PRIFYSGOL ABERYSTWYTH UNIVERSITY

**Aberystwyth University**

# *Combining Support Vector Machines and Information Gain Ranking for Classification of Mars McMurdo Panorama Images*

Barnes, David Preston; Shang, Changjing

*Publication date:*
2010

*Citation for published version (APA):*
Barnes, D. P., & Shang, C. (2010). *Combining Support Vector Machines and Information Gain Ranking for Classification of Mars McMurdo Panorama Images*. 1061-1064. http://hdl.handle.net/2160/5736

# COMBINING SUPPORT VECTOR MACHINES AND INFORMATION GAIN RANKING FOR CLASSIFICATION OF MARS MCMURDO PANORAMA IMAGES

*Changjing Shang and Dave Barnes*

Dept. of Computer Science, Aberystwyth University, UK
email: {cns, dpb}@aber.ac.uk

**ABSTRACT**

This paper presents a novel application of support vector machine (SVM) based classifiers for Mars terrain image classification. SVMs are applied in conjunction with information gain ranking (IGR) that allows the induction of informative feature subsets from sample descriptions of feature vectors of a higher dimensionality. Such an integrated use of IGR and SVMs helps to enhance the effectiveness and efficiency of the classifiers, minimizing redundant and noisy features. This work is supported with comparative studies – the resultant SVM-based classifiers generally outperform MLP and KNN-based classifiers and those which use PCA-returned features.

***Index Terms***— Mars image classification, support vector machines, information gain ranking, feature selection.

## 1. INTRODUCTION

Automated detection and classification of objects within Mars images, including different types of rocks and their surroundings, is of practical significance to the exploration of Martian environment [2, 10]. However, Mars images vary considerably in terms of intensity, scale and rotation, and are of large scale and blurred with noise [8]. These factors make Mars image classification a very challenging problem, demanding both effective and efficient techniques.

One critical step to successfully classify Mars images is to extract and use informative features only. To capture essential image characteristics, many features may have to be extracted without explicit prior knowledge of what properties might best represent the original image. Yet, generating more features increases computational complexity and in the mean time, not all such features may be essential to perform classification. The use of extra features may even cause the reduction of the overall discrimination ability of the feature set [4] and hence, that of the classification accuracy [5]. Thus, it is desirable to employ feature selection methods that can find the most significant features, based on sample measurements, to simplify and improve the classification process.

This paper presents an integrated approach for performing large-scale Mars image classification, by exploiting the potential of advanced classification and feature selection techniques. In particular, support vector machines (SVMs) [11] are employed for image classification. This is due to the recognition of their high generalization performance in complex data sets [1]. Information gain-based ranking (IGR) is adopted for feature selection, due to its computation simplicity and proven performance [9]. The resulting integrated approach helps to improve the effectiveness and efficiency of SVM-based image classifiers. This is because only those informative features are required to be generated in performing actual classification, minimizing both the feature measurement noise and the computational complexity (of both feature extraction and feature pattern-based classification). Such a property is of great importance to on-board image classification in future Mars rover missions. This is because flight projects demand least memory requirement and simplest computation possible (in order to minimize loads and increase software reliability).

The paper is organized as follows. Section 2 introduces the Mars images under investigation. Sections 3, 4 and 5 outline the key techniques used, including feature extraction, selection and classification. Section 6 shows the experimental results, supported by systematic comparative studies (with MLP and KNN-based classifiers that use IGR-selected or PCA-returned features). The paper is concluded in Section 7, where prospects for further research are discussed.

## 2. IMAGE DATABASE

The *McMurdo* panorama image obtained by NASA's Mars Exploration Rover is shown in Fig. 1, sized $22348 \times 5771$. This image, excluding the rover's instrument and shadowed areas, is used for the work here. Sixteen significant image types (i.e. classes) are listed in Table 1 and illustrated in Fig. 2. The ultimate task of this research is to detect and recognize image regions of such classes.

## 3. FEATURE EXTRACTION

In this work, local histograms and the first and second order statistics of color and grey-level images are used to produce feature vectors representing each pixel. Such features are effective in depicting the underlying image characteristics, efficient in computation, robust to image translation and rotation [1, 6], thereby suitable for classification of Mars images.

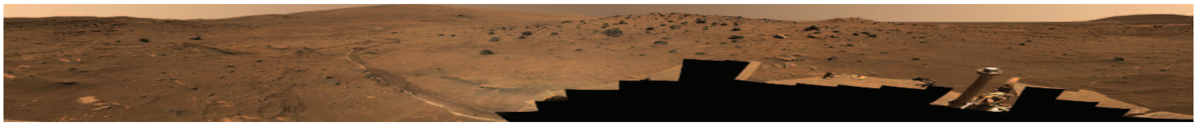

**Fig. 1**. Mars McMurdo panorama image with the size of $22348 \times 5771$

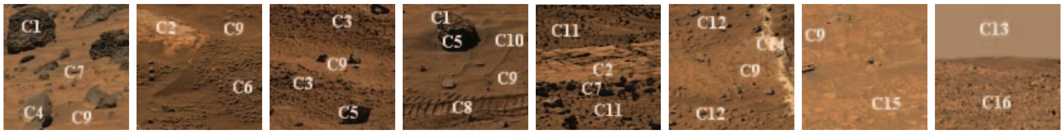


**Fig. 2**. Image types (classes)

| Class | Label | Class | Label |
|---|---|---|---|
| textured dark rock | C1 | bedding rock | C2 |
| mud | C3 | gray smoothed rock | C4 |
| black smoothed rock/shadow | C5 | gravel-1 | C6 |
| gravel-2 | C7 | rover tracks-1 | C8 |
| sand | C9 | sand ripple | C10 |
| gravel-3 | C11 | gravel-4 | C12 |
| sky | C13 | rover tracks-2 | C14 |
| gravel-5 | C15 | gravel-6 | C16 |

**Table 1**. Image classes and their labels

### 3.1. Color Statistics-Based Features

Color images in the RGB (Red, Green and Blue) space are first transformed to those in the HAS (Hue, Saturation and Value) color space [6]. Features are then generated per pixel, by computing the mean ($M$) and the standard deviation ($SD$) with respect to each of the R, G, B, H, S and V color components, from a neighborhood of the pixel. The 12 resulting features are denoted by: $R_M$, $R_{SD}$, $G_M$, $G_{SD}$, $B_M$, $B_{SD}$, $H_M$, $H_{SD}$, $S_M$, $S_{SD}$, $V_M$, $V_{SD}$.

### 3.2. Local Histogram-Based Features

For each pixel, a number of color histogram-based features can be computed, with respect to each color component, given a fixed bin size and neighborhood [1]. Similarly, another set of local grey-level (GL) histogram features can be generated by first transforming color images to GAL ones. In this work, the bin size for computing color histogram features and that for GL histogram features are set to 8 and 16. The resulting color histogram-based features are denoted by $H_{Hi}$, $S_{Hi}$, and $V_{Hi}$, $i = 1, 2, ..., 8$, regarding the H, S, and V components, respectively. The GL features are denoted by $GL_{Hj}$, $j = 1, 2, ..., 16$. Two further GL statistic features, mean and STD, which are denoted by $GL_M$ and $GL_{SD}$, are also generated.

## 4. INFORMATION GAIN-BASED RANKING

Let $D_X$ be the value set of feature $X$ and $D_C$ be the label set of class variable $C$. The entropies of the class before and after observing $X$ are respectively defined by:

$$H(C) = -\sum_{c \in D_C} p(c) log_2 p(c)$$

$$H(C|X) = -\sum_{x \in D_X} p(x) \sum_{c \in D_C} p(c|x) log_2 p(c|x)$$

The amount by which the entropy of the class decreases after observing a certain feature reflects the additional information about the class that feature provides, and is called the information gain: $IG = H(C) - H(C|X)$. It measures how well a given feature separates data points with respect to their underlying class labels. Thus, all extracted features $X_k$, $k = 1, 2, ..., N$, can be ranked with regard to the $IG$ values of observing themselves: $IG_k = H(C) - H(C|X_k)$. Such ranking can be arranged in descending order, reflecting the fact that the higher an $IG$ value is, the more information the corresponding feature has to offer regarding the class. A subset of $M$ most informative features, $M \leq N$, can therefore be selected by choosing the first $M$ in the rank list.

## 5. SVM-BASED CLASSIFICATION

Support vector machines (SVMs) are used to perform image classification. Such a classifier seeks to find the optimal separating hyperplane among different classes by focusing on those training points (named support vectors), which are placed at the edge of the underlying feature vectors and whose removal would change the solution to be found. Radial Basis function (RBF) kernel is adopted in the SVMs. Here, the sequential minimal optimization algorithm of [7] is used to train the SVMs. Detailed SVM learning mechanism is omitted, but can be found in the literature (e.g. [11]).

In order to increase the efficacy of the SVM classifiers, IGR is used to rank the extracted features and to select those most informative during the training phase. This is of practical significance as for on-board application, classifiers are expected to be built with mature technologies (rather than totally new mechanisms that have limited experimental data). SVMs are proven high-performance classifiers, but they rely on quality input features. Adding SVMs with IGR-based feature selection helps to improve the quality of their input.

## 6. EXPERIMENTAL RESULTS

A set of 270 non-overlap images, of a size $512 \times 512$ each, subdivided from the large image of Fig. 1 (excluding regions that represent instruments and shadows) are used in this experimental investigation. In developing each classifier, a collection of 2870 pixel points are selected as training data, and another set of $4070$ points as testing data. Each point is labeled (by expert) with an identified class index, one of those 16 as given in Table 1, and is originally represented by a vector of 54 features (see Section 3). The size of a pixel's neighborhood used for generating feature vectors is set to $15 \times 15$. The SVM penalty parameter is set to 100, with standard Gaussian Radial Basis function (RBF) used. In the following comparative studies, the results of KNNs are first obtained with $K$ set to 1, 3, 5, 8, and 10, while for MLPs, only those of one hidden layer are considered, with the number of hidden nodes first set to 24, 28, 32, or 36. Then, those classifiers which have the highest accuracy, with respect to a certain number of $K$ or hidden nodes, are taken to facilitate fair comparison.

### 6.1. Comparison with the Use of Unreduced Feature Sets

For the given training data, the IGR method ranks the original $54$ features in the following descending order: $B_M$, $S_{SD}$, $GL_M$, $G_M$, $V_M$, $R_M$, $S_M$, $B_{SD}$, $G_{SD}$, $GL_{SD}$, $S_{H2}$, $S_{H5}$, $V_{H2}$, $S_{H4}$, $S_{H7}$, $R_{SD}$, $V_{SD}$, $S_{H6}$, $H_{H7}$, $H_{SD}$, $GL_{H4}$, $H_{H6}$, $GL_{H2}$, $GL_{H10}$, $V_{H3}$, $GL_{H3}$, $V_{H1}$, $GL_{H11}$, $H_M$, $V_{H5}$, $H_{H5}$, $GL_{H5}$, $H_{H8}$, $GL_{H9}$, $V_{H4}$, $GL_{H8}$, $V_{H7}$, $GL_{H12}$, $GL_{H6}$, $V_{H6}$, $GL_{H13}$, $GL_{H15}$, $H_{H7}$, $GL_{H7}$, $GL_{H14}$, $GL_{H1}$, $H_{H4}$, $V_{H8}$, $S_{H3}$, $H_{H6}$, $S_{H1}$, $GL_{H16}$, $S_{H8}$, $H_{H1}$. Fig. 3 shows the classification accuracy over the testing set, in relation to how many top-ranked features (by IGR) are used. The right-most case is the result of using all of the $54$ original features.

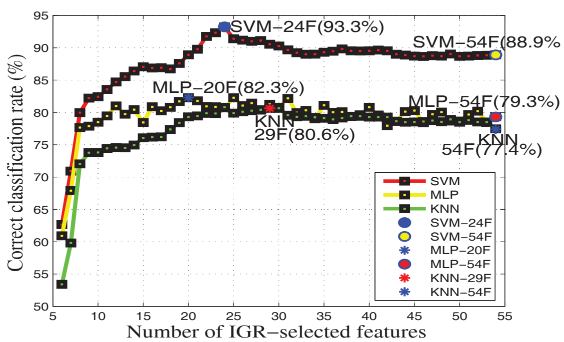


**Fig. 3**. Accuracy vs. number of IGR-selected features

These results demonstrate that all three types of classifier can have higher classification accuracy when IGR-selected features are used, than using the full set of (54) original features. This is generally true when the number of IGR-selected features is greater than 18 for both SVMs and KNNs, and 11 for MLPs. In particular, the SVM that uses 24 IGR-selected features performs the best with a classification rate of $93.3\%$ (as marked in Fig. 3). Comparatively, the best MLP and KNN which use 20 and 29 top-ranked features respectively, just reach a rate of $82.3\%$ and $80.6\%$. Also, the classification rate using full features is $88.9\%$ for SVM, $79.3\%$ for MLP, and $77.4\%$ for KNN. The employment of IGR not only reduces redundant feature measurements (thereby simplifying classification process), but also minimizes the noise associated with such measurements (thereby improving the classification accuracy) in SVM, MLP and KNN classifiers. The combined use of SVM and IGR techniques offers the best performance.

Based on the above results, the SVM which employs those 24 IGR-selected features (trained by the use of 2870 feature vectors) is taken to classify the entire Mars McMurdo image of Fig. 1 (again, excluding equipment and shadows). As an illustration, six classified and segmented image parts are shown in Fig. 5, where 16 different colors represent those 16 image types (see Fig 2). From these classified images, it can be seen that all image types vary in terms of their size, rotation, contrast, shape, and texture. For human eyes it can be very difficult to identify boundaries between many of such regions, such as those between types of sand gravel, between sand and sky, between mud and track sign, and those between rock classes. However, the classifier is able to perform under such circumstances (with respect to the ground truth painstakingly identified by domain experts).

### 6.2. Comparison with the Use of PCA-returned Features

As principal component analysis (PCA) [3] is arguably one of the most popular dimensionality reduction methods (although it was not initially designed to obtain discriminatory features), it is adopted here as the benchmark for comparison. Classifiers that are aided with IGR are systematically compared to those supported by the use of PCA. The results are summarized in Fig 4. In particular, of the same dimensionality per type of classifier (i.e. by the use of 24 features for SVMs, 20 for MLPs, and 29 for KNNs), the optimal classifiers which employ IGR-selected features have a substantially higher classification accuracy than those using PCA-returned features as listed in Table 2. Furthermore, the best performers that use PCA-returned features only reach a classification rate of $87.2\%$, $81.2\%$ and $79\%$, whilst requiring the use of 49, 37 and 45 features, respectively.

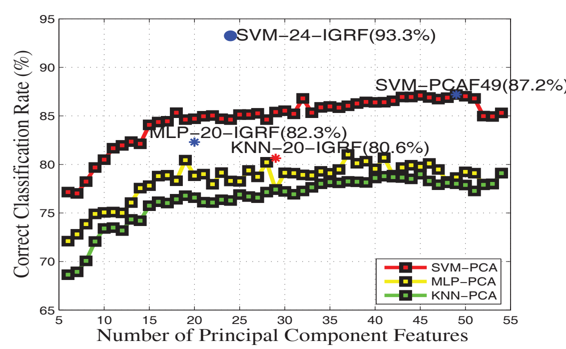


**Fig. 4**. IGR-selected vs. PCA-returned features

## 7. CONCLUSION

This paper has presented a study on Mars terrain image classification, using SVMs supported by information-gain based

| Method | SVM | MLP | KNN |
|---|---|---|---|
| IGR | 93.3% | 82.3% | 80.6% |
| PCA | 84.6% | 78.8% | 77.4% |

**Table 2**. Use of IGR-selected vs. PCA-returned features

feature selection. For the first time, these two techniques are integrated to help addressing challenging problems in space engineering where the real-world images are of many classes and of large-scale. The resultant SVM-based classifiers generally outperform MLP and KNN-based classifiers and those which use PCA-returned features. This is confirmed by systematic experimental investigations. The employment of IGR not only simplifies classification process, but also improves the classification accuracy. This work is, therefore, of significant potential for classification and analysis of real images on board in future Mars rover missions.

Interesting further research remains. This includes: comparing the present work with the use of alternative feature selection methods (e.g. fuzzy-rough set-based [10]), determining how the number of optimal IGR-selected features may vary with respect to different Martian terrain images, and incorporating IGR into the SVM formulation (instead of using it as a preprocessing tool for SVM classifiers).

## 8. REFERENCES


[1] O. Chapelle, P. Haffner, V. N. Vapnik. Support vector machines for histogram-based image classification. *IEEE Tran. on Neural Networks*. 10:1055-1064, 1999.

[2] H. Dunlop, D. Thompson and D. Wettergreen, Multi-scale features for detection and segmentation of rocks in Mars images. *IEEE Int. Conf. CVPR*. 1-7, 2007

[3] R. O. Duda, P. E. Hart and D. G. Stork. *Pattern Classification* (2nd edition), Wiley & Sons, New York, 2001.

[4] K. Huang and S. Aviyente. Wavelet feature selection for image classification. *IEEE Trans. Image Proc.*, 17:1709-1720, 2008.

[5] R. Jensen, Q. Shen. Are more features better? *IEEE Trans. Fuzzy Systems*, 17(6):1456-1458, 2009.

[6] D. R. Martin, C. C. Fowlkes, J. Malik. Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE Trans. PAMI*. 26:530-549, 2004.

[7] J. Platt. Macines using sequential minimal optimisation. in *Advances in kernel methods: support vector learning book contents*. 185 - 208, MIT Press Cambridge, MA, USA. 1999.

[8] http://marswatch.astro.cornell.edu/pancam_instrument/mcmurdo_v2.html.

[9] D. Iqbal and N. Lee, Efficient feature selection based on information gain criterion for face recognition, *Proc. Int. Conf. on Information Acquisition*. 523-527, 2007.

[10] C. Shang, D. Barnes and Q. Shen. Effective feature selection for Mars McMurdo image classification. *Proc. 9th Int. Conf. on Intelligent Systems Design and Applications*. 1419-1424, 2009.

[11] V. Vapnik. Statistical learning. New Yark, Wiley. 1998.


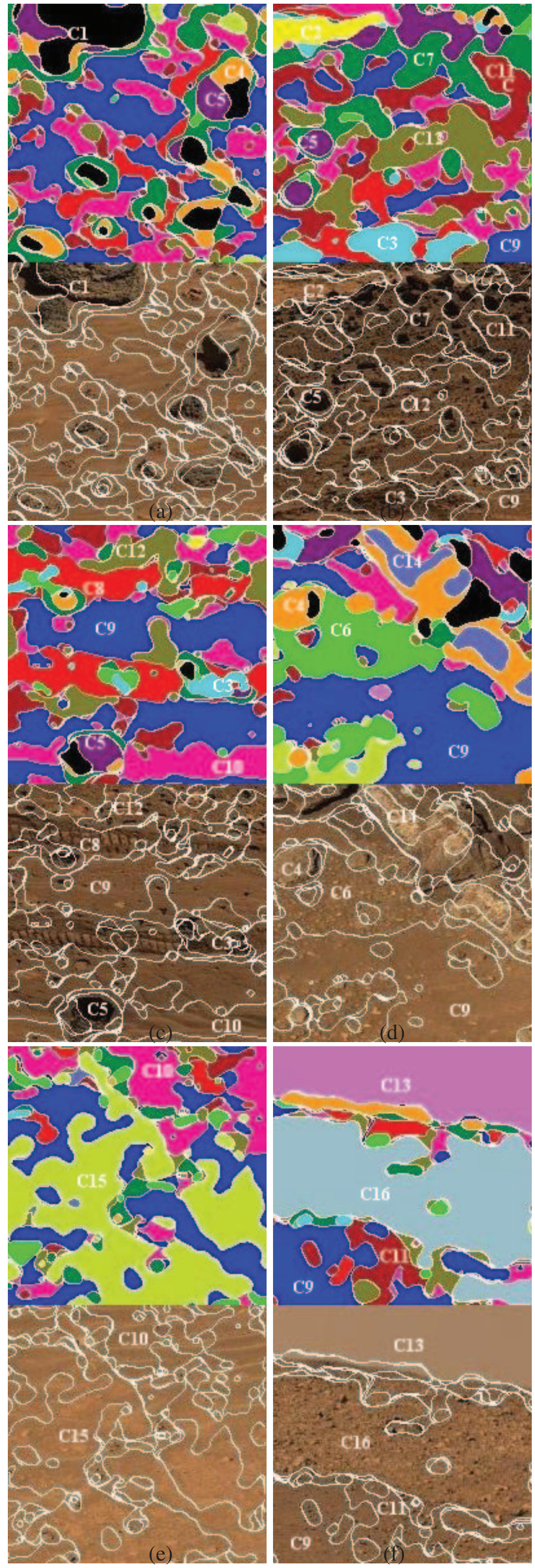


**Fig. 5**. Classified and segmented image