MARS MULTISPECTRAL IMAGE CLASSIFICATION USING MACHINE LEARNING TECHNIQUES

Lilan Pan, Chen Gui, Dave Barnes, and Changjing Shang

Department of Computer Science, Aberystwyth University, UK

ABSTRACT

This paper presents a novel application of machine learning techniques for Mars rock detection using multispectral data. The feature set contains spectral data captured from the NASA MER Pancam instruments. The slope features, PCA features, statistic features and features in different colour space derived from the raw multispectral data are also added to the full feature set in order to enlarge the searching range of optimized features. Fuzzy-rough feature selection (FRFS) is employed to generate good feature sets with lower dimension. Some machine learning classification methods (1NN, 5NN, Bayes, SMO and Dtree) and cluster method (FCM) are utilized to classify the rock from soil using the selected feature. The experimental results show that the FRFS can produce a low-dimensional feature set with improved classifying and clustering results thereby enhancing the efficacy and accuracy of rock detection.

Key words: Mars image classification; FRFS; Multispectral.

1. INTRODUCTION

Nowadays, the rovers on Mars are equipped with a number of instruments, and hence they are able to collect a considerable quantity of scientific data. These data are transmitted back to Earth where scientists can analyse the data to gain scientific information. In addition, scientists send the command sequence to Mars to control the rovers after analysing these data. However the transmitting bandwidth between Mars and Earth is limited. Therefore, to enhance the efficiency, applying automatic technology for Mars exploration is appropriate.

In the variety of data, image data are numerous and informative. For example, Curiosity, the lastest rover which landed on Mars on August 2012, carries 17 camera instruments which are eight HazCams, four NavCams, two MastCams (analogous to Pancam), a ChemCam, MARDI and MAHLI. If the automated analysis of Mars images was implemented, the transmission quantity could be significantly reduced.

Panoramic Camera (Pancam, in Curiosity called MastCam) is one of most important image capture instruments. The Pancam system of the Mars Exploration Rover Mission (MER, including two rovers: Spirit and Opportunity) is a multispectral, stereoscopic, panoramic system consisting of two digital cameras capable of obtaining color images (synthesized using multispectral data) to constrain the mineralogic, photometric, and physical properties of surface materials [1]. As to the forthcoming ESA ExoMars project, the Panoramic Camera (PanCam) imaging system is also designed to have the ability to obtain high-resolution colour and wide angle multispectral stereoscopic panoramic images [2]. The multispectral data accessed from Pancam instruments provides abundant information for autonomous science analysis.

The major scientific objectives of the Pancam images are the identification of surface targets such as outcrops, ridges and troughs and the variety of rocks. Thereinto, one of key elements is to detect rocks from the captured images. However, rock objects exhibit diverse morphologies, colours and textures on Mars. They are also often covered in dust or partially embedded in the terrain thereby increasing the difficulty of identification. Some similar rocks may look different with variable coatings and dust mantles. A rock may show various appearances under different angles of sunlight. Shadows may lead to a negative effect for rock detection as well.

Although the reasons above weaken the relationship between some features and the essence of a rock, we believe that there are some features that exist to discriminate rocks from soil. There are many features applied in rock detection such as edge-based features [3, 4], morphological features [5] and statistics-based features [6]. Here we propose that the multispectral data captured by a Pancam instrument and the derived features could be use to characterise a rock. However, not all features facilitate rock detection. Some irrelevant and random features will reduce the efficiency and even decrease the accuracy of classification. Thus, we propose an approach that uses the raw multispectral data to produce many features from where we can select the most effective features to perform rock detection.

2. GENERATION AND SELECTION OF FEATURES

We downloaded the MER Spirit multispectral image data in *.img* format from the NASA planetary data system (PDS) archives to generate the features for selecting. The data can be represented as an image sized of 512×512 pixels. For the feature selection method, we adopted a fuzzy-rough feature selection algorithm based on fuzzy similarity relations.

2.1. Raw Data

Each Pancam camera from MER is equipped with an eight position filter wheel, providing the multispectral imaging capabilities. The detailed wavelength and band pass of each filter is shown in Table 1. Among all the 16 filters, the filters L2-L7 and R1-R7 are designed for the geology purposes. In other words, the spectral data captured by these filters can provide information relating to Mars geology. Thus by analysing these data, we can find the distinction between the rocks and soil (regolith). Some instances of spectral data are illustrated in Figure 1.

Table 1. MER Pancam Characteristics

Name	Wavelength (nm)	Band Pass (nm)
	Left Camera	
L1	739	338
L2	753	20
L3	673	16
L4	601	17
L5	535	20
L6	482	30
L7	432	32
L8	440	20
	Right Camera	
R1	436	37
R2	754	20
R3	803	20
R4	864	17
R5	904	26
R6	934	25
R7	1009	38
R8	880	20

In Figure 1, it can be seen that the spectral value of rock and soil varies slightly when the wavelength is greater than 700 nm. Thus, to simplify the data generating process, we chose multispectral data only from the left geology cameras (i.e. L2-L7, the spectral range from 432 nm to 753 nm). As to the type of these data, we used the RAD data (which can be radiometrically-corrected calibrated to absolute radiance unit), and then converted to R* (R-star) data. R* was defined as "the brightness of the surface divided by the brightness of an RT (Radiometric Calibration Target) scaled to its equivalent Lambert reflectance" [8]. It can be calculated as following:

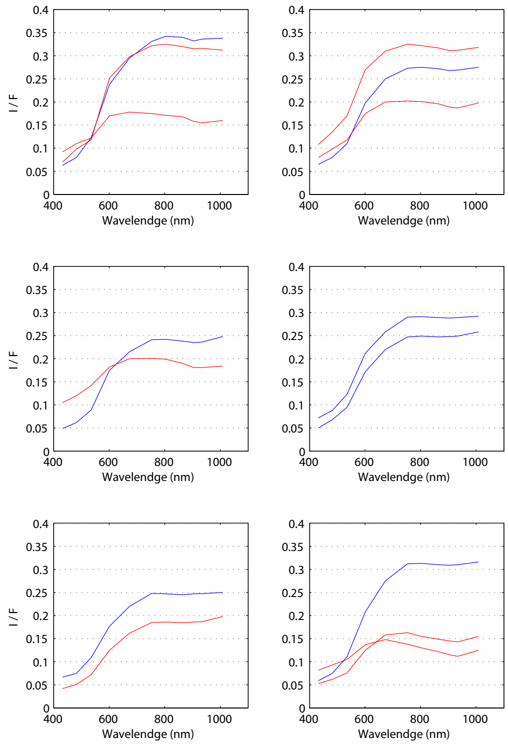

Figure 1. Some examples of multispectral curves captured by the Spirit Pancam. Red curves represent rocks, blue curves represent soils. Detailed information can be found in [7].

$$Rstar = DN * RSF + RO$$

In which, DN is the digital number value (intensity) of the image data, RSF is the radiance scaling factor and RO is the radiance offset (All these parameters are stored in the *.img* files). R* data are useful for classification by the reason that they allow for direct comparison between spectra taken at different times of day.

2.2. Feature Generating

The original data did not consider the intensity and direction of the light so that they can hardly become the optimized feature set for classification. A rock in frontlight may exhibit a different appearance when it is in backlight. Hence we made some preliminary features from the original data in order to search the optimized feature set to recognize the rocks from soil.

Firstly we extracted the slope features between each adjacent sample spectrum. For example, the slope between 432nm and 482nm is $(R*_{482} - R*_{432})/(482 - 432)$. Because all of the features should be normalized before classification, the function was simplified to $(R*_{482} - R*_{423})$. Finally 5 features of slope were obtained.

Table 2. *Description of each feature*

Feature No.	Meaning
1-6	Original spectral radiance data
7-11	Slope between 2 adjacent spetra
12-13	Mean and variance of original data
14-15	The PCA first 2 components
16-18	CIEXYZ
19-21	RGB
22-24	CIELab

In addition, we computed the mean value and variance and regarded them as 2 features. The mean value reflects the intensity of illumination, and the variance reflects the fluctuation of spectral data to some degree.

Principal component analysis (PCA) which is found to be a useful tool for interpreting compositional variation has been applied to MER datasets [10]. In our current work, all 6 original R* data were subjected to PCA to find the components including more information. Here we picked the first 2 components from PCA as features. The cumulative energy of these 2 components was more than 99.9%.

Additionally, we converted the multispectral data to three different colour spaces: CIEXYZ, RGB and CIELab. CIEXYZ can reflect the light tristimulus values to the human eye. CIELab represents lightness and colour information in different channels independently. RGB is the most popular space to synthesize colour image. Each of these colour spaces contains three channels, and hence we obtained 9 features.

From the above feature generating methods, we obtained 24 features in total. All features were normalized to range from 0 to 1 for classification. An example of all the normalized features represented by grey-scale images is shown in Figure 2.

For easy cross-referencing, Table 2 lists the reference numbers of the features that may be selected.

2.3. Fuzzy-Rough Feature Selection

After we established the full feature set containing 24 features, a Fuzzy-Rough feature selection (FRFS) method [9] was applied to find a good feature subset for classification. It has also be used in the work of *McMerdo* image classification [6]. This FRFS method uses a fuzzy similarity measure to calculate the degree of dependency. The subsets with a high degree of dependency perform better classification than those with low dependency.

The *QuickRUDUCT* algorithm has been applied to accelerate the progress of feature selection. This algorithm chooses the feature which increases mostly the dependency of the current subset and adds the feature into it, and thus the subset added by this feature has a higher dependency than the previous one. The algorithm will terminate when the addition of any remaining feature does not increase the dependency.

For the FRFS parameter of our work, the Lukasiewicz t-norm and implicator were taken as fuzzy connectives. The similarity relation function used is as follows, in which σ_a is the variance of feature a.

$$\mu_{R_a} = \max(\min\Big(\frac{a(y)-(a(x)-\sigma_a)}{a(x)-(a(x)-\sigma_a)}, \frac{(a(x)+\sigma_a)-a(y)}{(a(x)+\sigma_a)-a(x)},\Big), 0)$$

3. EXPERIMENTS AND RESULTS

We conducted experiments using the MERA (Spirit rover) multispectral data of Sols 601-602. The purpose of our experiment focused upon the rock and soil detection. The detailed experimental results and analysis are presented below.

3.1. Feature Selection and Classification in Local Image

Five classic machine learning algorithms were applied to test the performance of our approach. These were: 1-nearest neighbors algorithm (1NN), 5-nearest neighbors algorithm (5NN), naive bayes algorithm (Bayes), decision tree J48 algorithm (Dtree) and SVM by Sequential Minimal Optimization (SMO). In every image (multispectral data), 50 pixel points of rock and soil were selected into the training set for feature selection respectively. The feature selection and classification process of one image were separate to the process of the other images. That is to say, an image only selects "local" feature(s) from itself and for the classification of itself. Some classified results with the related selected features are illustrated in Figure 3. From visual inspection, it can be seen in the figure that the 5NN, Bayes and SMO classifier gained better results than the Dtree and 1NN.

Moreover, to prove that the features selected by our method can classify the rock and soil effectively and efficiently, we compared the results between the classifications using different features (selected features, full features (1-24), origin features(1-6) and some random features). The random feature sets have the same number of features as the selected feature sets. Since the results using different classifiers are similar, we only show the classified results by Bayes in Figure 4. In the comparison between the classification results using different feature sets, it can be seen that classified results using the selected feature set are approximate to the full and original feature sets, but the amount of features used are reduced. In addition, as to the random feature sets, which contain the same number of features to our features selected by FRFS, cannot obtain results as accurate as the one using our selected feature set.

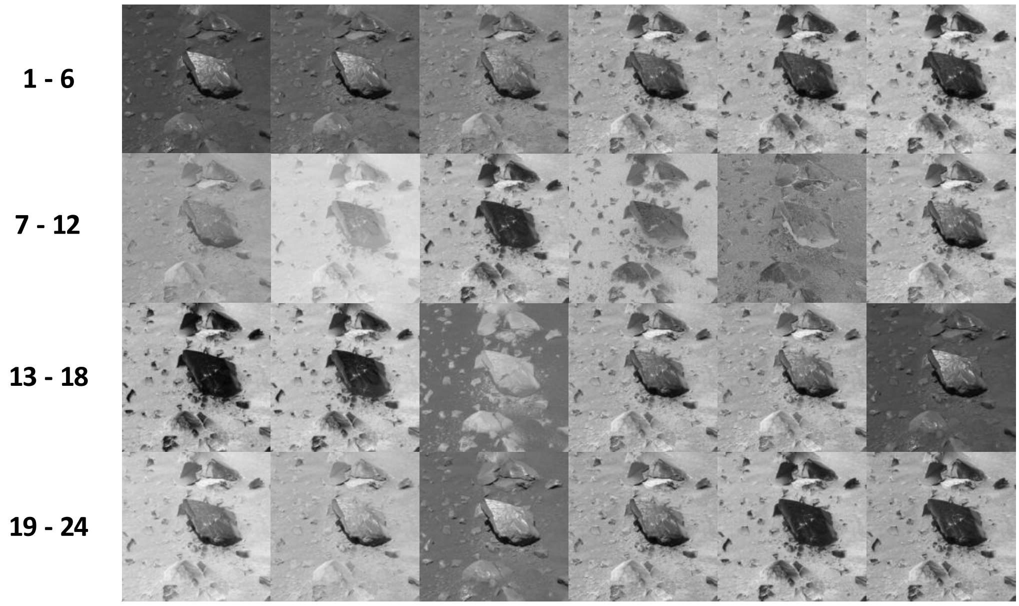

Figure 2. An example of all 24 feature of a image (see Table 2).

3.2. Global Classifier for Rock Detection in a Scene

In the classification above, although the method provides a competent performance, we focus only on the classification of an image itself. It is of lesser significance to reduce the quantity of transmission in the respect that we have to transmit images to Earth for generating a training set. Therefore, we have attempted to establish a trained classifier for a certain scene on Mars. Essentially, we need to find a "global" feature set for the classification problem.

To build the global classifier with less features, we picked 50 pixel points of rock and 50 pixel points of soil from 5 different multispectral data images respectively. Thus we produced a training set which had 250 instances of rock and 250 instances of soil. We applied FRFS to the training set with 24 features to find the most effective feature set for classifying. We obtained a reduced feature set that contains feature 1, 4, 7, 10, 11, 13, 15 and 23. Then we used the training set containing only these features to establish the global classifier to detect rock. The Bayes classified results of images from which we picked pixel points as training set are shown in Figure 5.

In addition, we tested the global classifier on Martian images in similar scenes from which we have not extracted points to generate features and obtained qualified results. A result example is shown in Figure 6.

Figure 6. A classified result of an image from which we have not extracted pixel points for training.

3.3. Cluster Results by Selected Feature

Besides the classification approaches used here, the issue of detecting rock from soil can be solved by cluster methods. The 2 class clusterer can replace a classifier to deal with the problem of a faster computative speed and without the need for training sets. Thus we used fuzzy C-means (FCM) to cluster the data to verify if FRFS can select suitable feature(s) for clustering. Both local and global selected feature sets were used for clustering. In comparison, the clustering results using full features, original features and some random features are given. The comparative results are shown in Figure 7.

It is clear to see that the results of the feature set selected by our algorithm, whether global features or local features, gained an improved performance. The rocks covered by dust can be also clustered into the class of rock using our selected features, while the rocks in the clustered results using other feature sets are incomplete. We found that the clustering result using selected features performed even better than the full feature sets. It is

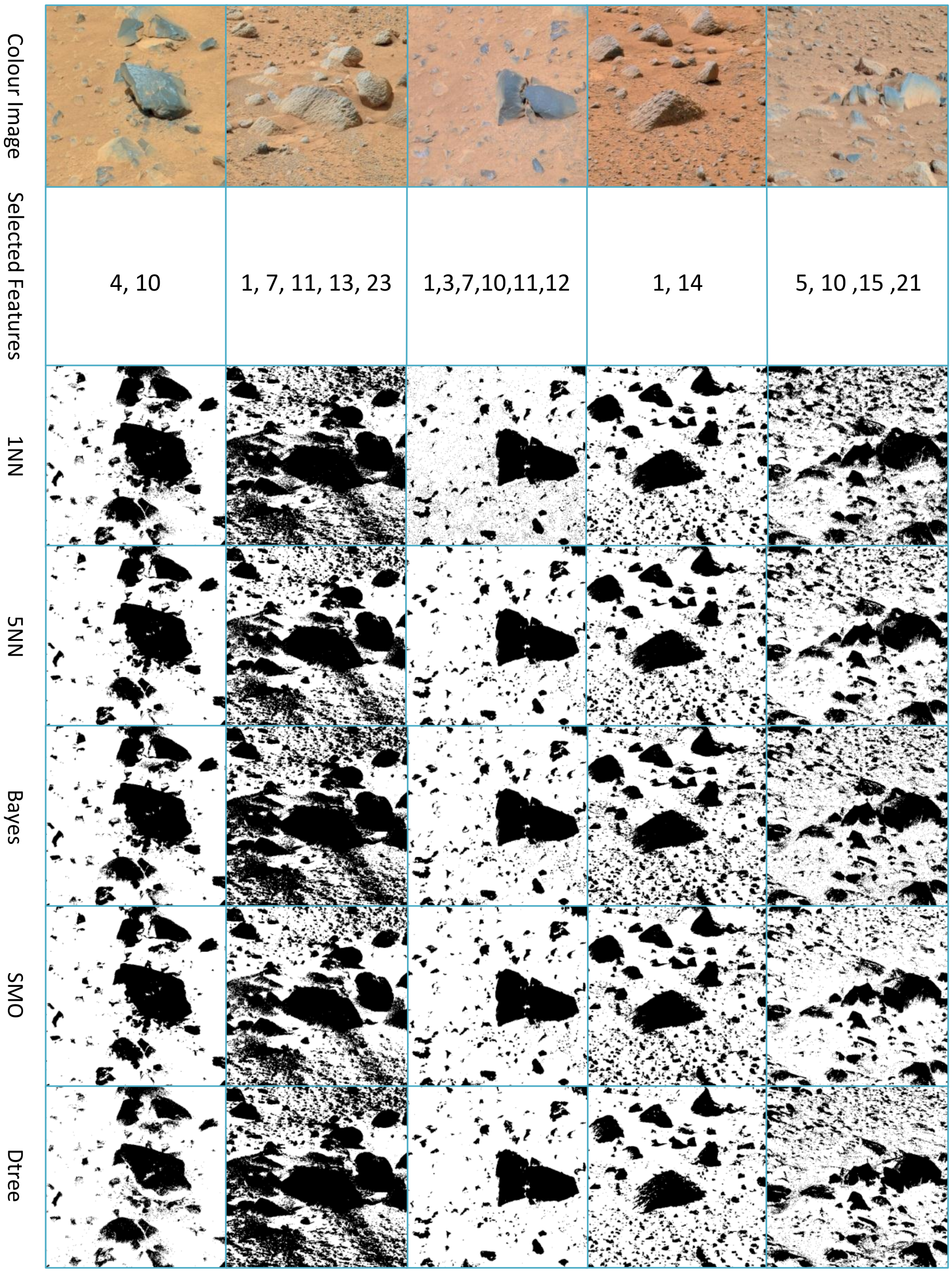

Figure 3. The selected features and the classification results by these features using different classification algorithms.

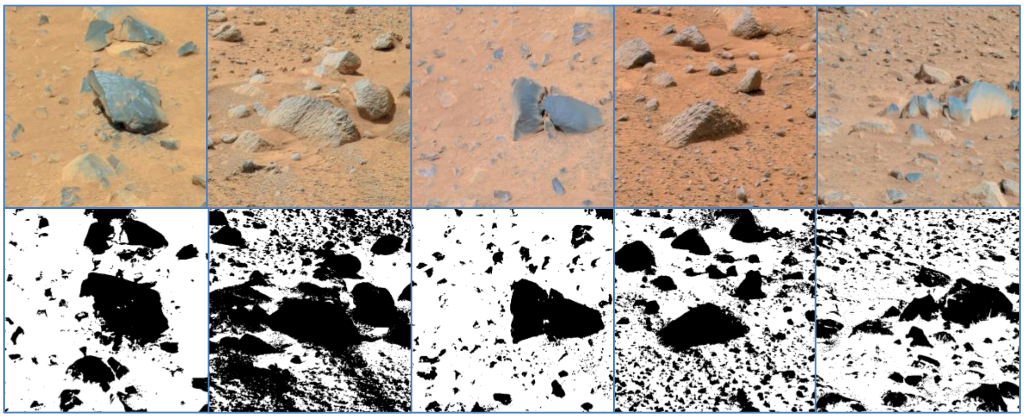

Figure 5. Global classification results.

proven that the FRFS is capable not only to reduce the number of features but also to remove the random and unrelated features which may disturb the cluster result.

4. CONCLUSION

In this paper, we have used Martian multispectral data captured by MER Spirit to generate several multispectral-derived features. The FRFS algorithm was applied in seeking suitable and optimized feature set for classifying and clustering. The results showed that our method can deal with problem of the rock detection effectively. In future work, other features such as band depth and red/blue ratio [10] will also be introduced into our method in order to find the most optimized feature subset thereby being adapted to more complicated environments.

REFERENCES

[1] J.F. Bell III et al., 2003, Mars Exploration Rover Athena Panoramic Camera (Pancam) investigation, J. Geophys. Res 108.

[2] Griffiths A.D, Coates A.J, Jaumann. R, et al., 2006, Context for the ESA ExoMars rover: The Panoramic Camera (PanCam) instrument. International Journal of Astrobiology, 5(3), pp.269-275.

[3] D.R. Thompson and R. Castano, 2007, Performance comparison of rock detection algorithms for autonomous planetary geology, Aerospace Conference.

[4] C. Gui, D. Barnes and L. Pan, 2012, An Approach for Matching Desired Non-feature Points on Mars Rock Targets Based on SIFT. TAROS 2012: 418-419.

[5] Mark Woods, Andy Shaw, Dave Barnes, et al., 2009, Autonomous science for an ExoMars Rover-like mission, J. Field Robotics, 26(4), pp. 358-390.

[6] Changjing Shang and Dave Barnes, 2013, Fuzzy-rough feature selection aided support vector machines for Mars image classification, J. Computer Vision and Image Understanding, 117(3), pp.202-213.

[7] J.F Bell, S.W Squyres, R.E Arvidson, et al, 2004, Pancam multispectral imaging results from the Spirit Rover at Gusev Crater. Science, 305.

[8] R. Reid, P. Smith, M. Lemmon, et al. Imager for Mars Pathfinder (IMP) image calibration. Journal of Geophysical Research-Planets, 104(E4):8907C 8925, Apr. 25 1999.

[9] Jensen R. and Shen Q., 2009, New Approaches to Fuzzy-rough Feature Selection. IEEE Trans. Fuzzy Syst. 17(4), pp. 824-838.

[10] R.B. Anderson, J.F. Bell III, 2013, Correlating multispectral imaging and compositional data from the Mars Exploration Rovers and implications for Mars Science Laboratory, Icarus, 223(1), pp. 157-180.

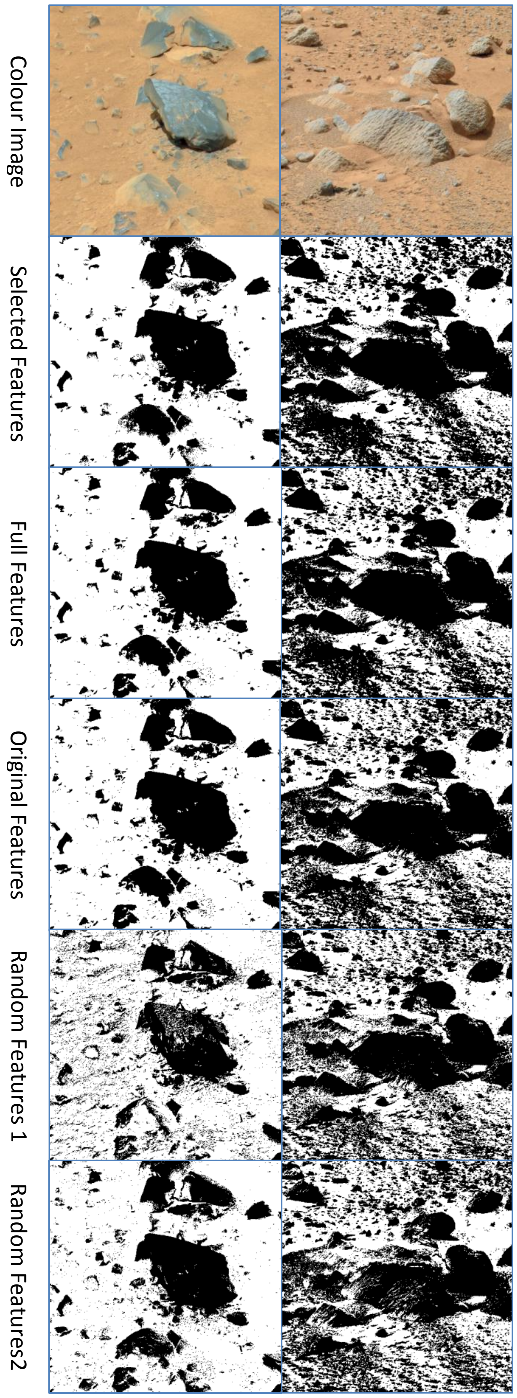

Figure 4. The comparison between selected feature subset to other feature sets.

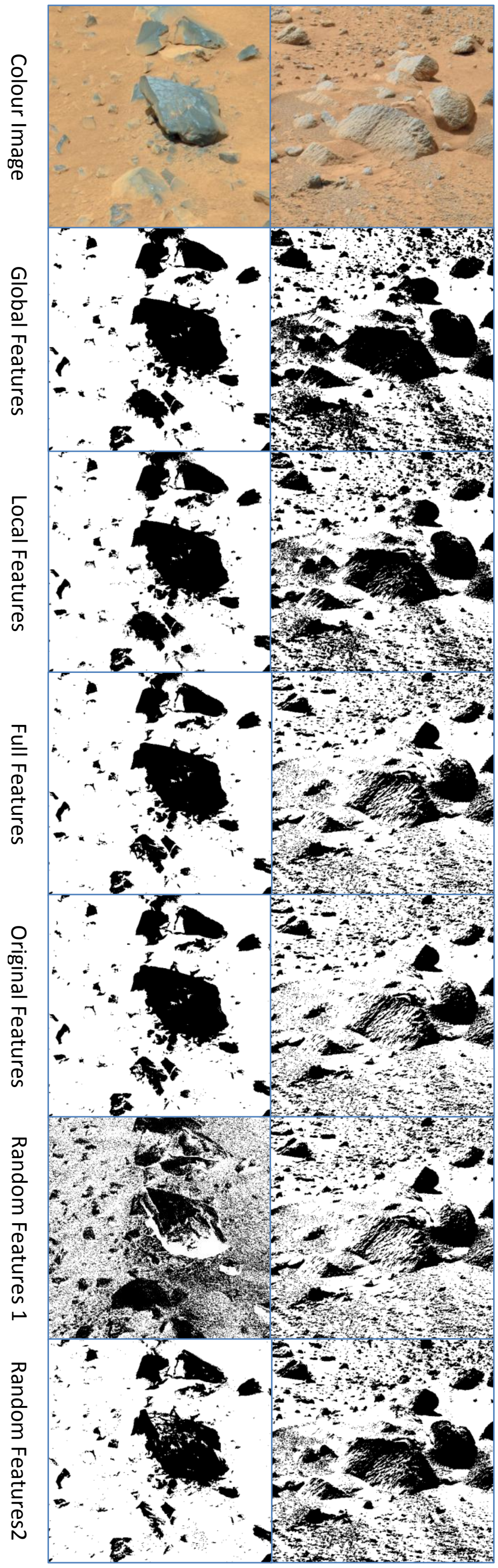

Figure 7. The clustering results using different feature sets.