Available online at www.sciencedirect.com
ScienceDirect
Advances in Space Research 53 (2014) 1768–1782
Advances in Space Research
(a COSPAR publication)
www.elsevier.com/locate/asr

Crater detection via genetic search methods to reduce image features

Joseph Paul Cohen, Wei Ding *

Department of Computer Science, The University of Massachusetts Boston, 100 Morrissey Blvd., Boston, MA 02125-3393, USA

Available online 16 May 2013

Abstract

Recent approaches to crater detection have been inspired by face detection's use of gray-scale texture features. Using gray-scale texture features for supervised machine learning crater detection algorithms provides better classification of craters in planetary images than previous methods. When using Haar features it is typical to generate thousands of numerical values from each candidate crater image. This magnitude of image features to extract and consider can spell disaster when the application is an entire planetary surface. One solution is to reduce the number of features extracted and considered in order to increase accuracy as well as speed. Feature subset selection provides the operational classifiers with a concise and denoised set of features by reducing irrelevant and redundant features. Feature subset selection is known to be NP-hard. To provide an efficient suboptimal solution, four genetic algorithms are proposed to use greedy selection, weighted random selection, and simulated annealing to distinguish discriminate features from indiscriminate features. Inspired by analysis regarding the relationship between subset size and accuracy, a squeezing algorithm is presented to shrink the genetic algorithm's chromosome cardinality during the genetic iterations. A significant increase in the classification performance of a Bayesian classifier in crater detection using image texture features is observed.

Keywords: Shrinking feature set cardinality; Machine learning; Genetic algorithms; Crater detection; Bayesian classifier; Simulated annealing

1. Introduction

Craters are among the most studied geomorphic features in the Solar System because they yield important information about the past and present geological processes and provide the only tool for measuring relative ages of observed geologic formations. The work presented in this paper utilizes machine automation for crater detection.

A major challenge for planetary scientists is the magnitude of data available from missions to other planets such as Mars. With every mission to mars capturing higher and higher resolution imagery, the data to process will only grow. The High Resolution Stereo Camera HRSC (n.d.) camera captures 12.5 m per pixel and it's successor High Resolution Imaging Science Experiment HiRISE (n.d.) captures 50 cm per pixel. Automatic methods to process this data are not trivial.

A previous study by Tanaka (1986) concluded that the distribution of craters is exponential in relation to their size on Mars. This work aims to detect sub-kilometer craters and so must be able to deal with an exponential growth in complexity to analyze and draw significant new results on existing regions. This could result in billions of craters being identified on the surface of Mars during future planetary research.

If a scalable high-accuracy solution can be found it will enable many studies on large regions of planetary bodies including determining the geologically active regions of a planet, relatively dating sections of a planet, and determining both landing and exploration sites for interplanetary probes/machines.

Classifiers primarily operate on numerical instance vectors. The challenge is to create an effective mapping between crater detection and machine learning classification problems. The classifier is trained with instance vectors

* Corresponding author. Tel. +1 617 287 6428.
E-mail addresses: joecohen@cs.umb.edu (J.P. Cohen), ding@cs.umb.edu (W. Ding).

0273-1177/$36.00 © 2013 COSPAR. Published by Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.asr.2013.05.010

that are labeled and can then label instance vectors themselves. A trivial mapping into a standard machine learning problem would be to use sliding windows to generate candidates and have each pixel be an feature in the candidate. This is infeasible due to the large size and number of planetary images. This problem space must be reduced to something solvable in a reasonable amount of time.

Using techniques described by Bandeira et al. (2010) we can generate crater candidates using the highlight and shadow regions as well as the circular shape to narrow down the crater search space of a large image. This method of generating candidates is usable because it retains high recall. However it must be complemented with machine learning to increase the accuracy to a usable level. Haar feature masks were a breakthrough tool in face detection by Viola and Jones (2004) because of their high accuracy and ability to achieve the speed necessary for real time face detection. These candidates can be processed using Haar feature masks to generate instance vectors for a classifier.

This previous research has reduced crater detection to machine learning. There are still problems that inhibit this pipelines ability to achieve high accuracy. One impediment is the number of Haar feature masks that must be generated for high accuracy. Crater detection is a harder task than face detection. The distinction between faces and other objects is more apparent than cratered ground and noncratered ground. Craters have rims that are circular but when aged they appear as piles of rubble. While detecting craters, both fresh and aged craters must be detected with high accuracy. Rubble piles can be mistaken for rims of aged craters which makes it very hard to draw a solid line between crater and noncrater. The blur of this line is noise to the classifier that will both increase runtime with irrelevant crater features as well as reduce performance.

There are many crater features available in visible imagery from both simple and complex craters Pike (1980). For instance ejecta blankets surrounding craters provide a consistent frame for cropped crater imagery. Center uplifts and central peeks provide distinctive discrimination between unimpacked and overimpacked soil. The algorithm is designed to detect multi-anglular age-invariant impact craters with one single generic supervised algorithm that is robust enough for planetary scientists to automate their work.

The state of the art method of crater detection involves utilizing texture and contrast features of crater image candidates described by Ding et al. (2011) and Bandeira et al. (2011). This is achieved by extracting many numerical features from an image, each representing a particular texture or contrast, constructing a representative instance vector, and then applying machine learning to decide if potential crater images are in fact craters. Haar features, a gray-scale image texture features, are especially useful because of their ability to be calculated in near constant time using a data structure called integral images described by Viola and Jones (2004).

The challenge in using Haar features is that the number of Haar features can easily be tens of thousands. Many Haar features are redundant or even irrelevant. Ding et al. (2011) used AdaBoost to select crater image features and boost performance to between 79% and 90%. This paper aims to validate and improve those results.

Our work of machine learning applied to discriminative image features makes certain assumptions that are not considered when using AdaBoost. Each feature generated from the candidate crater images can be broken down into discriminate features and indiscriminate features. Discriminate features contain information that is useful during classification. Indiscriminate features provide no information to the classifier or misguiding information. Instead of weighting these features some negligible amount and continuing to extract them; they are simply removed.

To solve this new problem we want to find a subset of crater features that consists of only relevant features. By removing redundant and irrelevant features we both speed up classification and improve accuracy. There are two methods of feature selection we can use. Filter methods use domain knowledge to remove crater features without consideration of the class label. On the other hand, wrapper methods explicitly use the class label to validate the quality of a subset of features. Wrapper methods provide higher accuracy when finding subsets because they take the global classification rates into account. The downside to wrapper methods is that they have a $O(2^n)$ price tag to find the optimal solution. This computational cost can be mitigated using a variety of methods.

In this paper we present feature subset selection methods to find a high performing subset of features for a given classifier. Feature subset selection is known to be NP-hard. Exhaustive search is the only way to find the optimal subset of a set of features. To find, for certain, the optimal solution all permutations must be considered. The search space is 2^f where f is the number of features. For an example with only 58 features it would take 91 million years to compute all classifiers if a classifier took 0.10 s to create and evaluate.

2. Rational and approach

By the design of our feature extraction process, thousands of contrast image features can be extracted and many will be redundant or even irrelevant. A subset of these features can achieve higher accuracy when selected properly (Cohen et al., 2011; Liu et al., 2011). Our feature selection methods apply beyond the scope of crater detection and can apply to any wrapper based feature selection problem.

Once Haar features have been extracted from the candidate image we choose only those that are relevant and discriminative to our crater objective. This will not only increase the speed of the classifier but will increase the accuracy as well. Removing redundant and irrelevant features saves time during the feature extraction process because they do not have to be written to disk or sent over