

An improved recognition algorithm for lunar terrain based on CCD image

Zhankai Li^{1,a}, Jiaheng Chen^{1,b}, Yiren Chang^{1,c} and Xiaolin Tian^{1,2,d}

¹ Faculty of Information Technology, Macau University of Science and Technology, Macau 999078, China;

² Lunar and Planetary Science Laboratory/Space Science Institute, Macau University of S and T, Macau.

^ajanlzk@126.com; ^bchenjiaheng1992@126.com; ^c1521485949@qq.com; ^dxltian@must.edu.mo.

Keywords: CCD image; Recognition algorithm; Lunar terrain; Mare and Highland

Abstract. With the development of space technology, topographic study of celestial bodies becomes increasingly important. In order to better carry out geomorphologic analysis and landing site selection of celestial bodies, the terrain classification becomes particularly critical. This paper provides an improved algorithm and proved its better in identifying the lunar mare area and the highland area of the CCD images with four features used in *k*-means clustering. We chose two typical areas: ‘H010’ and ‘SI’ areas of lunar terrain to research. And the result of the improved algorithm is analyzed from two different block size with different number of testing points. And also the whole recognition rate and Cohen’s kappa coefficient are both better than the result of previous algorithm in using DEM or CCD data. Especially in the ‘H010’ area, the average whole recognition rate is 91.4325%, and the average Cohen’s kappa coefficient is 0.813.

1. INTRODUCTION

The moon is Earth's nearest celestial body, has long attracted world attention^[1]. 17th century, Galileo's first telescope to the Moon, opened a prelude to human lunar exploration^[2]. In the 1960s to 1970s, the former Soviet Union and the United States have used data and images from spacecraft and ground-based telescopes which called *Luna*, *Apollo* and *Lunar Orbiter* to compile and publish a series of lunar atlas that covering the front and back of the moon, and it has become an important reference works of lunar science^[3~5]. Lunar surface topography characteristics were produced by endogenic and external agency, and it directly reflects the ups and downs of the lunar surface and spatial distribution characteristics^[6].

Geomorphic surface morphology is the most intuitive form of expression. Since geomorphology from the sprout, geomorphologists used the height and steepness to describe the qualitative attribute information of landscape entity, and used relative position to describe the spatial location information of landscape entity^[7,8].

In this paper, we use the CCD data of Chang'E-1 downloaded from *The Science and Application Center for Moon and Deep space Exploration*. And we improved the algorithm which based on Jiang's method^[10], we choose the suitable block size and set four features to distinguish and recognize the lunar mare and highland, and we compared two results from different block size. The result has been compared with the geological data published by USGS which has shown that the improved algorithm does well in blocking image for terrain classification.

2. THE IMPROVED RECOGNITION ALGORITHM

This section has been divided into three parts: the flowchart of the improved algorithm, details about four features and the clustering.

2.1. The flowchart of the improved algorithm

In order to show the whole processing of the improved algorithm, the flowchart has been shown in the Fig. 1.

*This work is supported by the Science and Technology Development Fund of Macao (No. 059/2013/A2).

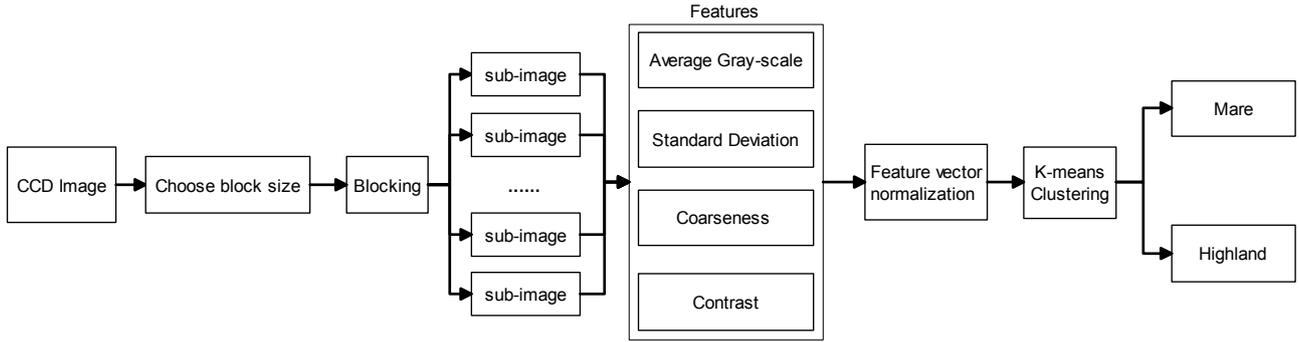


Fig. 1. The flow chart of improved algorithm proposed

2.2. Features

According to the characteristics of lunar terrain CCD image, we summarize some significant features. And in this part, four features of CCD data used for clustering in this paper will be discussed. They are the *Average Gray-scale*, the *Standard Deviation of gray-scale*, the *Coarseness* and the *Contrast*. Especially, the *Standard Deviation of gray-scale* and the *Coarseness* are the first time to be used in the CCD data for lunar terrain recognition.

2.2.1. The Average Gray-scale

In the lunar terrain CCD image, what is the most obvious difference between mare and highland? It's undoubted that the mare is much darker than the highland. Because of the main components of lunar mare are basalts, and the highland are Light-colored plagioclases^[9], the reflectivity of these two main components is different. So, we inevitably choose the Average Gray-scale to describe this difference. Formula to calculate the Average Gray-scale has been shown below:

$$F_{AGS} = \frac{\sum_{i=1}^m \sum_{j=1}^n G(i,j)}{m \times n} \quad (1)$$

where $G(i, j)$ is the gray-level of each pixel. The m is the length of image and n is the width of image.

2.2.2. The Standard Deviation

We always seen some new formula use the *Standard Deviation*, like *Light reflection property*^[10] and *Slope*^[11,19], but the *Standard Deviation of gray-scale* is the most effective way in lunar terrain recognition. Standard deviation is a measure of the degree of dispersion of a set of data's average value. When we get a large standard deviation, this means that there is a large difference between most value and its average, so the degree of dispersion is high. And a small standard deviation means that there is a small difference between most value and its average, so the degree of dispersion is low. As we all know, the same average does not mean the same standard deviation. Therefore, we can use standard deviation to distinguish two different terrain areas which have the same average gray-scale. For example, due to the large undulation of highland, it always has large gray-scale difference in an area, and maybe the average gray-scale in this area is the same as the area of mare. So, the standard deviation solves this problem. Formula to calculate the Standard Deviation has been shown below:

$$F_{SD} = \sqrt{\frac{\sum_{i=1}^m \sum_{j=1}^n (G(i,j) - \bar{G})^2}{m \times n}} \quad (2)$$

$$\bar{G} = F_{AGS} \quad (3)$$

2.2.3. Coarseness

Coarseness is a value, reflecting the granularity of the texture, which is the most basic texture feature. When the only difference between two texture patterns is primitives, the one with larger size of primitive model gives more rough feeling. The calculation method is as follows^[12]:

Step 1: Calculate the average gray-scale of $2^k \times 2^k$ pixels in active window.

In formula: $k = 0, 1, 2, 3, 4, 5$; $g(i, j)$ is gray-value which located in pixel (i, j) .

$$A_k(x, y) = \sum_{i=x-2^{k-1}}^{x-1+2^{k-1}} \sum_{j=y-2^{k-1}}^{y-1+2^{k-1}} g(i, j) / 2^{2k} \quad (4)$$

Step 2: For each point, take differences between pairs of averages corresponding to pairs of

non-overlapping neighborhoods just on opposite sides of the point in both horizontal and vertical orientations.

$$E_{k,h} = |A_k(x + 2^{k-1}, y) - A_k(x - 2^{k-1}, y)| \quad (5)$$

$$E_{k,y} = |A_k(x, y + 2^{k-1}) - A_k(x, y - 2^{k-1})| \quad (6)$$

$E_{k,h}$ is the differences in the horizontal case, $E_{k,y}$ is the differences in the vertical case.

According to each pixel points, we can pick the best size which gives the highest output value. In the other words, we use k , the max value of E to configure the best size S_{best}

$$S_{best}(x, y) = 2^k \quad (7)$$

$$E_k = E_{max} = \max(E_1, E_2, E_3, \dots, E_h) \quad (8)$$

Step 3: Finally, F_{crs} can be calculated using the average value of S_{best} of the image, m and n is the width and height of the picture

$$F_{crs} = \frac{1}{m \times n} \sum_{i=1}^m \sum_{j=1}^n S_{best}(i, j) \quad (9)$$

2.2.4. Contrast

Contrast refers brightness level difference between the brightest white in bright area and the darkest black in dark area in an image. The larger the range of pixel difference, the larger the contrast and vice versa. Its size is decided by following factors: gray dynamic range, the polarization degree of the histogram in black and white, the sharpness of the edge and the cycle of the repetitive patterns. Generally, contrast usually refers to the first two factors. There are many different methods for calculating the value of contrast due to different goals in their own files, like Weber contrast^[20]. But in this paper, we use the contrast which from Tamura^[12]. The calculation as shown in formula:

$$F_{con} = \frac{\sigma}{\alpha_4^{1/4}} \quad (10)$$

It can be easily seen that the variance σ^2 or standard deviation σ about the mean of the gray-levels probability distribution is more preferable as contrast. As is well known, α_4 represents the peak value of image gray value, defined by

$$\alpha_4 = u_4 / \sigma^4 \quad (11)$$

and u_4 is the fourth moment about the mean, the $\sigma = F_{SD}$.

3. CLUSTERING

Those four features: Average gray-scale, Standard deviation, Coarseness and Contrast are clustered by *k-means* method.

3.1. Clustering Distance

The weighted Euclidean distance model is considered as one of the most popular multidimensional scaling methods^[13]. The formula of the Euclidean distance model is^[14]:

$$d(i, j) = \sqrt{|X_{i1} - X_{j1}|^2 + |X_{i2} - X_{j2}|^2 + \dots + |X_{ip} - X_{jp}|^2} \quad (12)$$

where $i, j = (1, 2, 3 \dots p)$ are two object data in p dimension. Then a weight W is added for each variable according to the importance. The formula of the weighted Euclidean distance model is:

$$dw(i, j) = \sqrt{W_1 |X_{i1} - X_{j1}|^2 + W_2 |X_{i2} - X_{j2}|^2 + \dots + W_p |X_{ip} - X_{jp}|^2} \quad (13)$$

3.2. K-means clustering

K-means method is used as the clustering rule in this paper. Given a set of observations (x_1, x_2, \dots, x_n) , where each observation is a d -dimensional real vector, *k-means* clustering aims to partition the n observations into k ($\leq n$) sets $S = \{S_1, S_2, \dots, S_k\}$ so as to minimize the within-cluster sum of squares (WCSS) (sum of distance functions of each point in the cluster to the K center). In other words, its objective is to find^[15]:

$$V = \sum_{i=1}^k \sum_{x_j \in S_i} (x_j - \mu_i)^2 \quad (14)$$

where μ_i is the mean of points in S_i .

4. TESTING RESULT

There are two test areas of Chang'E-1 used in this paper: ‘SI’ area and ‘H010’ area (Table 1) which was downloaded from *The Science and Application Center for Moon and Deepspace Exploration*. The ‘SI’ area is 100m per pixel and the ‘H010’ area is 200m per pixel. And they are selected manually in order to compare with the previous algorithm^[10, 16].

Table 1. The details of the testing area

Name	Longitude	Latitude
H010	0° W ~ 18° W	0° S ~ 14° S
SI	26° W ~ 39° W	40° N ~ 50° N

4.1. Choosing block size

We have cut and tested a small region from ‘SI’ area to choose a suitable block size for our algorithm in this paper. Finally, we decide to use two block sizes in research, they are 50*50pxl and 100*100pxl in Fig. 2(a) to (d). So, we will analysis our result based on these two block sizes.

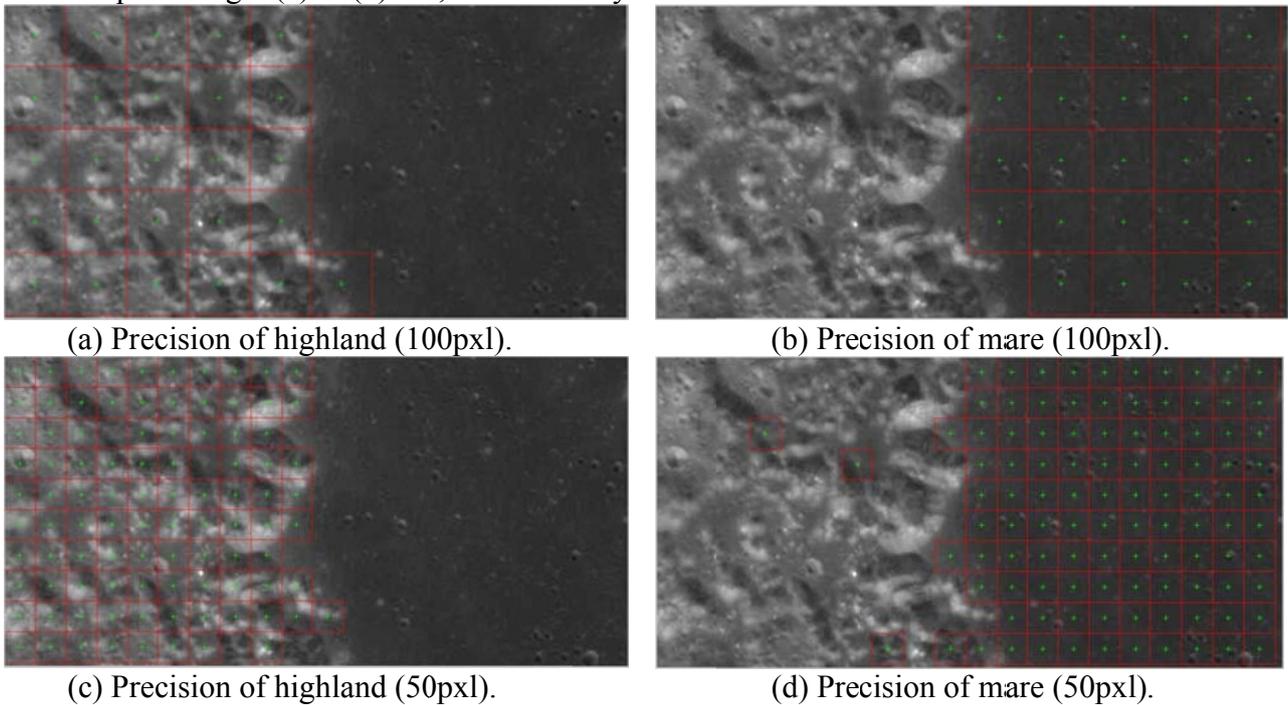
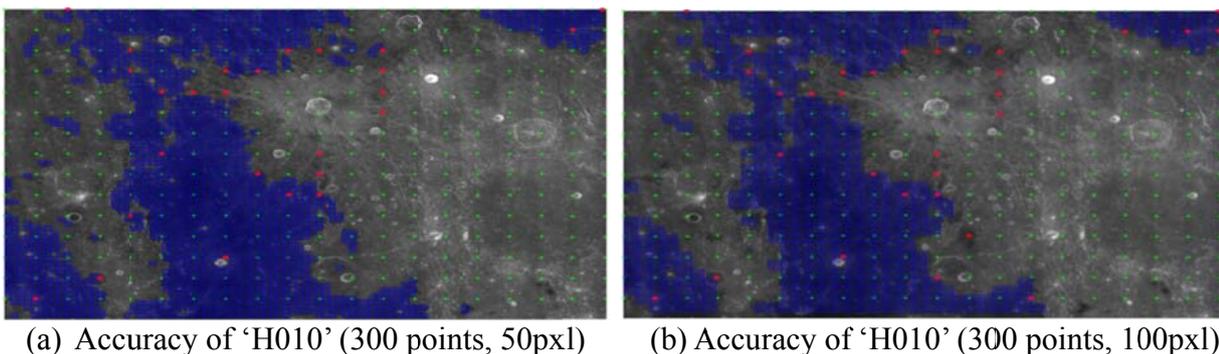
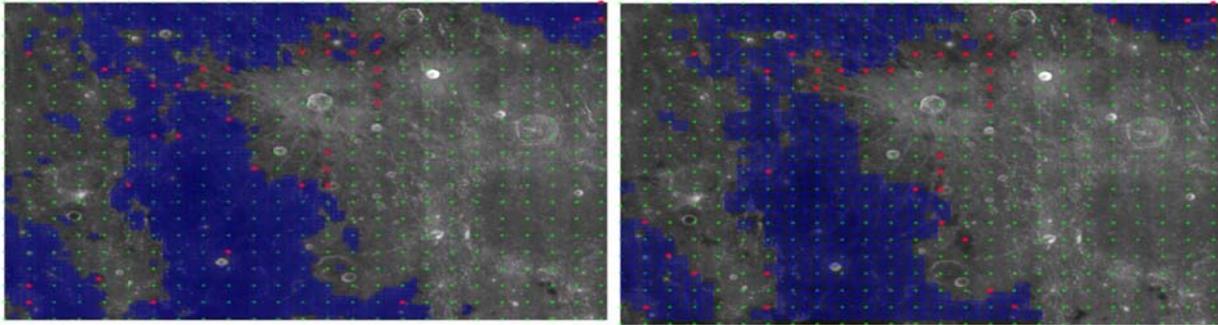


Fig.2. Two testing block sizes from ‘SI’ area

4.2. Testing Result

Because of the size of CCD data is different, and then there are different numbers of points select for each image to test how many areas can be recognized correctly by the improved algorithm. According to our test, we set 300 and 500 points for ‘H010’ area, set 340 and 525 (except the last column) points for ‘SI’ area. Fig. 3 (a) to (d) is the accuracy of ‘H010’, and Fig. 4 (a) to (d) is the accuracy of ‘SI’. The *blue* area is mare, and the point we set is *green*. By comparison with the data of USGS (NASA), the point will be marked as a *red* point in the area for warning the error.



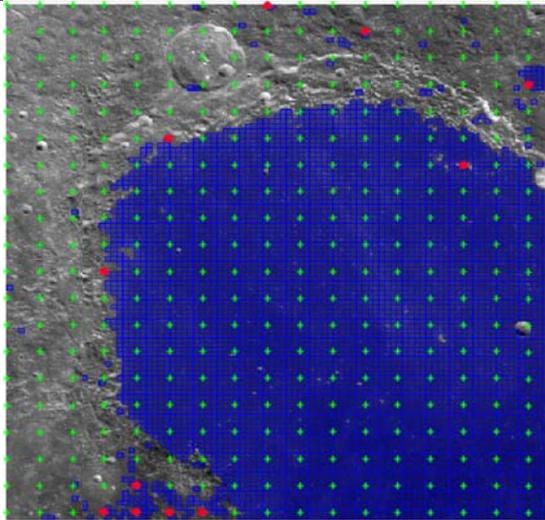


(c) Accuracy of 'H010' (500 points, 50pxl) (d) Accuracy of 'H010' (500 points, 100pxl)

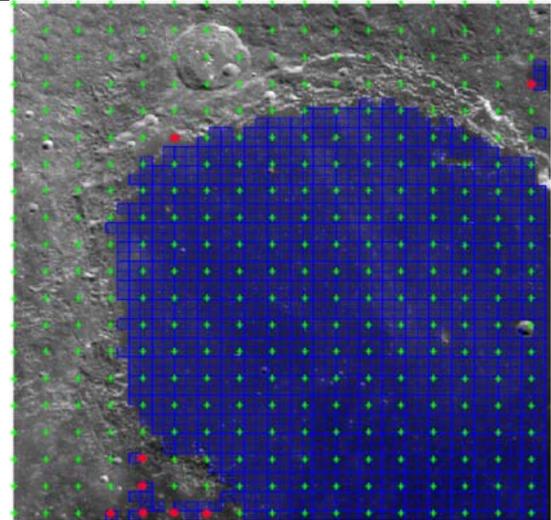
Fig.3. Testing results of 'H010' area

Table 2. Accuracy of 'H010'

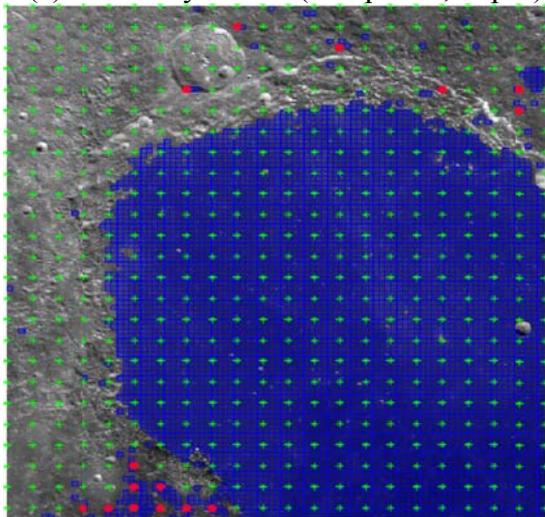
	Mare recognition rate	Highland recognition rate	Whole recognition rate	Cohen's Kappa coefficient
(a)	88.99%	92.67%	91.33%	0.8059
(b)	87.16%	91.62%	90%	0.7843
(c)	90.11%	93.71%	92.4%	0.8362
(d)	86.81%	94.97%	92%	0.8256



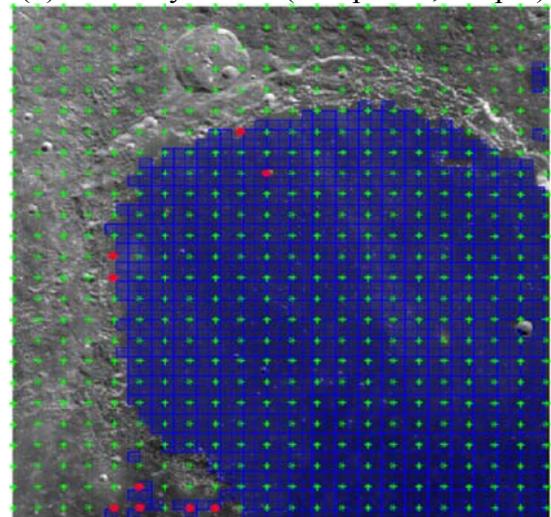
(a) Accuracy of 'SI' (340 points, 50pxl)



(b) Accuracy of 'SI' (340 points, 100pxl)



(c) Accuracy of 'SI' (525 points, 50pxl)



(d) Accuracy of 'SI' (525 points, 100pxl)

Fig.4. Testing results of 'SI' area

Table 3. Accuracy of ‘SI’

	Mare recognition rate	Highland recognition rate	Whole recognition rate	Cohen's Kappa coefficient
(a)	98.88%	94.41%	96.76%	0.9349
(b)	99.44%	95.65%	97.64%	0.9528
(c)	99.64%	94.33%	97.14%	0.9424
(d)	98.92%	97.57%	98.29%	0.9654

From the table 2 and the table 3, we can clearly see that the accuracy of each area, and the recognition rate shows good result for lunar terrain recognition.

5. CONCLUSION

We can get some conclusions from the testing results: There is a very high recognition rate in our algorithm which recognizes the mare and highland on the ‘H010’ and ‘SI’ of lunar terrain. There is different accuracy from different block size, but larger block size will lose many details and smaller block size will improve the erroneous recognition rate. Error in recognition can be found easily in the edge between mare and highland, also in the block, some areas in the highland have the same features as areas in the mare.

In order to evaluate the result and make sure the high accuracy is not an accident, so we had calculated the Cohen's kappa coefficient^[18] for each condition. Finally, we get the average Cohen's kappa coefficient for ‘H010’ and ‘SI’ respectively. The value for ‘H010’ area is 0.813 and ‘SI’ area is 0.9489. Both of them are not only higher than the value given by other previous algorithms^[16] that is 0.78 with the best conditions in the same area or some similar surface, but also better than the result given by Jiang^[10] which is 0.802. And the average whole recognition rate of ‘H010’ is 91.4325% and ‘SI’ is 97.4575%. They are also better than the latest result given by Wang^[17] which is 85.7%.

However, there are still some problems, which need to be solved, like how to maintain the recognition rate in the case of lifting the details. We will continue work on them and will try to improve the algorithm further in the future.

REFERENCES

- [1] Ziyuan OuYang, Introduction to lunar science (in Chinese). Beijing: China Aerospace Press, 2005
- [2] Xiaogang Ning, King B, Jixian Zhang. Study of Digital Lunar Mapping Based on Clementine Imagery and ULCN2005 (in Chinese), 2009, 38: 290-295
- [3] Akademia N, Barabashov, Nokolai P, et al. Atlas Obratnoi Storony Luny. Moskva: Izd-vo-Akademii Nauk SSSR, 1960
- [4] Lewis H A G. The Times Atlas of the Moon. London: Times Newspapers, 1969
- [5] Bowker D E, Hughes J K. Lunar Orbiter Photographic Atlas of the Moon. Houston: Lunar and Planetary Institute, 1971
- [6] Wenzhe Fa, TTheoretical Modeling and Parameters Inversion for Microwave Remote Sensing of Lunar Surface (in Chinese), Ph.D. Thesis. Shanghai: School of Information Science, Fudan University, 2008
- [7] Juzhang Li, A preliminary study on the quantitative index of basic types of geomorphologic form in china (in Chinese). Journal of Geographical Sciences, 1982, 37: 17-26
- [8] Chenghu Zhou, et al. Digital topography analysis and mapping (in Chinese). Beijing: Science press, 2009.

- [9] Zongqing Chen, Introduction of the lunar surface topography (in Chinese). Taiwan: Astronomical education bimonthly, 2009
- [10] Hongkun Jiang, Xiao-Lin Tian and Ao-Ao Xu, A new segmentation algorithm for lunar surface terrain based on CCD image, Research in Astron. Astrophys. 2014 December 30.
- [11] Xuejun Liu, WANG Ye-fei, CAO Zhi-dong, LI Jun-feng, TANG Guo-an, 2004. A study of error spatial distribution of slope and aspect derived from Grid DEM. journal of Bulletin of Surveying and Mapping, 0494-0911(2004)12-0011-03.
- [12] Tamura, H., Mori, S., & Yamawaki, T. (1978). Textural features corresponding to visual perception. Systems, Man and Cybernetics, IEEE Transactions on, 8(6), 460-473.
- [13] Jan de Leeuw. 1978. A new computational method to fit the weighted euclidean distance model. PSYCHOMETRIKA-VOL.43, NO.4. DECEMBER, 1978
- [14] DONG Xu, WEI Zhen-jun. 2005. A clustering method of euclid distance with weights. Journal of information engineering university, vol 6, no 1.
- [15] JA Hartigan, MA Wong, Algorithm AS 136: A k-means clustering algorithm, Applied statistics, 1979-JSTOR
- [16] Zhou Z P, Cheng W M, Zhou C H, et al. Characteristic analysis of the lunar surface and automatically extracting of the lunar morphology based on CE-1 (in Chinese). Chinese Science Bulletin (Chinese Ver), 2011, 56: 18—26, doi: 10.1360/972010-1375
- [17] Chenzhi Wang, et al, A Method for Identifying the Lunar Morphology Based on Texture from DEMs (in Chinese), Vol.17, No.1. Jan., 2015.
- [18] Cohen J. A coefficient of agreement for nominal scales. EducPsychol Meas. 1960; 20(1): 37–46
- [19] Clapham, C.; Nicholson, J. 2013. Oxford Concise Dictionary of Mathematics, Gradient. Addison-Wesley.
- [20] Denis G. Pelli , Peter Bex .2013. Measuring contrast sensitivity. Volume 90, Vision Research Pages 10–14.