*, 2010, pp. 1–10.
- [20] M. A. Beyer and D. Laney, "The Importance of 'Big Data': A Definition," *Gart. Rep. June Version ID*, no. G00235055, 2012.
- [21] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A K-Means Clustering Algorithm," *J. R. Stat. Soc. Ser. C Appl. Stat.*, vol. 28, no. 1, pp. 100–108, Jan. 1979.
- [22] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognit. Lett.*, vol. 31, no. 8, pp. 651–666, Jun. 2010.
- [23] M. Slaney and M. Casey, "Locality-Sensitive Hashing for Finding Nearest Neighbors [Lecture Notes]," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 128–131, 2008.