Robotics and Autonomous Systems 62 (2014) 151–167

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

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

A survey on recent object detection techniques useful for monocular vision-based planetary terrain classification

Yang Gao *, Conrad Spiteri, Minh-Tri Pham, Said Al-Milli

Surrey Space Centre, University of Surrey, United Kingdom

HIGHLIGHTS

- The aim of this paper is to bridge the gap between monocular vision-based terrain classification and object detection in computer vision, by presenting a broad and structured overview of recent computer vision techniques behind the successes of object detection.
- We believe that these techniques provide great potentials for terrain classification using monocular vision processing.

ARTICLE INFO

Article history:
Received 11 October 2012
Received in revised form
3 November 2013
Accepted 22 November 2013
Available online 4 December 2013

Keywords:
Remote terrain classification
Autonomous navigation
Object detection
Monocular vision
Planetary rovers

ABSTRACT

Direct terrain classification from monocular images for autonomous navigation of planetary rovers is a relatively new and challenging research area, not only because of the hardware limitation of a rover, but also because the rocks and obstacles to be detected exhibit diverse morphologies and have no uniform properties to distinguish them from background soil. We present a survey of recently developed object detection techniques that can be useful for terrain classification for planetary rovers. We start with summarizing current vision-based terrain classification methods. We then provide a comprehensive and structured overview of recent object detection techniques, focusing on those applicable to terrain classification.

1. Introduction

The ability to perceive surrounding objects is an essential aspect of autonomous localization and navigation for planetary rovers, since the terrain's physical properties can strongly influence navigation abilities of the rovers.

Knowledge of nearby rocks and impassable obstacles allows a rover to adapt its control and plan traveling strategies to avoid collisions thus improving the mission's scientific return. Sloped and natural terrains typically affect a rover's mobility: the rover might easily traverse a region of packed soil, but become entrenched in loose drift material. For example, in 2005 and again in 2006, NASA's Mars Exploration Rover (MER) Opportunity became entrenched in loose drift material and was immobilized for several weeks. Recently, since May 2009, MER Spirit has become stuck in soft soil for months, despite many carefully analyzed attempts to free it from the soil. Fig. 1 presents some examples of the Martian terrains that exploration rovers have to handle.

Near-term scientific goals for Mars surface exploration are expected to focus on understanding the planet's climate history, surface geology, and potential for past or present life. To accomplish these goals, planetary rovers will be required to safely access rough terrain with a significant degree of autonomy. Planetary rovers suffer from intermittent communications, transmission latency, bandwidth constraints, and limited hardware resources. Onboard image understanding allows rovers to collect and interpret science imagery autonomously when human guidance is unavailable.

The analysis can significantly save bandwidth by selectively sending key summaries of data products for downlink. In addition, onboard analysis could recognize different terrain types, triggering opportunistic sensor measurements in response to novel morphology. The Mars Exploration Rovers have collected over 300,000 images for manual interpretation, but the relevant geologic analyses are labor intensive and to date only a small subset has undergone comprehensive study.

Vision-based terrain classification has been an attractive solution for perceiving nearby objects for planetary rovers. Visual data often provides information at a further range than other types of sensory data. Visual processing for terrain classification can arbitrarily be segregated into two categories: stereopsis processing and monocular processing. Stereopsis (or stereo vision) takes advantage of the binocular disparity between two rigidly mounted cameras pointing at the same scene, similar to the human visual arrangement. This disparity allows the vision processing mechanism to recover 3D point clouds from the scene. However,

---

* Corresponding author. Tel.: +44 1483 683446.
E-mail address: yang.gao@surrey.ac.uk (Y. Gao).

0921-8890/$ – see front matter Crown Copyright © 2013 Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.robot.2013.11.003

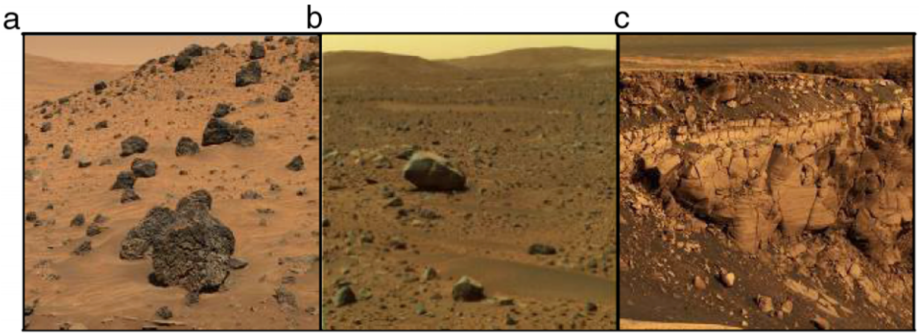

Fig. 1. Image samples of Martian terrain: (a) sloped sandy terrain with a scatter of large rocks (i.e., obstacles), (b) flat silt terrain with a wide variation of rock sizes, and (c) trenches leading to craters.

the point cloud recovery becomes less reliable as distance from the rover increases and the accuracy is a direct relation to the separation between optical axes of the two cameras (parallax) [1]. Stereo vision processing also involves a high computational cost [2]. This is especially true for stereo vision implementations using global algorithms [3], although dedicated hardware such as FPGAs and GPUs can be introduced to reduce the burden on the main processor.

Monocular vision, on the other hand, operates on a single camera. Processing a monocular image is more challenging than stereo vision processing because the depth information is not recoverable from one camera if no prior knowledge is provided. However, monocular vision processing does not necessarily require as much computational power as stereo vision processing and can detect objects at a much further distance than stereo vision processing. Being able to identify long-range obstacles and other terrain parameters however is a relatively new area of research, whose results will be of clear relevance for planetary explorations. Thus, monocular vision processing is a very promising, but difficult-to-use tool for vision-based terrain classification.

Recent advances in computer vision have shown many great successes in detecting objects from monocular images. For example, real-time frontal face detection has been generally considered as a solved problem [4]. The aim of this paper is to bridge the gap between monocular vision-based terrain classification and object detection in computer vision, by presenting a broad and structured overview of recent computer vision techniques behind the successes of object detection. We believe that these techniques provide great potentials for terrain classification using monocular vision processing.

The paper is organized as follows: Section 2 reviews current vision-based terrain classification methods used by Earth ground rovers and planetary rovers. In Section 3, we present a survey on techniques used in object detection that are applicable to monocular vision-based terrain classification and object detection. Finally, our conclusions are presented in Section 4.

2. Review of vision-based terrain classification

Terrain classification starts with the analysis of raw data from its sensors. Any kind of sensor is useful for the assessment of the terrain. Cameras (both monocular and stereo) are the most common sensors used for terrain classification, followed by LIDARs, often found in terrestrial applications. They provide direct information on terrain at mid-range to long-range distances. Motor encoders and other proprioceptive sensors such as vibration sensors, current sensors, and contact sensors provide indirect information on terrains at short-range or zero distance.

A wide variety of methods have been proposed for terrain classification, converting all ranges of distances, from zero distance to long-range distances. These include methods that work exclusively for a certain type of sensor or range, as well as those that fuse information from different types of sensors, and from different ranges. Survey papers on vibration-based terrain classification techniques can be found at [6,7]. Sancho-Pradel and Gao have summarized the majority of recent terrain classification methods in [5]. They group the terrain classification methods into five categories: geometric analysis, appearance analysis, soil analysis, and scientific interest analysis operating in real time while Semantic Mapping runs off-line at a ground station. Fig. 2 illustrates the functional configuration proposed by Sancho-Pradel and Gao for autonomous planetary robots, which incorporates a terrain assessment module that can combine multiple features (i.e. geometry, appearance, terramechanic parameters and scientific interest index) in a decentralized fashion. These features are extracted by a set of specialized blocks derived from existing techniques and can be extended to accommodate new approaches and sensors.

In this section, however, we focus on methods that work for vision-based terrain classification. Vibration-based techniques are often limited to zero or short-range distances. Geometric information coming from sensors like LIDARs or stereo vision work well for short-range to mid-range distances. However, the estimated 3D information (i.e.; 3D point clouds) becomes less reliable as the distance increases, and may not be sufficient to optimally navigate a rover over long distances. At long-range distances, 2D images are probably the only reliable sensory input available. Image cues can provide valuable information to complement geometric methods in the analysis of a scene. Although they contain rich information for terrain classification, they are harder to process than other types of sensory input. Low level classifiers refer to classification methods that rely on a single visual feature such as color, texture, and geometric properties [8]. In unstructured environments such as Mars, none of these features are sufficient on their own for robust classification of the terrain. Color-based classifiers suffer from cases of shadowing and reflectance and in a Martian-like environment, the color variation is narrowed due to the lack of moisture and almost homogeneous dust covering the planet. Image texture is a measure of special variation in intensity and is often used to detect structural, roughness or regularity differences in an image. Texture feature classifiers suffer from range sensitivity and from significant intra-class variability. Geometric-based classifiers suffer from the lack of regularity of objects found within a natural environment setting. In this paper, we divide vision-based terrain classification methods into two groups, those working at short-range and mid-range distances, and those working at long-range distances (see Fig. 3 for examples).

2.1. Short range and mid-range

A variety of terrain types is likely to be encountered thus preemptive planning based on input data is desirable [9,10]. The analysis of visual information (e.g. color and texture) from small terrain patches often provides a good balance between sensor complexity and terrain classification accuracy.