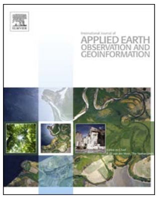
Int J Appl Earth Obs Geoinformation xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Int J Appl Earth Obs Geoinformation

journal homepage: www.elsevier.com/locate/jag



Comparing experts and novices in Martian surface feature change detection and identification

Jessica Wardlaw[a,*], James Sprinks[a], Robert Houghton[a], Jan-Peter Muller[b], Panagiotis Sidiropoulos[b], Steven Bamford[a], Stuart Marsh[a]

[a] University of Nottingham, Nottingham, UK
[b] Imaging Group, The Mullard Space Science Laboratory, University College London, UK

ARTICLE INFO

Keywords:
Crowd sourcing
Citizen science
Volunteered geographic information
Planetary science
Change detection
Image analysis

ABSTRACT

Change detection in satellite images is a key concern of the Earth Observation field for environmental and climate change monitoring. Satellite images also provide important clues to both the past and present surface conditions of other planets, which cannot be validated on the ground. With the volume of satellite imagery continuing to grow, the inadequacy of computerised solutions to manage and process imagery to the required professional standard is of critical concern. Whilst studies find the crowd sourcing approach suitable for the counting of impact craters in single images, images of higher resolution contain a much wider range of features, and the performance of novices in identifying more complex features and detecting change, remains unknown.

This paper presents a first step towards understanding whether novices can identify and annotate changes in different geomorphological features. A website was developed to enable visitors to flick between two images of the same location on Mars taken at different times and classify 1) if a surface feature changed and if so, 2) what feature had changed from a pre-defined list of six. Planetary scientists provided “expert” data against which classifications made by novices could be compared when the project subsequently went public.

Whilst no significant difference was found in images identified with surface changes by expert and novices, results exhibited differences in consensus within and between experts and novices when asked to classify the type of change. Experts demonstrated higher levels of agreement in classification of changes as dust devil tracks, slope streaks and impact craters than other features, whilst the consensus of novices was consistent across feature types; furthermore, the level of consensus amongst regardless of feature type. These trends are secondary to the low levels of consensus found, regardless of feature type or classifier expertise. These findings demand the attention of researchers who want to use crowd-sourcing for similar scientific purposes, particularly for the supervised training of computer algorithms, and inform the scope and design of future projects.

1. Introduction

Detection of change in satellite images of Earth and other planetary bodies is of significant scientific interest in the monitoring of environmental and climate change. Automating the detection of surface features over different spatial and temporal scales, however, remains complex and computationally expensive. Variation in the quality and coverage of images render them difficult for computers to process, in addition to the atmospheric and morphological influences on the “visibility” of features (Kim et al., 2005). Although we anticipate the development of increasingly subtle and powerful image processing and machine learning systems (Sidiropoulos and Muller, 2016), there remains a role for the human analyst particularly when variability is emphasised and human aptitudes of flexibility and judgement are called into play (e.g. interpreting rare events or features to make serendipitous discoveries). However, there is currently a clear, growing and profound imbalance between the number of expert observers and the sheer volume of satellite data available to the wider scientific community (See et al., 2016). One solution to this is to crowdsource analysis of imagery − a process often discussed within the realm of Citizen Science (Bonney et al., 2009). However, the viability of this solution rests on the fundamental question of whether a collection of suitably equipped amateurs can generate data of comparable quality to that produced by experts (Salk et al., 2016).

This paper investigates the potential power of novices to address two challenges that face the future application of a crowd-sourcing approach for the analysis of satellite imagery: detection of a wider range of surface features and changes in the appearance of these

---

[*] Corresponding author.

http://dx.doi.org/10.1016/j.jag.2017.05.014
Received 31 October 2016; Received in revised form 28 March 2017; Accepted 29 May 2017

Please cite this article as: Wardlaw, J., Int J Appl Earth Obs Geoinformation (2017), http://dx.doi.org/10.1016/j.jag.2017.05.014

features that reflect dynamic changes on the surface. Crowdsourcing has successfully classified surface features in Earth Observation, through calibration with ground truth (Zhao et al., 2014; See et al., 2016). The crowd is commonly used to count craters for estimating the age of lunar surfaces, a task which implicitly assumes that craters can be reliably identified, and further relies on measurements of crater diameter for age calculations (Robbins et al., 2014). In lunar images factors such as atmospheric distortion and the range of surface features are reduced so that the effects of human subjectivity can be isolated (Gault, 1970; Kirchoff et al., 2011). Robbins et al. (2014) investigated the consistency of expert classifications of craters in relation to terrain type, size and frequency, across different user interfaces. For all variables, only annotations of the smallest craters (< 10 pixels in diameter) were significantly different. They concluded "volunteers are approximately as good as experts in identifying craters…so long as enough volunteers examine the image to derive a robust result," with the caveat that accuracy for any single crater or cluster of craters is not important (Robbins et al., 2014; 126). Comparison of automated feature detection with the subjectivity introduced by humans has found differences between and within the classifications of individuals, for example on different days (Tar and Thacker, 2016). Successful cataloguing of geological landmarks could facilitate the filtering of imagery according to features of interest but the future utility of any automated process for this would require a significant human effort to label examples for training the algorithm (Wagstaff et al., 2012; Wagstaff et al., 2015). Whilst the work of Robbins suggests that novices can produce comparable annotations of impact craters to experts, their ability to identify other surface features of interest remains untested.

The present study extends previous work to detecting changes in images of the surface of Mars, in which features change at different rates, from rapidly moving dust devils, seasonal and inter-annual fluctuations of the polar ice caps and recurring slope lineae (indicating contemporary water activity), and slowly shifting sand dunes. Scientific interest in detecting changes in features such as impact craters (Kim et al., 2005; Bue and Stepinski, 2007; Li et al., 2015), gullies (Stepinski and Collier, 2004) and sand dunes (Bandeira et al., 2013) on Mars is high because changes reveal the evolution of the climate and geology of the planet; repeat image coverage for change detection is increasingly available and, until surface data can be validated with any certainty, alternative approaches are needed.

Although beyond the scope of this study, the introduction of human analysts, even within the context of the crowd-sourcing approach, brings into play other potential confounds on performance. Visual search is known to be affected by feature complexity (Lloyd and Hodgson, 2002) and size (Warner et al., 2015), scene context (Castelhano and Heaven, 2010), information density and presentation method (Chang et al., 2012), in addition to the human factors associated with performing visual search for a prolonged period of time (See, 2012). Change detection studies are also relevant in this context (Rensink, 2002), as well as those concerned with the quality of Volunteered Geographic Information (Haklay, 2010; Foody et al., 2013).

The ultimate goal of the on-going development of the algorithm is to achieve fully automated change detection and characterisation. Such a task is typically tackled with a supervised learning approach using a ground-truth dataset, but no publicly available ground-truth currently exists for planetary surfaces. The crowdsourcing this paper presents is thus intended to produce annotations for developing a fully automatic change detection algorithm. More information about the co-registration and the change detection algorithm can be found in Sidiropoulos and Muller (2016).

Section 2 now sets out the approach used to study these questions. Section 3 will present the consensus found within and between novices' and experts' classifications of change, and feature type that changed. Section 4 will discuss key findings and their implications for the remote sensing community, and designers of crowd-sourcing platforms for the classification of geomorphological features.

2. Method

2.1. Experimental design

To investigate novice performance in detecting 1) more complex features and 2) changes in features over time, this work presents the results of a Citizen Science project built with the project builder 'Panoptes' on Zooniverse.org and tested with experts and novices to directly their classifications of dynamic geological changes in Martian images, with a task designed for participants to compare two images of the same location but at different times (Bowyer et al., 2015).

The current interest in Martian exploration and the volume of images that have amassed since the planet was first imaged forty years ago represent an outstanding opportunity for the investigation presented. The images under study were processed from genuine images of the surface of Mars, so that participants would not anticipate what they would see. Prior to public release, doctoral Planetary Science students and post-docs classified images within a workshop at University College London's (UCL) Mullard Space Science Laboratory. Their exclusive access over the two days enabled separation of their "expert" classifications from those of volunteer "novices" over the following months.

2.2. Apparatus/materials

The study used images extracted from high-resolution image strips acquired by four orbital cameras described in Table 1.

First, the raw images were projected, or "co-registered", to a single coordinate system, to enable comparison. Since no high-resolution global datum exists for Mars, a mix of High-Resolution Stereo Camera (HRSC) Orthorectified Images (ORI) and Digital Terrain Models (DTMs), covering almost 50% of Mars, was selected for use as a baseline (Sidiropoulos and Muller, 2015). The co-registration technique was developed to achieve a fast and fully automatic co-registration of large volumes of data for generating an abundant input for change detection (Sidiropoulos and Muller, 2016). The subsequent set of co-registered images comprised of overlapping image pairs, which were then processed by an algorithm for detection of "regions-of-interest" (Sidiropoulos and Muller, 2016). The algorithm selected 868 regions-of-interest, each 512 × 512 pixels in size, as surface change candidates.

The change detection algorithm used is a "late fusion classification scheme" (Ye et al., 2012), and defines four types, or "classifiers", of change. Each classifier models a distinct type of surface change and produces a single, independent output in the form of a "confidence score" (Ye et al., 2012) from 0 to 1 for the probability of a positive classification, with 1 meaning 100% certainty that a pair of images

Table 1
Description of the cameras that took the images used in this study.

Camera	Spacecraft	Dates of Operation	Resolution	Reference
Context Camera (CTX)	Mars Reconnaissance Orbiter	2006-present	6m/pixel	Bell et al. (2013)
High-Resolution Stereo Camera (HRSC)	Mars Express	2004-present	12.5m/pixel	Jaumann et al. (2007)
Thermal Emission Imaging System (THEMIS)	Mars Odyssey	2002-present	17.5m/pixel	Christensen et al. (2004)
Mars Orbiter Camera – Narrow Angle (MOC–NA)	Mars Global Surveyor	1997–2006	1.5-12m/pixel	Malin et al. (2010)