
Data Mining System for Planetary Images

- Crater Detection and Categorization -

Rie Honda
Osamu Konishi
Ryushi Azuma
Hisashi Yokogawa
Syuta Yamanaka

Dept. Inform. Sci., Kochi University, 2-5-1, Kochi, JAPAN

HONDA@IS.KOCHI-U.AC.JP
KONISHI@IS.KOCHI-U.AC.JP
TATAZUMA@IS.KOCHI-U.AC.JP
YOKOGAWA@IS.KOCHI-U.AC.JP
YAMANAKA@IS.KOCHI-U.AC.JP

Yuichi Iijima

IIJIMA@PLANETA.SCI.ISAS.AC.JP

Institute of Space and Astronautical Science, Yoshinodai 3-1-1, Sagamihaya, Kanagawa, JAPAN

Abstract

The surfaces of solid planets are covered with a tremendous number of craters of various shape and size that reflect the geological history and rheology of the planetary surfaces. In the present study we examined crater detection and categorization system for planetary images in view of data mining from a large-scale scientific image database. Crater detection by combinational Hough transforms and generic algorithm and categorization by self-organizing mapping are investigated as component processes in detail.

1. Introduction

A large amount of remote-sensing image data has recently been obtained by satellites due to the improvement of sensors and telemetry systems. These conditions have led scientists to request the construction of a large-scale scientific image database, to extract the characteristics of these images, and to construct a secondary database without manual operation. Studies aimed at the automatic extraction of objects of interest and classification based on a machine-learning method are reported recently.(Fayyad et. al., 1996; Smyth et. al., 1996; Burl et. al., 1998, 1999)).

In this study we consider a knowledge discovery support system for the lunar multi-spectral images taken by the U.S. spacecraft Clementine (Nozette et. al., 1994), particularly focusing on crater detection and categorization problems.

In the following, section 2 provides backgrounds of

the crater analysis of planetary surfaces, section 3 describes a system overview as well as details of the method of extraction of crater candidates and their categorization using self-organizing mapping. Section 4 presents conclusions.

2. Backgrounds

It is widely known that solid planet surfaces are covered with a tremendous number of craters of various sizes and shapes (Melosh, 1988). Most of these craters have been formed as a result of meteoroid impacts. Assuming a size-frequency distribution of meteoroids and its time variations, we can estimate the relative formation age of the local area of a solid planet surface by counting and sizing craters. The craters also have a variety of shapes, including bowl-floored craters, flat-floored craters, flat-floored craters with central peaks, and degraded craters, reflecting the surface rheology and subsurface structure. Despite its importance and the tremendous amount of data available, crater analysis remains dependent on human vision and manual operation.

In 1994, the U. S. Spacecraft Clementine obtained lunar global images by means of UV-VIS sensors. These data are archived in 88 volumes of CD-ROMs and the U. S. Geological Survey has composed a mosaic images called lunar digital image model (LDIM, 15 volumes of CD-ROMs) from the 750-nm band data.

By using the LDIM, we examine the automatic extraction and categorization of craters as a prototype of an datamining system for planetary imagery. We believe that this concept will be easily extended to the other features such as volcanoes, grooves, and lineaments.

3. Method

3.1 Overview

An overview of the image-mining system for craters is illustrated in Figure 1. The system includes image database, preprocessing by filtering and binarization, circular object detection by a genetic algorithm (GA) and Hough transforms(HT), and categorization using self-organizing map (SOM). Additional leaning and clustering utilizes the feature map previously learned by SOM.

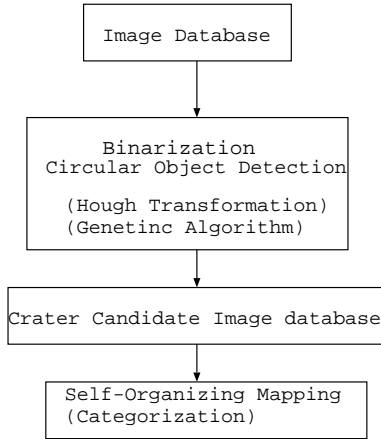


Figure 1. Overview of crater detection and categorizing support system.

The image database is composed of a Clementine LDIM consisting of a set of images whose height and width are equal to a 7 degree in latitude. Thus the image database is composed by tuple as follows: {Image name, Reticle longitudes, Reticle latitudes }. LDIM are grayscale images, most of which are taken when the spacecraft and the sun are situated in the same direction in relation to the lunar surface. Images around 60 degree in latitude are taken at a solar altitude of approximately 30 degree and are thus appropriate for terrain analysis.

Since we recognize craters based on spatial intensity changes, we preprocess the images by operating an edge detection filter (Robert filter). The images are transformed into binary data by determining the threshold for binarization statistically from the intensity histogram or interactively.

3.2 Circular object detection

After the binarization, the crater extraction is considered in terms of circular signal detection by a pattern-matching method. We examined the following two ap-

proaches for a comparison.

3.2.1 COMBINATIONAL HOUGH TRANSFORMS

The Hough transforms is a commonly used method for extracting geometrically simple parametric figures from binary images. The signals on images are projected into the parameter space of the figure. In this case, the parameters are the center of the circle and its radius. Watanabe and Shibata(1990) have reported combinational Hough Transform (CHT) that uses pyramid images and a pair of signals in a restricted region to simplify the projection to a parameter space, showing that CHT can significantly reduce computational time and improve the solution accuracy. We therefore implemented this method for crater detection.

The algorithm of crater detection based on CHT are summarized as follows.

1. The original binary image are preprocessed by using some of the following methods: isolated noise reduction, expansion and shrinking, thinning by Hilditch's algorithm, pyramid-like signal reduction.
2. The image is degraded using the $W \times W$ pixel filter matrix.
3. The image is divided into the $L \times L$ pixels blocks.
4. The radius of the target circle is set to be $r = L/4$. The following process from 5 to 8 are proceeded increasing r by 1 while $r \leq L/2$.
5. The processes of 6 and 7 are performed for all blocks.
6. Among pairs of black pixels in the block extended by 50% ($2L \times 2L$ pixels), $P_{i1}(x1, y1)$ and $P_{i2}(x2, y2)$, the pairs that satisfy $r \leq |P_{i1}P_{i2}| < 2r$ are selected as solution candidates.
7. The center of the circle (x_{ci}, y_{ci}) is calculated for each pair assuming they exist on a circle rim with radius of r .
8. The poll to the (x_{ci}, y_{ci}, r) cell in the parameter space is increased by 1.
9. The cells are sorted concerned with number of polls. If the poll is larger than 0, a circle of (x_{ci}, y_{ci}, r) is projected on the image, and the normalized poll and the matching ratio are calculated. The definition of both values are given by

$$NP = p/npp^2, \quad (1)$$

$$M = bp/npp, \quad (2)$$

where NP is the normalized poll, p is the poll, npp is the number of pixels on the rim of projected circle, M is the matching ratio, bp is the number of black pixels of the rim of projected circle. Furthermore, to exclude the false solutions caused by random noises, we introduced the internal noise ratio IM within the circle with the radius of hr , where $0 < h < 1$, and typically $h = 0.6$.

10. The cells satisfying $NP > NP_{threshold} \cap M > M_{threshold} \cap IM < IM_{threshold}$ are extracted as the solutions.

Since the radius of circle is restricted by L , we utilize the multiresolution image of the original grayscale image to detect the circle with the radius larger than $L/2$.

3.2.2 GENETIC ALGORITHM

A genetic algorithm is commonly used to obtain the single solution in the optimization problem. In order to implement this algorithm for circular object detection, we set the gene as follows:

$$\{x_i, y_i, r_i\},$$

where (x_i, y_i) is the center of the circle represented by i -th gene and r_i is the radius of the circle. The adaptation for the i -th gene, g_i , is calculated by projecting the circle of the i -th gene on the binary image and by checking the overlap:

$$g_i = n_i / N_i, \quad (3)$$

where n_i is the number of black pixels on the circle of i -th gene, and N_i is the total number of pixels on the circle.

In order to avoid random noise being incorporated into the solution, we modified g_i as follows:

$$g'_i = g_i - g_{i,r=fr_i}, \quad (4)$$

where $0 < f < 1.0$, typically $f = 0.3$.

Since it is possible to have many craters in a single image, we have included a process to erase the detected circles from the original images, and the same process is continued iteratively. The algorithm of crater detection by GA is summarized as follows.

1. The original image is degraded using $W \times W$ pixel filter matrix.
2. Initial populations of genes are generated.
3. The following process from 4 to 6 are iterated for a given number of generations.
4. The adaptations of each gene, g'_i , are calculated.
5. The genes are selected, crossed, and mutated.
6. The genes with the same attributes are unified.
7. The genes with $g'_i > g'_{threshold}$ are detected as solutions.
8. The solutions are projected as circle rims on the image. The intensity of pixels on the projected circle rims are changed to 1 (white).
9. The processes from 2 to 8 are iterated for a given number of times.

3.3 Experimental result of circle detection by CHT and GA

The above described method of crater detection, CHT and GA, are examined for a test image selected from USGS's Mars Digital Image Model. Figure 2 show the original grayscale image and the binary image. The size of image is 296×320 pixels and number of correct solutions which should be detected are eight.

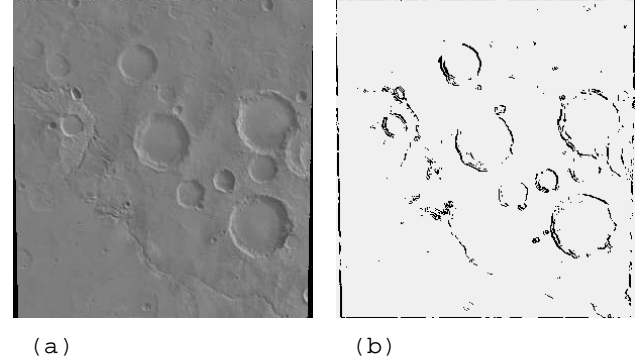


Figure 2. Test images for crater detection (mg25s017 from NASA's Mars Digital Image Model). The figure (a) is the original grayscale image and the figure (b) is the binarized image.

As described in 3.2.1, CHT has free parameters of three thresholds for NP , M , IM and the size of filter matrix size of degradation, choice of noise reduction method and signal reduction. We mainly examined the effect of noise reduction, signal reduction, and size of degradation matrix providing $NP_{threshold}=0.064$, $M_{threshold}=0.60$, $IM_{threshold}=0.04$, and $L=36$.

Table 1 summarizes the result of experiments for crater detection by CHT. The results are sorted by the decreasing number of true solutions and the increasing number of false solutions. The results suggest that CHT can detect clear craters successfully by preprocessing images adequately and optimizing parameters. We can observe that noise reduction strongly affects both numbers of true and false solutions. In particular, thinning and pyramid-like signal reduction are very effective to reduce number of false solutions and computational time without losing true solutions.

The same image is processed by crater detection algorithm based on GA. We focused on the effect of parameters to the number of iteration required until all correct solutions are obtained. We chose the selection rate of 0.5 and mutation rate of 0.2 based on the preliminary experiments.

Table 2 summarizes the experimental results on crater

Table 1. Summary of result of crater detection by CHT. The notation of I, ES and TH indicate isolated noise reduction, expansion and shrinking, and thinning, respectively. Calculation time is measured by using Ultra Sparc 167MHz.

noise reduction	signal reduction	W	no. solutions		cpu time (sec)
			true	false	
TH	Y	5	7	0	115
I+ES+TH	Y	9	7	1	40
N/A	Y	5	7	3	162
I+TH	Y	9	7	3	174
N/A	Y	7	7	4	112
N/A	Y	7	7	5	189
I+TH	Y	7	6	0	66
I	Y	5	6	1	120
I+ES+TH	Y	7	6	1	40
I+TH	Y	5	4	0	68
I+ES+TH	Y	5	4	0	41
TH	N	5	0	0	366

detection by GA, which are sorted by the increasing number of false solutions. The results suggests that number of false solutions depend on $g'_{threshold}$ (adaptation) most strongly. Thus we consider we should take relatively high threshold for g'_i to improve the solution accuracy. Furthermore iteration numbers indicate that 1-3 true solutions are detected in one trial in average, which indicates the modification of GA algorithm for multi solutions problem are effective.

Table 2. Summary of result of crater detection by GA as an average of 10 trials. The notation of iter represents number of iterations required to find all correct solutions.

popula- tion	no. of genera- tion	$g'_{threshold}$	W	no. of false solu- tion	iter	cpu time (sec)
100	200	60	9	1.2	4.2	85
200	200	60	9	1.5	3.0	152
200	150	60	9	1.8	3.5	123
100	150	60	9	1.9	6.0	92
100	150	50	5	3.0	6.5	99
100	150	50	9	6.4	3.4	52

Binarized image of LDIM includes many noises because LDIM image has wider dynamic range than mars images. Figure 3 represents a example of circle detection for one of LDIM images. The results suggest that both method based on CHT and GA tend to detect many false solutions and to fail to detect some of clear craters due to the effect of noises. Among the two method CHT appears to be more robust than GA. We consider evaluation based on experiments using many lunar images and investigation of the optimiza-

tion of parameters are required in the future study, in order to make these methods more robust as component method for datamining system.

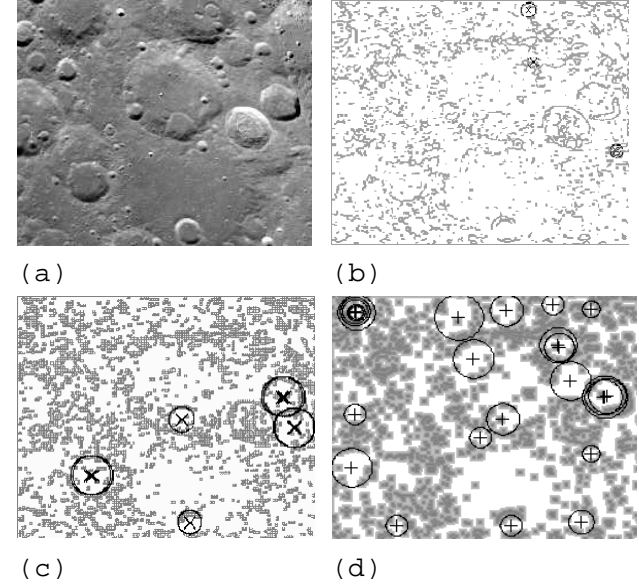


Figure 3. The results of circle detection by CHT and GA for bi59n262 of USGS LDIM. The figure (a) is the original grayscale image, (b) and (c) are the result of CHT for original image and pyramid image, respectively, and (d) is the results of GA. Crosses and solid circles represent the center of the detected craters and its rims, respectively

3.4 Categorization of craters by SOM

In order to categorize the craters, we cut out the crater images of their diameter width and height from original 'grayscale' images.

We implement Kohonen's self-organizing mapping (Kohonen, 1992) for crater categorization, since this method has previously been implemented for clustering of weather (cloud) images by GMS 5 (Katayama and Konishi, 1999), and it has been demonstrated that it can successfully categorize the image characteristics. Kohonen's self-organizing mapping uses a two-dimensional feature map whose cells have feature vectors. The distance between the input vector and each cell's feature vector is calculated and the input vector is placed into the closest cell. At the same time, the feature vector in the cells adjacent to the closest cell is modified so that it moves closer to the input vector. As a result of this iterative learning process, we obtain a feature map that reflects variations in the input vectors and at the same time adequate clustering of the input vectors.

We considered three types of input vectors for crater analysis.

- Image vectors normalized with respect to intensity and size. The normalized intensity is represented by

$$p'(i, j) = \{p(i, j) - p_{mean}\} / \sigma, \quad (5)$$

where $p(i, j)$ is the intensity of the i -th and j -th pixel for each crater image, and p_{mean} and σ are the average and the standard deviation, respectively.

- An intensity histogram calculated from the normalized image.
- Amplitude vector of FFT components calculated from the normalized image.

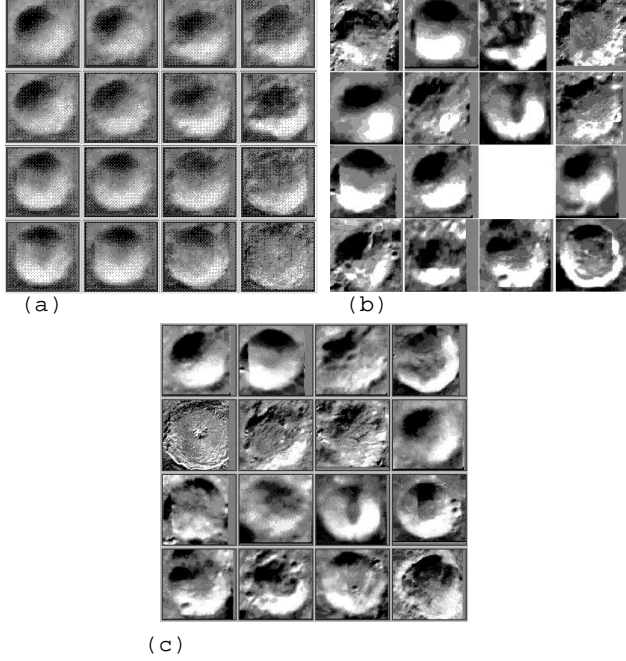


Figure 4. Feature maps for 37 craters obtained by SOM. The input vectors of (a), (b), (c) are normalized image vectors, intensity histogram, and FFT amplitude vector, respectively. Images on each cells of (a) is generated from feature map. Images for (b) and (c) are represented by the cut images that minimize the distance in each cell.

As preliminary experiments we processed two LDIM images. The craters are selected both visually and manually since neither the algorithm of GA or CHT is robust enough for extremely noisy lunar images at present. We selected 37 craters, including simple bowl-shaped craters, degraded craters, and flat floored craters. Each image is normalized to the size of 100-pixel square, and the intensities are normalized providing $p_{mean} = 128$ and $\sigma = 40$. The SOM block size

is taken to be 4×4 , the number of iteration is 5000, and the neighborhood distance at the iteration of t is given by $2(1 - t/5000)$.

Figure 4(a) shows thus obtained feature map taking the normalized image vector as the input. The simple bowl-shaped craters are clustered into the upper left cells, and the flat-floored craters are clustered in the lower left corner cells. This result indicates that normalized image vectors can successfully distinguish the variations in crater shape that are recognized in human vision. Figure 4(b) shows the feature map of the representative images after learning by SOM when the intensity histogram vectors are taken as the input vectors. This figure shows no significant meaning in the distribution of cut images in the feature map and thus a failure to learn the crater patterns from input vectors. Figure 4(c) shows the feature map of the representative images when the amplitude vectors of FFT component are taken as the input vectors. The obtained feature map reflects that characteristics of terrain bumps well. In particular, craters containing higher frequency bumps are clustered into the cells around the left, upper-middle cells.

Comparison between Figure 4(a)(b)(c) suggests that normalized image vectors and amplitude vectors of FFT component can classify crater images more successfully than histogram vector. In particular, normalized image vectors generates most comprehensive feature map in relation to human vision.

Furthermore, we have extended the above-described experiments to 170 craters located in the higher middle latitudes ranging from 56 degrees to 63 degrees N. Figure 5(a) shows the feature map obtained by SOM after 5000 iterations. The result represents that a similar patterns in the feature map as those observed previously in Figure 4(a), including a clustering of bowl-shaped craters and flat-floored craters in the corner cells on the diagonal line.

By visually classifying the input images in each of the cells, we were able to group the cells into four super-categories as shown in Figure 5(b). The number of the cells in each category ranges from two to ten. Table 3 presents a semantical description of these four categories. Categories 1 and 2 are composed of flat-floored craters, and 50 % of category 1 and 15 % of category 2 are central-peaked craters, respectively. Category 3, the intermediate between categories 2 and 4, consists of cone-shaped craters. Category 4, which accounts for 72 % of all the craters, consists of bowl-shaped craters.

The similarity of the results from two independent data sets (Figures 4(a) and 5(a)) suggests that the nor-

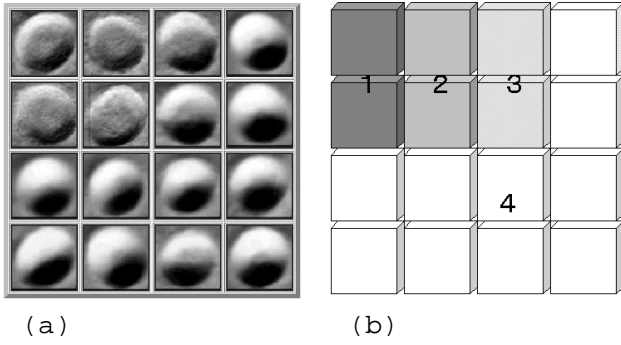


Figure 5. The result of categorization for 170 craters in the northern hemisphere (ranging from 58 degrees N to 63 degrees N) of the moon. (a) is the same as in Figure 4(a), and (b) represents its super-categories. The semantical description (b) is given in Table 3

Table 3. Semantical description of the crater categories.

Category	Number	Description	Percentage of central-peaked craters
1	19	Flat-floored.	50
2	13	Flat-floored.	15
3	16	Cone-shaped.	–
4	122	Bowl shaped.	–

malized image vectors are efficient as inputs for crater categorization problems. We have two other problems on normalization on imaging conditions: sun direction and solar altitude. Since both are known as the imaging conditions in most cases, the former problem is resolved by the rotation of images before the cut out, and the use of a map with higher resolution and grouping of super-category on it is effective with regard to the solar altitude problem.

4. Concluding Remarks

We examined the crater detection and categorization process in view of data mining from a large-scale scientific image database. The genetic algorithm and the combinational Hough transforms are studied as crater detection modules. Presently, CHT appears to be more robust for highly noisy lunar images, however, the optimization of various parameters for both methods should be considered in future studies.

Furthermore we investigated self-organizing mapping in relation to the categorization of craters, by taking normalized image vectors, intensity histograms, and amplitude vectors of FFT components, as input

vectors, respectively. The experimental results show that normalized image vectors produce the most meaningful feature maps that group simple bowl-shaped craters, cone-shaped craters, flat-floored craters and flat-floored craters with a central peak.

We continue to work on enhancement of robustness on each component process and the integration of the whole system. The adaptation of the domain-expert knowledge and knowledge discovery from its spatial distribution are also subjects in the future study.

5. Acknowledgements.

This research is supported by grants-in-aid for intensive research (A) (2) (Project 1130215) from the Ministry of Education, Science, and Culture of Japan.

References

- Burl, M. C. et. al. (1998), Learning to Recognize Volcanoes on Venus Machine Learning ,30, 165–195.
- Burl, M. C. et al.(1999), Mining for Image Content, *Systems, Cybernetics, and Informatics / Information Systems: Analysis and Synthesis*, (Orlando, FL)
- Fayyad, U. M. et. al. (1996). Automating the Analysis and Cataloging of Sky Survey. *Advances in Knowledge Discovery and Data Mining*, (pp. 471–493). AAAI Press.
- Katayama, K. & O. Konishi.(1999). Construction of Satellite Image Databases for supporting Knowledge Discovery(in Japanese). *IPSJ, Trans. on Database, 40-SIG5(TOD2)*, 69–78.
- Kohonen, T. (1997). *Self-Organizing Maps*, 2nd eds., Springer.
- Melosh, H. J.(1988). *Impact Cratering*, Oxford University Press.
- Nozette, S. et. al. (1994). The Clementine Mission to the Moon: Scientific overview. *Science*, 266 , 1835–1839.
- Smyth, P. et. al. (1996). Modeling subjective uncertainty in image annotation . *Advances in Knowledge Discovery and Data Mining*, (pp. 517-539). AAAI Press.
- Watanabe, T. & Shibata, T. (1990). Detection of broken ellipse by the Hough transforms and Multiresolutional Images(in Japanese), *Trans. IEICE, J83-D-2*, 159–166.