Field Testing of an Integrated Surface/Subsurface Modeling Technique for Planetary Exploration

Paul Furgale[1], Timothy D Barfoot[1], Nadeem Ghafoor[2], Kevin Williams[3] and Gordon Osinski[4]

Abstract

While there has been much interest in developing ground-penetrating radar (GPR) technology for rover-based planetary exploration, relatively little work has been done on the data collection process. Starting from the manual method, we fully automate GPR data collection using only sensors typically found on a rover. Further, we produce two novel data products: (1) a three-dimensional, photorealistic surface model coupled with a ribbon of GPR data, and (2) a two-dimensional, topography-corrected GPR radargram with the surface topography plotted above. Each result is derived from only the onboard sensors of the rover, as would be required in a planetary exploration setting. These techniques were tested using data collected in a Mars analogue environment on Devon Island in the Canadian High Arctic. GPR transects were gathered over polygonal patterned ground similar to that seen on Mars by the Phoenix Lander. Using the techniques developed here, scientists may remotely explore the interaction of the surface topography and subsurface structure as if they were on site.

Keywords
Visual odometry, ground-penetrating radar, computer vision, robotic mapping

1. Introduction

The use of ground-penetrating radar (GPR), together with a stereo camera, on planetary exploration rovers has been proposed several times (Barfoot et al. 2003; Fong et al. 2008) and is now in development for the European Space Agency's (ESA) ExoMars project (Vago et al. 2006). Used together, surface and subsurface imaging will aid in the search for liquid water and evidence of life. The ESA mission proposes using the stereo camera for site selection and survey, while the GPR will then be used to characterize the subsurface stratigraphy and to select sites for drilling.

Despite this interest, there are still several open issues regarding the use of GPR on a rover platform:

1. Rovers must be able to deliver information about the surface (topography, substrate particle size distribution, and/or the presence of any existing outcrops) that enables the operator to give local geologic context to the subsurface data. Surface and subsurface information must be coupled (i.e. the location of the GPR traverse must be known with respect to the surface data captured by the rover) so that the scientific interpretation of the data is as close as possible to a direct (human) site survey.
2. For a more complete interpretation of GPR data, the radargram (i.e. the two-dimensional subsurface profile)

should be corrected for topography (e.g. Busby and Merritt (1999); Lunt and Bridge (2004); Annan (2009) and Cassidy (2009)). As planetary exploration rovers have no access to a global positioning system (GPS) equivalent, topographic profiles must be generated using other onboard sensors.
3. A flight-ready GPR antenna must satisfy size, mass, and power consumption constraints and the integration must minimize interference from the rover's metal chassis.

The contribution of this paper is to address points 1 and 2 above by using stereo imagery to enhance GPR data. Specifically, we propose an architecture that enables the automated construction of coupled surface/subsurface models using the specific sensors planned for deployment on near-term robotic exploration missions. Stereo cameras have been deployed on the Mars Exploration Rovers (Maki et al. 2003) and are planned for both the Mars Science

[1] University of Toronto, Institute of Aerospace Studies, Toronto, Canada
[2] MDA Space Missions, Brampton, Canada
[3] Buffalo State College, Department of Earth Sciences, Buffalo, NY, USA
[4] University of Western Ontario, Departments of Earth Science, Physics and Astronomy, London, Canada

Corresponding author:
Paul Furgale, University of Toronto, Institute for Aerospace Studies, 4925 Dufferin Street, Toronto, Ontario, Canada M3H 5T6
Email: paul.furgale@utoronto.ca

Laboratory (Malin et al. 2005) and the ExoMars Mission (Vago et al. 2006). The modeling pipeline uses visual odometry (VO) (Barfoot 2005; Nistér et al. 2006; Konolige et al. 2007; Maimone et al. 2007; Howard 2008)—full 6-degree-of-freedom motion estimation using a stereo camera as the primary sensor—to fully automate the GPR data collection procedure and produce two novel data products: (1) a three-dimensional, photorealistic surface model coupled with a ribbon of GPR data, and (2) a two-dimensional, topography-corrected GPR radargram with the surface topography plotted above. Each result is derived from only the onboard sensors of the rover, as would be required in a planetary exploration setting.

On Earth, producing a site survey using GPR on rough terrain involves several steps:

1. The GPR antenna is dragged along a transect to collect many GPR traces.
2. The position of the GPR at each trace is determined using traditional surveying techniques or differential global positioning (DGPS) integrated directly with the GPR.
3. GPR processing software is used to correct the horizontal spacing and vertical offset of the GPR traces along the transect.
4. The corrected traces are concatenated into a raster image called a *radargram*.

Our approach automates this process by using VO in the place of GPS, thereby enabling the technology to be used in a planetary exploration context.

These techniques have been tested using data gathered at two sites near the Haughton-Mars Project Research Station (HMP-RS) on Devon Island, Nunavut, Canada (Lee et al. 2007). The sites exhibit polygonally-patterned ground, a periglacial landform often indicative of subsurface ice deposits (Mackay 1999). Figure 1 shows the polygonal terrain seen by the Phoenix lander and at one of our test sites.

Stereo images were captured during GPR transects and our integrated surface/subsurface modeling techniques were applied to the resulting data. Our work here is part of a larger project to develop a mission concept to sample ground ice in Martian polygonal terrain (Barfoot et al. 2009).

The rest of the paper is organized as follows. Related work regarding VO, terrain mapping, and the development of GPR for planetary exploration is discussed in Section 2. Our integrated surface/subsurface modeling technique is described in Section 3. We describe our VO algorithm (3.1), our three-dimensional surface/subsurface modeling technique (3.2) and our method of two-dimensional radargram correction (3.3). Sections 4 and 5 outline our field tests on Devon Island and the associated results. Our conclusions are provided in Section 6. A video outlining the project is available as Extension 1. It includes footage of the manual data collection process, the raw data used in our experiments, and renders of the resulting data products.

2. Related Work

The architecture proposed in this paper draws on previous work in the areas of VO, terrain modeling, and ground-penetrating radar. The related work will be reviewed separately for each of these areas.

2.1. Visual Odometry

The VO problem has some relation to the simultaneous localization and mapping (SLAM) problem. In both cases, a robot is navigating in an unknown environment and trying to localize with respect to a fixed reference frame. However, VO systems have no requirement to return a map, and this has led to different solution formulations and testing paradigms.

Early work estimating motion from image sequences may be divided into two categories: (1) optical flow algorithms and (2) feature tracking algorithms (Aggarwal and Nandhakumar 1988). While there is some recent work recovering motion (or structure and motion) using optical flow (Corke et al. 2004; Benoit and Ferrie 2007; Comport et al. 2010), the techniques are not widely used in unstructured environments. This is due to the difficulty of jointly estimating the scene structure and camera motion without finding explicit correspondences or making assumptions about the nature of the scene (e.g., planarity). The planetary exploration context requires full six-degree-of-freedom motion estimation as planar camera motion cannot be guaranteed in unstructured terrain. Hence we have chosen to estimate motion by tracking a sparse set of feature points through the image sequence.

The scale of the motion can not be recovered solely from monocular sequences (Hartley and Zisserman 2000; Nistér 2003) so metric odometry using a monocular camera is only possible using some other measurement of scale, such as an inertial measurement unit (Strelow and Singh 2004) or wheel odometry. A stereo camera provides metric structure within each pair of images and this simplifies the motion estimation problem. The basic sketch of stereo VO from sparse feature tracks was detailed in Moravec (1980). His work included all of the major processing blocks still used for VO today: image capture and de-warping, feature tracking, outlier detection, and a non-linear numerical solution for the pose change.

There are two basic categories of feature tracking pipelines used in VO algorithms. We will call them *correlation* pipelines and *descriptor* pipelines. Correlation pipelines start by finding point features in one image using an algorithm such as Harris corners (Harris and Stephens 1988) or the Förstner operator (Förstner and Gulch 1987). A small patch of the image around the point feature is used as a template for correlation-based matching in another image (Maimone et al. 2007; Matthies et al. 2007). Descriptor pipelines use algorithms such as the scale invariant feature transform (SIFT) (Lowe 2004). These algorithms detect regions of interest over different scales in the image and
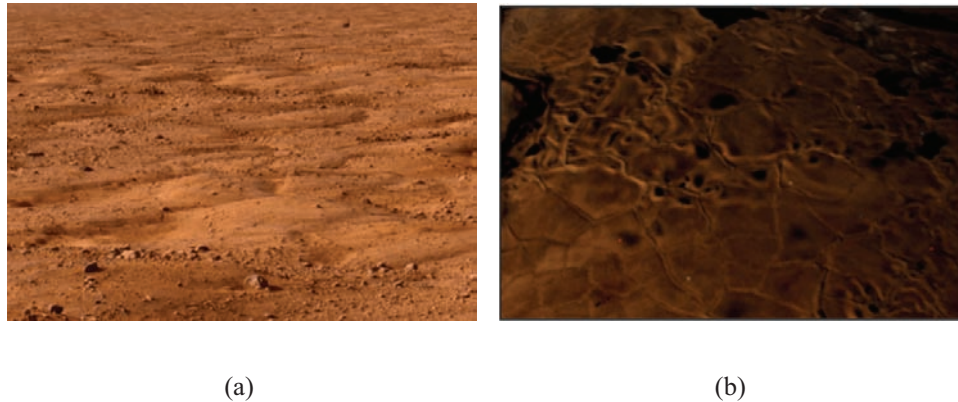
(a)   (b)

Fig. 1. Polygonal terrain is a common periglacial landform often indicative of subsurface ice deposits, seen here (a) from the Phoenix Lander on Mars (courtesy of NASA/JPL-Caltech/University of Arizona/Texas A&M University), and (b) at our test site on Devon Island, in the Canadian High Arctic.

compute a D-dimensional (usually $D = 64$ or $D = 128$) descriptor that accounts for scale and rotation changes. Matches are then found by looking for the nearest neighbors in descriptor space. Correlation pipelines are generally considered more computationally efficient but they are sensitive to viewpoint changes. This limits the amount a robot may move between images before the pipeline begins to fail (Olson et al. 2003). Descriptor pipelines are more robust to viewpoint changes but require more computation, and, in the context of near-term planetary exploration, a correlation pipeline would almost certainly be used. Our vision pipeline uses the speeded-up robust feature (SURF) detector and descriptor (Bay et al. 2006) because of its capability for robust matching over wide viewpoint changes. This capability was not necessary for this work but was developed to enable other projects using the same vision pipeline (Furgale and Barfoot 2010a,b).

Feature tracks inevitably contain mismatches and it is important to find them and remove them to get the best motion estimate. Many approaches to outlier detection have been developed in the computer vision community, mostly to address the problem of estimating the fundamental matrix relating a pair of images. The random sample consensus algorithm (RANSAC) (Fischler and Bolles 1981) approaches this problem by repeatedly (1) generating a model from a randomly-selected minimal set of data, and (2) scoring the model by counting the number of data points with error below a fixed threshold. Many improvements to the original RANSAC algorithm have been proposed. For example, MLESAC (Torr and Zisserman 2000), scores models using a squared-error cost function, and MAPSAC (Torr 2002) includes a step to choose the best model parameters. Motion estimation (or structure and motion estimation) using a calibrated stereo rig is a significantly simpler task than fundamental matrix estimation because there is no ambiguity as to which model parameters to choose. The intrinsic camera parameters are obtained through calibration, and the known baseline between the cameras resolves any scale ambiguities. As such, many VO implementations adopt the original RANSAC algorithm (Konolige et al. 2007; Maimone et al. 2007).

VO sequences can contain thousands of stereo images and the pose solution method must scale to trajectories of this size. Our current algorithm computes a frame-to-frame solution in the style of Matthies and Shafer (1987), who compound the incremental estimates into a trajectory. Other solutions to this problem have been based on SLAM research (Thrun et al. 2001; Barfoot 2005) or sparse bundle adjustment (SBA) (Triggs et al. 2000; Sünderhauf et al. 2006; Konolige et al. 2007; Sibley et al. 2008), which jointly estimates the camera trajectory and feature locations. Solution methods based on SBA are more accurate than frame-to-frame methods (Konolige et al. 2007) as they closely approximate the full batch maximum likelihood solution for structure and motion (Sibley et al. 2008). However, we have chosen to work with an algorithm similar to the one described by Maimone et al. (2007) as the computational complexity and memory usage patterns are more suitable to the space exploration context.

The VO problem has no requirement to close loops or return a map. As such, the most common error metric is the position error as a percentage of distance traveled. Most VO algorithms have been tested over distances less than 400 m. Barfoot (2005) reports 4% error over 38.8 m on sand and 0.5% error over 117.6 m on gravel. Maimone et al. (2007) quote 1.5% to 2.5% error over runs less than 30 m. Mallet et al. (2000) report 4% error over 25 m. Nistér et al. (2004) plot many runs of their algorithms against GPS but do not quantify the position errors. The evaluation performed by Konolige et al. (2007) is a notable exception. They have tested their algorithm on several multi-kilometer datasets, achieving 1.5% error over 10 km. When fusing the VO estimate with data from an inertial measurement unit, they report an error of 0.08% of distance traveled. This is clearly the most impressive test of VO published to date and it sets a precedent for future work in the field.

2.2. Terrain Modeling

Terrain modeling algorithms for planetary rovers have been developed for use in both onboard and offboard systems. Onboard systems use terrain models for autonomous rover behavior such as path planning and localization (Lacroix et al. 2001). Offboard systems generate immersive three-dimensional environments that human operators can use for operational and science planning (Wright et al. 2005). The two purposes have different requirements that guide the technology development. Systems generally differ along three lines: the data acquisition method, the data alignment method, and the data representation.

Terrain modeling starts with the acquisition of three-dimensional sets of points. Point sets can be acquired through dense-stereo processing, or using active ranging sensors. Stereo image processing algorithms use differences between the two images in a stereo pair to solve for the depth at every pixel. The depth information is more accurate for surfaces closer to the camera and algorithms may fail if the images lack sufficient texture (Matthies et al. 2007). Nevertheless, we use stereo cameras as they are cheap, reliable, and low-power. Terrain modeling systems based on stereo cameras have been successfully employed on rover platforms for both onboard and offboard terrain modeling (Zhang 1997; Biesiadecki and Maimone 2006; Olson et al. 2007). Active ranging devices have high power requirements and may be heavy and expensive. However, unlike stereo processing, they achieve nearly constant range-measurement accuracy (Blais 2004). Active devices based on light detection and ranging technology have made their way onto a number of recent rover prototypes including those of Bakambu et al. (2006) at the Canadian Space Agency and Fong et al. (2008) at NASA Ames.

If point sets have been acquired at multiple times, or from multiple sensors, they must be transformed into a common coordinate frame. When the rover is stationary during acquisition, the scans may be aligned using calibration data or from motorized pan-tilt units (Wright et al. 2005). Point clouds with unknown relative transformations and unknown correspondences between points can be aligned using the iterative closest point algorithm (Zhang 1997; Baumgartner et al. 1998). When a stereo camera is used for data acquisition, photogrammetric approaches become possible. Feature tracking algorithms generate a small number of points with known correspondences. If there are a small number of poses then this problem may be solved efficiently using sparse bundle adjustment (Triggs et al. 2000). As the number of poses grows, solving for all poses at once becomes computationally too expensive. This is why we have adopted the use of VO techniques to generate the pose estimates which are used to align the point sets (Barfoot et al. 2006; Se and Jasiobedzki 2008).

Aligned point sets must then be converted into a format useful for the application. NASA Ames has developed a multi-resolution format that uses octrees to combine point sets from multiple sources (Wright et al. 2005).

However, raw point data is not necessarily the most useful format for processing or display. Height maps or digital elevation models represent nominally planar surfaces as a two-dimensional grid of height values. This is an efficient representation and may be used for autonomous planning (Lacroix et al. 2001; Goldberg et al. 2002), but it is unable to represent complex three-dimensional spaces. Point sets can be converted into polygonal meshes and mapped with texture from the raw images to create a photorealistic surface. Meshes are able to represent complex surfaces and may be viewed interactively on commodity hardware. Consequently, this is the representation of choice for offboard systems such as ours (Pedersen et al. 2005; Barfoot et al. 2006; Fong et al. 2008; Se and Jasiobedzki 2008).

2.3. Ground-penetrating Radar

Perhaps the most common geophysical tool used in terrestrial geological investigations, GPR, relies on differences in subsurface materials' dielectric permittivity to determine underlying stratigraphy. The GPR transmitter emits a high-energy electromagnetic pulse into the ground at frequencies generally in the range of 10-1000 MHz (Degenhardt and Giardino 2003). When the signal encounters an interface between layers of differing permittivity, part of the energy is reflected back towards the surface while the remainder is refracted into the subjacent medium. The reflection/refraction process continues until the signal has attenuated completely or the user-defined time window—the amount of time that the GPR receiver is programmed to search for a return signal—has elapsed (Moorman et al. 2003). Based on the two-way travel time of each reflected pulse, a trace is produced illustrating a series of reflector intensities located beneath the unit, whereby the amplitude of the reflection is proportional to the relative difference in permittivity between adjacent materials (Arcone et al. 1995). When the GPR survey is conducted along a surface transect, individual traces can be combined to produce a radargram, a two-dimensional profile showing continuous subsurface reflective layers, which allows for enhanced stratigraphic interpretation.

The imaging of the subsurface of Mars has, to date, been performed by orbital sounding radars—The Mars Advanced Radar for Subsurface and Ionosphere Sounding (MARSIS) on Mars Express (Picardi et al. 2005), and the Mars Shallow Radar (SHARAD) on Mars Reconnaissance Orbiter (Seu et al. 2007). These instruments have revealed subsurface information to kilometers and hundreds of meters depths, respectively, at resolutions that reveal large-scale features such as buried impact craters, 250 km across (Picardi et al. 2005). Although the scans returned by SHARAD have an order of magnitude better depth resolution than those collected by MARSIS (Seu et al. 2007), it is still too coarse to pinpoint sampling locations for surface-rover drilling operations. As such, GPR deployed on a surface rover is a logical complement to the currently available orbital data.
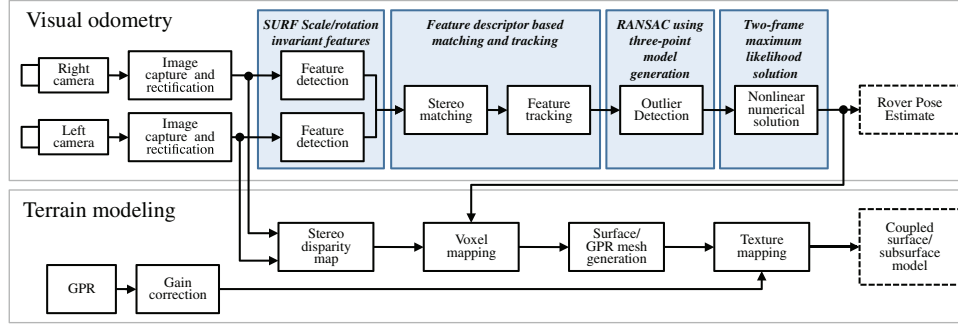
Fig. 2. An overview of the major processing blocks of our system. A VO pipeline (top) processes stereo image pairs to solve for the pose of the camera at every image. The pose estimate is used by the modeling pipeline (bottom) to create a three-dimensional coupled surface/subsurface model.

The application of GPR to frozen terrain was pioneered by Annan and Davis (1976) (cf. Ross et al. (2005)) and is becoming increasingly widespread. Given its established utility in some of the Earth's most extreme environments such as Antarctica (Arcone et al. 2002) and the Canadian Arctic (dePascale et al. 2008), rover-based GPR has thus been proposed for development on a variety of planetary missions and will be included on ESA's upcoming ExoMars mission (Vago et al. 2006). While previous studies have focused primarily on hardware development and testing (Leuschen et al. 2002; Grant et al. 2003; Kim et al. 2006), understanding the physics of dielectric signal loss in Mars-type substrates (Pettinelli et al. 2007), and possible applications to Mars analogue environments (Arcone et al. 2002; Degenhardt and Giardino 2003; Williams et al. 2005), little effort has been directed towards the integration of subsurface data with surface-based distance estimates (Barfoot et al. 2003). Because the interpretation of GPR data relies heavily on knowledge of the survey's spatial orientation and relation to surface features, it is important that any autonomous GPR study have such a capability. Previous advances automating GPR data acquisition have used DGPS (Rial et al. 2005; Fong et al. 2008), a laser theodolite (Lehmann and Green 1999), or proprioception (Freese et al. 2007) to track the pose of the GPR during data collection. To the best of our knowledge, this work is the first to use VO-derived motion estimates to align GPR data.

3. Integrated Surface/Subsurface Modeling

This section will describe our integrated surface/subsurface modeling system. Data flow through the main processing blocks of our system can be seen in Figure 2. The images captured from a calibrated stereo camera are first undistorted and rectified. This process accounts for lens distortion and aligns the images as if they came from perfect pinhole cameras with parallel optical axes. The images are then used to create the motion estimate, which is the enabling technology for the rest of the system. Using the

rover's pose estimate, we are able to create and couple the surface and subsurface models.

3.1. Visual Odometry

We define $\underset{\rightarrow}{\mathcal{F}}_j$ to be a coordinate frame attached to the left camera of a stereo pair at time j. The attitude of the rover at this time may be described by $\mathbf{C}_{1,j}$, the rotation matrix that transforms vectors in $\underset{\rightarrow}{\mathcal{F}}_j$ to $\underset{\rightarrow}{\mathcal{F}}_1$. Similarly, we define the rover's position as $\boldsymbol{\rho}_1^{j,\,1}$, a vector from the origin of $\underset{\rightarrow}{\mathcal{F}}_1$ to the origin of $\underset{\rightarrow}{\mathcal{F}}_j$ (as indicated by the superscript) expressed in $\underset{\rightarrow}{\mathcal{F}}_1$ (as indicated by the subscript). The goal of our VO algorithm is to estimate $\{\mathbf{C}_{1,j}, \boldsymbol{\rho}_1^{j,\,1}\}$ for each pose $j = 2 \ldots M$. The maximum-likelihood formulation below is based on the work of Matthies (1989), which eventually became the VO algorithm deployed on the Mars Exploration Rovers (Maimone et al. 2007), and will continue to be used on the Mars Science Laboratory (Johnson et al. 2008). However, our approach to linearization uses the small angle approximation at the matrix level, avoiding the need to compute Jacobians at the scalar level.

3.1.1. Observation Model Let $\bar{p}_j^{i,\,j} = [x \; y \; z]^{\mathrm{T}}$ be the three-dimensional location of feature i with respect to and expressed in $\underset{\rightarrow}{\mathcal{F}}_j$. Our observation model, $\mathbf{h}(\cdot)$, projects $\bar{\mathbf{p}}_j^{i,\,j}$ into the rectified images of an axis-aligned stereo camera

$$\bar{\mathbf{y}}_j^i := \mathbf{h}(\bar{\mathbf{p}}_j^{i,\,j}) = \begin{bmatrix} u_l \\ v_l \\ u_r \\ v_r \end{bmatrix} = \frac{1}{z} \begin{bmatrix} f_u x \\ f_v y \\ f_u(x-b) \\ f_v y \end{bmatrix} \quad (1)$$

where $\bar{\mathbf{y}}_j^i$ is the noise-free observation, f_u and f_v are the horizontal and vertical focal lengths (in pixels), and b is the camera baseline (in meters). The projected coordinates in the left and right images are (u_l, v_l) and (u_r, v_r) respectively. We model noise on the observations, $\delta \mathbf{y}_j^i$, as drawn from a zero-mean Gaussian density with covariance $\mathbf{Y}_j^i$

$$\delta \mathbf{y}_j^i \sim \mathcal{N}(\mathbf{0}, \mathbf{Y}_j^i).$$

Hence, we define $\mathbf{y}_j^i$, the observation of feature i at time j, as

$$\mathbf{y}_j^i := \mathbf{h}(\bar{\mathbf{p}}_j^{i,j}) + \delta\mathbf{y}_j^i$$
$$= \bar{\mathbf{y}}_j^i + \delta\mathbf{y}_j^i. \qquad (2)$$

The assumption of zero-mean noise is a strong one, but it is standard practice in sparse-stereo processing (Konolige et al. 2007; Maimone et al. 2007).

3.1.2. Inverse Observation Model Because we are using a stereo camera, the observation model (1) is invertible

$$\bar{\mathbf{p}}_j^{i,j} = \mathbf{g}(\bar{\mathbf{y}}_j^i) = \mathbf{h}^{-1}(\bar{\mathbf{y}}_j^i) = \frac{b}{u_l - u_r} \begin{bmatrix} u_l \\ (f_u/f_v)\,v_l \\ f_u \end{bmatrix} \qquad (3)$$

where $\mathbf{g}(\cdot)$ is the inverse observation model. Given an observation of a point feature in a stereo pair of images, (3) may be used to triangulate the position of the feature. We may model how the noise in an observation (2) affects the triangulation equation. If we assume $\delta\mathbf{y}_j^i$ is small, we can approximate the perturbation by a first-order Taylor expansion

$$\mathbf{g}(\bar{\mathbf{y}}_j^i + \delta\mathbf{y}_j^i) \approx \mathbf{g}(\bar{\mathbf{y}}_j^i) + \left.\frac{\partial\mathbf{g}}{\partial\mathbf{y}_j^i}\right|_{\bar{\mathbf{y}}_j^i} \delta\mathbf{y}_j^i$$
$$= \bar{\mathbf{p}}_j^{i,j} + \mathbf{G}_j^i \delta\mathbf{y}_j^i$$
$$= \bar{\mathbf{p}}_j^{i,j} + \delta\mathbf{p}_j^{i,j}$$
$$=: \mathbf{p}_j^{i,j}$$

where we have defined $\mathbf{G}_j^i := \left.\frac{\partial\mathbf{g}}{\partial\mathbf{y}_j^i}\right|_{\bar{\mathbf{y}}_j^i}$ and $\delta\mathbf{p}_j^{i,j} := \mathbf{G}_j^i \delta\mathbf{y}_j^i$. The mean of $\delta\mathbf{p}_j^{i,j}$ is

$$E\left(\delta\mathbf{p}_j^{i,j}\right) = E\left(\mathbf{G}_j^i \delta\mathbf{y}_j^i\right) = \mathbf{G}_j^i E\left(\delta\mathbf{y}_j^i\right) = \mathbf{0}$$

where $E(\cdot)$ is the expectation operator. We define $\mathbf{P}_j^i$ to be the covariance of $\delta\mathbf{p}_j^{i,j}$

$$\mathbf{P}_j^i := E\left(\delta\mathbf{p}_j^{i,j}\delta\mathbf{p}_j^{i,j\mathrm{T}}\right) = E\left(\mathbf{G}_j^i \delta\mathbf{y}_j^i \delta\mathbf{y}_j^{i\mathrm{T}} \mathbf{G}_j^{i\mathrm{T}}\right)$$
$$= \mathbf{G}_j^i E\left(\delta\mathbf{y}_j^i \delta\mathbf{y}_j^{i\mathrm{T}}\right) \mathbf{G}_j^{i\mathrm{T}} = \mathbf{G}_j^i \, \mathbf{Y}_j^i \mathbf{G}_j^{i\mathrm{T}}.$$

3.1.3. Maximum Likelihood Formulation Given a set of N features tracked from time j to time $j+1$, our goal is to estimate the maximum likelihood rotation and translation between $\underrightarrow{\mathcal{F}}_j$ and $\underrightarrow{\mathcal{F}}_{j+1}$. In formulating the error function, we will refer to the quantities we want to estimate, $\{\mathbf{C}_{j,j+1}, \boldsymbol{\rho}_j^{j+1,j}\}$, as the *design variables*.

For a single observation, we may define an error, $\mathbf{e}_i$, as the discrepancy between $\mathbf{y}_j^i$ and the expected observation

$$\mathbf{e}_i := \mathbf{y}_j^i - \mathbf{h}\left(\mathbf{p}_j^{i,j}\right). \qquad (4)$$

However, this approach requires an estimate of $\mathbf{p}_j^{i,j}$, the three-dimensional feature location. Estimating the feature locations increases the number of design variables from 6 to $6+3N$, where N can be anywhere from 50 to 500 depending on the nature of the scene and the distance between images. The VO problem has no requirement to return a map, so eliminating the feature locations from the estimation formulation is very desirable. This is easy to accomplish as we are considering only frame-to-frame correspondences.

Note that we have exactly two observations of each feature i: $\mathbf{y}_j^i$ and $\mathbf{y}_{j+1}^i$. Using our inverse measurement model and the design variables, $\{\mathbf{C}_{j,j+1}, \boldsymbol{\rho}_j^{j+1,j}\}$, we can build $\hat{\mathbf{p}}_j^{i,j}$, an estimate of $\mathbf{p}_j^{i,j}$

$$\hat{\mathbf{p}}_j^{i,j} := \mathbf{C}_{j,j+1}\mathbf{g}\left(\mathbf{y}_{j+1}^i\right) + \boldsymbol{\rho}_j^{j+1,j}$$
$$= \mathbf{C}_{j,j+1}\mathbf{p}_{j+1}^{i,j+1} + \boldsymbol{\rho}_j^{j+1,j}. \qquad (5)$$

Substituting (5) into (4) results in an error term with no feature variables

$$\mathbf{e}_i = \mathbf{y}_j^i - \mathbf{h}\left(\hat{\mathbf{p}}_j^{i,j}\right). \qquad (6)$$

However, the remaining design variables are inside of the non-linear function, $\mathbf{h}(\cdot)$, which will complicate the minimization algorithm. Instead, we define an alternate error term, $\boldsymbol{\epsilon}_i$, by rearranging (6) and applying the inverse measurement model

$$\mathbf{y}_j^i = \mathbf{h}\left(\hat{\mathbf{p}}_j^{i,j}\right) + \mathbf{e}_i$$
$$\mathbf{g}\left(\mathbf{y}_j^i\right) = \mathbf{g}\left(\mathbf{h}\left(\hat{\mathbf{p}}_j^{i,j}\right) + \mathbf{e}_i\right).$$

We assume $\mathbf{e}_i$ is small and approximate the right-hand side with a first-order Taylor expansion

$$\mathbf{g}\left(\mathbf{y}_j^i\right) \approx \mathbf{g}\left(\mathbf{h}\left(\hat{\mathbf{p}}_j^{i,j}\right)\right) + \underbrace{\left.\frac{\partial\mathbf{g}}{\partial\mathbf{y}_j^i}\right|_{\mathbf{h}\left(\hat{\mathbf{p}}_j^{i,j}\right)} \mathbf{e}_i}_{=:\,\boldsymbol{\epsilon}_i}$$

$$\boldsymbol{\epsilon}_i = \mathbf{g}\left(\mathbf{y}_j^i\right) - \mathbf{g}\left(\mathbf{h}\left(\hat{\mathbf{p}}_j^{i,j}\right)\right)$$
$$= \mathbf{p}_j^{i,j} - \hat{\mathbf{p}}_j^{i,j}$$
$$= \mathbf{p}_j^{i,j} - \mathbf{C}_{j,j+1}\mathbf{p}_{j+1}^{i,j+1} - \boldsymbol{\rho}_j^{j+1,j}. \qquad (7)$$

Using the error term given in (7), we define the following error function, J, that we will seek to minimize

$$J(\mathbf{C}_{j,j+1}, \boldsymbol{\rho}_j^{j+1,j}) := \frac{1}{2}\sum_{i=1}^{N} \boldsymbol{\epsilon}_i^{\mathrm{T}}\mathbf{W}_i\boldsymbol{\epsilon}_i \qquad (8)$$

where $\mathbf{W}_i$ is some weighting matrix. If the inverse covariance of $\boldsymbol{\epsilon}_i$ is used for $\mathbf{W}_i$, J is a Mahalonobis distance and thus finding the variables that minimize it also maximizes the joint likelihood of all the data. To find $\mathbf{W}_i$

we must examine the properties of $\boldsymbol{\epsilon}_i$. The mean is

$$
\begin{aligned}
\overline{\boldsymbol{\epsilon}}_i &:= E(\boldsymbol{\epsilon}_i) \\
&= E\left(\mathbf{p}_j^{i,j} - \mathbf{C}_{j,j+1}\mathbf{p}_{j+1}^{i,j+1} - \boldsymbol{\rho}_j^{j+1,j}\right) \\
&= E\left(\bar{\mathbf{p}}_j^{i,j} + \delta\mathbf{p}_j^{i,j} - \mathbf{C}_{j,j+1}\left(\bar{\mathbf{p}}_{j+1}^{i,j+1} + \delta\mathbf{p}_{j+1}^{i,j+1}\right) - \boldsymbol{\rho}_j^{j+1,j}\right) \\
&= \bar{\mathbf{p}}_j^{i,j} - \mathbf{C}_{j,j+1}\bar{\mathbf{p}}_{j+1}^{i,j+1} - \boldsymbol{\rho}_j^{j+1,j}.
\end{aligned}
$$

Noting that $\boldsymbol{\epsilon}_i - \overline{\boldsymbol{\epsilon}}_i = \delta\mathbf{p}_j^{i,j} - \mathbf{C}_{j,j+1}\delta\mathbf{p}_{j+1}^{i,j+1}$, and assuming that $E\left(\delta p_j^{i,j}\delta p_{j+1}^{i,j+1\mathrm{T}}\right) = \mathbf{0}$, the covariance of $\boldsymbol{\epsilon}_i$ is

$$
\begin{aligned}
\boldsymbol{\Sigma}_i &:= E\left((\boldsymbol{\epsilon}_i - \overline{\boldsymbol{\epsilon}}_i)(\boldsymbol{\epsilon}_i - \overline{\boldsymbol{\epsilon}}_i)^{\mathrm{T}}\right) \\
&= E\left(\left(\delta\mathbf{p}_j^{i,j} - \mathbf{C}_{j,j+1}\delta\mathbf{p}_{j+1}^{i,j+1}\right)\left(\delta\mathbf{p}_j^{i,j} - \mathbf{C}_{j,j+1}\delta\mathbf{p}_{j+1}^{i,j+1}\right)^{\mathrm{T}}\right) \\
&= E\left(\delta\mathbf{p}_j^{i,j}\delta\mathbf{p}_j^{i,j\mathrm{T}} + \mathbf{C}_{j,j+1}\delta\mathbf{p}_{j+1}^{i,j+1}\delta\mathbf{p}_{j+1}^{i,j+1\mathrm{T}}\mathbf{C}_{j,j+1}^{\mathrm{T}}\right) \\
&= \mathbf{P}_j^i + \mathbf{C}_{j,j+1}\mathbf{P}_{j+1}^i\mathbf{C}_{j,j+1}^{\mathrm{T}},
\end{aligned}
$$

from which we can define

$$
\mathbf{W}_i := \boldsymbol{\Sigma}_i^{-1}.
$$

3.1.4. Feature Tracking and Outlier Detection

Our algorithm uses scale-and-rotation-invariant feature detectors and descriptors for both *matching* (across stereo pairs) and *tracking* (over time). This is a class of feature pioneered by Lowe (2004). Lowe's SIFT algorithm has been used previously for object recognition (Nistér and Stewenius 2006), simultaneous localization and mapping (Folkesson and Christensen 2007; Se et al. 2002), and VO (Barfoot 2005). The SURF is a similar algorithm that is much faster to compute because it uses integral images to approximate the operations used to generate SIFT features (Bay et al. 2006). The SURF algorithm produces a list of features from an image. Each feature has a normalized 64-dimensional description vector (descriptor) which is abstracted from the scale and orientation of the feature in the image. Feature descriptors are used for data association in both matching and tracking. A descriptor comparison is considered a match if the angle between the descriptors is small and other descriptors are sufficiently far away (in descriptor space). The left and right image coordinates of a successful match at time j are stacked into a candidate observation.

After two consecutive stereo pairs have been matched, features are tracked between frames. Feature descriptor matches between the consecutive left images are used as candidate tracks. As is evident in Figure 3(c), candidate feature tracks contain a number of outliers. Our error function (8) is sensitive to outliers and so inlying feature tracks must be identified at this stage of the pipeline. We use a version of RANSAC (Fischler and Bolles 1981) to simultaneously reject outlier feature tracks and produce an initial motion estimate that is used to initialize our maximum likelihood solution. For each iteration of RANSAC, three pairs are selected randomly from the set of candidate tracks. The observations are triangulated using the inverse measurement model (3) and the efficient three-point algorithm developed by Fiore (2001) is used to solve $\{\mathbf{C}_{j,j+1}, \boldsymbol{\rho}_j^{j+1,j}\}$. Given this estimate of the rotation and translation, (6) is used to evaluate all candidate tracks. Tracks with $\|\mathbf{e}_i\|$ less than a threshold are considered inliers. The rotation and translation with the highest number of inliers is used as the initial guess for Gauss–Newton minimization.

3.1.5. Gauss–Newton Solution Method

To minimize (8), we use the Gauss–Newton method. This approach is very similar to the one developed in Matthies (1989). However, we parameterize rotation using a local small-angle formulation, which is not subject to singularities (Triggs et al. 2000). First, we introduce the rotation vector, $\boldsymbol{\theta}$, and use it to decompose $\mathbf{C}_{j,j+1}$ into the product of a small rotation, $\delta\mathbf{C}(\boldsymbol{\theta})$, and a large average rotation, $\overline{\mathbf{C}}_{j,j+1}$

$$
\mathbf{C}_{j,j+1} = \delta\mathbf{C}(\boldsymbol{\theta})\overline{\mathbf{C}}_{j,j+1} \tag{9}
$$

$$
\delta\mathbf{C}(\boldsymbol{\theta}) := \mathbf{a}\mathbf{a}^{\mathrm{T}} + (\mathbf{1} - \mathbf{a}\mathbf{a}^{\mathrm{T}})\cos\theta - \mathbf{a}^{\times}\sin\theta \tag{10}
$$

$$
\theta := \|\boldsymbol{\theta}\|
$$

$$
\mathbf{a} := \frac{\boldsymbol{\theta}}{\theta}
$$

where $\mathbf{1}$ is the identity matrix and $(\cdot)^{\times}$ indicates the cross product operation given by

$$
\begin{bmatrix} x \\ y \\ z \end{bmatrix}^{\times} := \begin{bmatrix} 0 & -z & y \\ z & 0 & -x \\ -y & x & 0 \end{bmatrix}.
$$

When evaluating (10), if θ is close to zero, $\delta\mathbf{C}(\boldsymbol{\theta}) = \mathbf{1}$. When $\boldsymbol{\theta}$ is small, we may approximate $\mathbf{C}_{j,j+1}$ by

$$
\mathbf{C}_{j,j+1} \approx (\mathbf{1} - \boldsymbol{\theta}^{\times})\overline{\mathbf{C}}_{j,j+1}. \tag{11}
$$

With the approximation in (11), the error function is quadratic in $\boldsymbol{\theta}$ and $\boldsymbol{\rho}_j^{j+1,j}$

$$
\begin{aligned}
&J(\boldsymbol{\theta}, \boldsymbol{\rho}_j^{j+1,j}) \\
&\approx \frac{1}{2}\sum_{i=1}^{N}(\mathbf{A}_i\boldsymbol{\theta} - \mathbf{b}_i - \boldsymbol{\rho}_j^{j+1,j})^{\mathrm{T}}\mathbf{W}_i(\mathbf{A}_i\boldsymbol{\theta} - \mathbf{b}_i - \boldsymbol{\rho}_j^{j+1,j}) \\
&\mathbf{A}_i := -(\overline{\mathbf{C}}_{j,j+1}\mathbf{p}_{j+1}^{i,j+1})^{\times} \\
&\mathbf{b}_i := \overline{\mathbf{C}}_{j,j+1}\mathbf{p}_{j+1}^{i,j+1} - \mathbf{p}_j^{i,j}.
\end{aligned}
$$

We may now derive a Gauss–Newton update step. First we take the derivative of J with respect to $\boldsymbol{\rho}_j^{j+1,j}$

$$
\frac{\partial J^{\mathrm{T}}}{\partial\boldsymbol{\rho}_j^{j+1,j}} = -\sum_{i=1}^{N}\mathbf{W}_i(\mathbf{A}_i\boldsymbol{\theta} - \mathbf{b}_i - \boldsymbol{\rho}_j^{j+1,j}).
$$

Defining the intermediate term,

$$
\mathbf{W} := \sum_{i=1}^{N}\mathbf{W}_i,
$$

(a) An image from the left
camera

(b) A sample of features
matched within a stereo pair

(c) Raw feature tracks found
in a single time step of our algorithm

(d) Feature tracks remaining
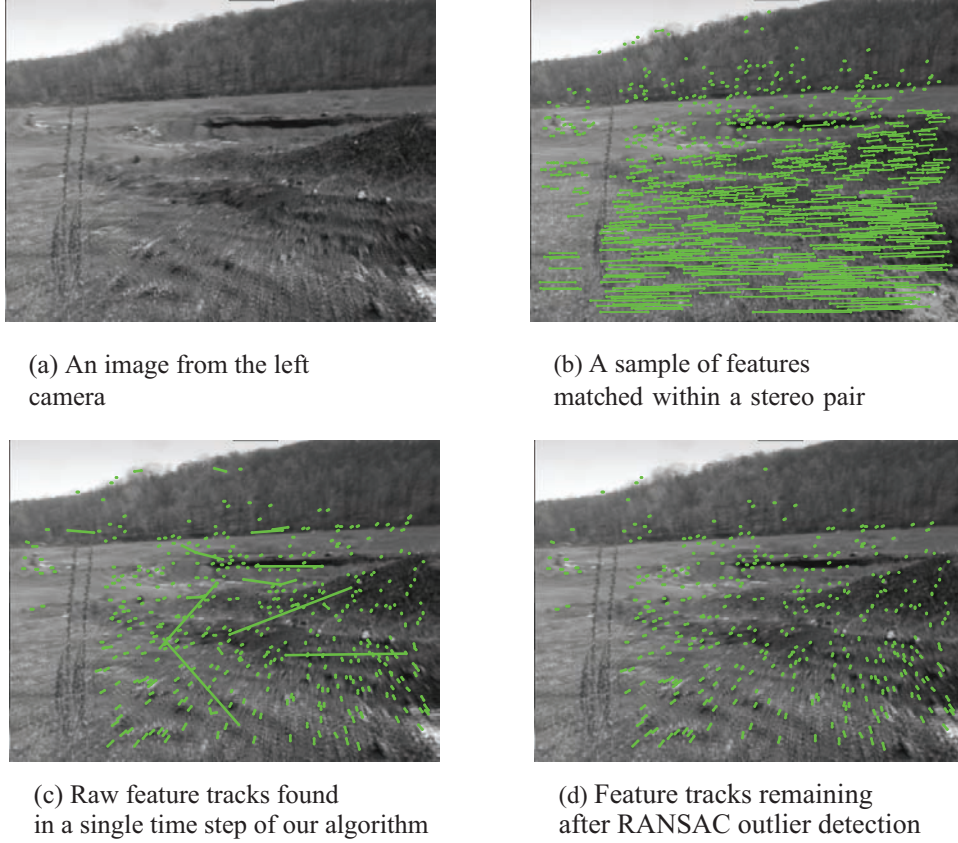after RANSAC outlier detection

Fig. 3. Features are tracked between consecutive left images using the SURF feature descriptors and then subjected to RANSAC outlier detection. The images used here are from the *Dunes* dataset (courtesy of Stanford Research Institute) described in Appendix B.

we set this to zero and solve for $\rho_j^{j+1,j}$

$$\rho_j^{j+1,j} = \mathbf{W}^{-1} \sum_{i=1}^{N} \mathbf{W}_i \left(\mathbf{A}_i \theta - \mathbf{b}_i \right). \tag{12}$$

Taking the derivative of J with respect to θ, we then substitute in the value of $\rho_j^{j,j+1}$ from (12),

$$\frac{\partial J}{\partial \theta}^T = \sum_{i=1}^{N} \mathbf{A}_i^T \mathbf{W}_i (\mathbf{A}_i \theta - \mathbf{b}_i - \rho_j^{j+1,j})$$

$$= \sum_{i=1}^{N} \mathbf{A}_i^T \mathbf{W}_i \left(\mathbf{A}_i \theta - \mathbf{b}_i - \mathbf{W}^{-1} \sum_{i=1}^{N} \mathbf{W}_i (\mathbf{A}_i \theta - \mathbf{b}_i) \right),$$

which we set to zero and solve for θ,

$$\theta = \left(\sum_{i=1}^{N} \mathbf{A}_i^T \mathbf{W}_i \mathbf{A}_i - \mathbf{L}^T \mathbf{W}^{-1} \mathbf{L} \right)^{-1}$$

$$\times \left(\sum_{i=1}^{N} \mathbf{A}_i^T \mathbf{W}_i \mathbf{b}_i - \mathbf{L}^T \mathbf{W}^{-1} \left(\sum_{i=1}^{N} \mathbf{W}_i \mathbf{b}_i \right) \right), \tag{13}$$

$$\mathbf{L} := \sum_{i=1}^{N} \mathbf{W}_i \mathbf{A}_i.$$

This is equivalent to one iteration of Gauss–Newton optimization and we use the above formulation to iteratively solve for $\mathbf{C}_{j,j+1}$. Each iteration consists of finding θ (13), computing $\delta \mathbf{C}(\theta)$ (10), and updating our guess (9). After convergence, θ is zero, so we use our optimal value of $\mathbf{C}_{j,j+1}$ in (12) to calculate $\rho_j^{j+1,j}$

$$\rho_j^{j+1,j} = \mathbf{W}^{-1} \sum_{i=1}^{N} \mathbf{W}_i (-\mathbf{b}_i)$$

$$= \mathbf{W}^{-1} \sum_{i=1}^{N} \mathbf{W}_i \left(-\mathbf{C}_{j,j+1} \mathbf{p}_{j+1}^{i,j+1} + \mathbf{p}_j^{i,j} \right) \tag{14}$$

Finally, to recover the pose of the robot at time $j+1$ in $\overrightarrow{\mathcal{F}}_1$, we compound this estimate with the previous estimate

$$\mathbf{C}_{1,j+1} = \mathbf{C}_{1,j} \mathbf{C}_{j,j+1} \tag{15}$$

$$\rho_1^{j+1,1} = \rho_1^{j,1} + \mathbf{C}_{1,j} \rho_j^{j+1,j} \tag{16}$$

3.1.6. Covariance of the Visual Odometry Estimate In this section we will describe track uncertainty in the VO estimate. We will first show how to recover the covariance from the converged Gauss–Newton estimate. Then we will derive recursive equations for updating the uncertainty

over the length of the traverse. As we will be considering the results of several individual solutions, the notation in Section 3.1.5 will have to be modified slightly. In particular, we will have to add subscripts to $\boldsymbol{\theta}$ so (11) may be re-written as

$$\begin{aligned}
\mathbf{C}_{j,j+1} &= \delta\mathbf{C}(\boldsymbol{\theta}_{j,j+1})\,\overline{\mathbf{C}}_{j,j+1} \\
&\approx \left(\mathbf{1} - \boldsymbol{\theta}_{j,j+1}^{\times}\right)\overline{\mathbf{C}}_{j,j+1} \;.
\end{aligned} \tag{17}$$

When the Gauss–Newton method described in Section 3.1.5 has converged, $\boldsymbol{\theta}_{j,j+1}$ in (13) is zero, and we have found the maximum likelihood rotation, $\mathbf{C}_{j,j+1}$, given all of the data. However, (17) still holds, and we may think of (13) as a least-squares solution for $\boldsymbol{\theta}_{j,j+1}$ at our best estimate for $\mathbf{C}_{j,j+1}$. From this insight, we can examine the properties of $\boldsymbol{\theta}_{j,j+1}$. Its mean is

$$E(\boldsymbol{\theta}_{j,j+1}) = \mathbf{0} \;,$$

where, again, $E(\cdot)$ is the expectation operator. We can read off its covariance, $\mathbf{Q}_{j,j+1}$, from (13)

$$\begin{aligned}
\mathbf{Q}_{j,j+1} &:= E(\boldsymbol{\theta}_{j,j+1}\boldsymbol{\theta}_{j,j+1}^{T}) \\
&= \left(\sum_{i=1}^{N}\mathbf{A}_i^{T}\mathbf{W}_i\mathbf{A}_i - \mathbf{L}^{T}\mathbf{W}^{-1}\mathbf{L}\right)^{-1} .
\end{aligned} \tag{18}$$

From this result and other equations in Section 3.1.5, we can define recursive equations that propagate the covariance through our trajectory estimates. But first, we must compute the covariance of our estimate of $\boldsymbol{\rho}_j^{j+1,j}$ from $\mathbf{Q}_{j,j+1}$. Substituting (17) into (14) gives

$$\begin{aligned}
\overline{\boldsymbol{\rho}}_j^{j+1,j} + \delta\boldsymbol{\rho}_j^{j+1,j} &= \mathbf{W}^{-1}\sum_{i=1}^{N}\mathbf{W}_i\left(-\overline{\mathbf{C}}_{j,j+1}\mathbf{p}_{j+1}^{i,j+1} + \mathbf{p}_j^{i,j}\right) \\
&\quad - \mathbf{W}^{-1}\sum_{i=1}^{N}\mathbf{W}_i\left(\overline{\mathbf{C}}_{j,j+1}\mathbf{p}_{j+1}^{i,j+1}\right)^{\times}\boldsymbol{\theta}_{j,j+1} \;.
\end{aligned}$$

Subtracting the nominal value, $\overline{\boldsymbol{\rho}}_j^{j+1,j}$, from both sides gives the result for $\delta\boldsymbol{\rho}_j^{j+1,j}$

$$\delta\boldsymbol{\rho}_j^{j+1,j} = \underbrace{-\mathbf{W}^{-1}\sum_{i=1}^{N}\mathbf{W}_i\left(\overline{\mathbf{C}}_{j,j+1}\mathbf{p}_{j+1}^{i,j+1}\right)^{\times}}_{=:\,\mathbf{B}_{j,j+1}}\boldsymbol{\theta}_{j,j+1}. \tag{19}$$

Again, we are interested in the mean and covariance of $\delta\boldsymbol{\rho}_j^{j+1,j}$. Its mean, $E(\delta\boldsymbol{\rho}_j^{j+1,j})$, is zero and its covariance, $\mathbf{R}_{j,j+1}$, is

$$\begin{aligned}
\mathbf{R}_{j,j+1} &:= E\left(\delta\boldsymbol{\rho}_j^{j+1,j}\delta\boldsymbol{\rho}_j^{j+1,j\,T}\right) \\
&= \mathbf{B}_{j,j+1}\mathbf{Q}_{j,j+1}\mathbf{B}_{j,j+1}^{T}.
\end{aligned}$$

Note that $\boldsymbol{\theta}_{j,j+1}$ and $\delta\boldsymbol{\rho}_j^{j+1,j}$ are not independent. We define the cross-covariance $\mathbf{S}_{j,j+1}$

$$\begin{aligned}
\mathbf{S}_{j,j+1}^{T} &:= E\left(\delta\boldsymbol{\rho}_j^{j+1,j}\boldsymbol{\theta}_{j,j+1}^{T}\right) \\
&= \mathbf{B}_{j,j+1}E\left(\boldsymbol{\theta}_{j,j+1}\boldsymbol{\theta}_{j,j+1}^{T}\right) \\
&= \mathbf{B}_{j,j+1}\mathbf{Q}_{j,j+1}.
\end{aligned}$$

Defining the column

$$\mathbf{x}_{a,b} := \begin{bmatrix}\boldsymbol{\theta}_{a,b} \\ \delta\boldsymbol{\rho}_a^{a,b}\end{bmatrix},$$

for some frames a and b, we can define the combined covariance of our estimate, $\mathbf{D}_{j,j+1}$:

$$\begin{aligned}
\mathbf{D}_{j,j+1} &:= E\left(\mathbf{x}_{j,j+1}\mathbf{x}_{j,j+1}^{T}\right) \\
&= \begin{bmatrix}\mathbf{Q}_{j,j+1} & \mathbf{S}_{j,j+1} \\ \mathbf{S}_{j,j+1}^{T} & \mathbf{R}_{j,j+1}\end{bmatrix}.
\end{aligned} \tag{20}$$

These results may be used derive recursive update equations for propagating the uncertainty through (15) and (16). At any point, we assume the following quantities are available:

$\mathbf{C}_{1,j}$: an estimate of the rotation up to frame j;

$\boldsymbol{\rho}_1^{j,\,1}$: an estimate of the translation up to frame j;

$\mathbf{D}_{1,j}$: the covariance associated with the complete estimate up to frame j, $E\left(\mathbf{x}_{1,j}\mathbf{x}_{1,j}^{T}\right)$;

$\mathbf{C}_{j,j+1}$: the maximum-likelihood rotation estimate from a single step of Gauss–Newton ;

$\boldsymbol{\rho}_j^{j+1,j}$: the maximum-likelihood translation estimate from a single step of Gauss–Newton ;

$\mathbf{D}_{j,j+1}$: the covariance $E\left(\mathbf{x}_{j,j+1}\mathbf{x}_{j,j+1}^{T}\right)$.

Substituting (17) into (15) gives

$$\begin{aligned}
\left(\mathbf{1} - \boldsymbol{\theta}_{1,j+1}^{\times}\right)\overline{\mathbf{C}}_{1,j+1} &= \left(\mathbf{1} - \boldsymbol{\theta}_{1,j}^{\times}\right)\overline{\mathbf{C}}_{1,j}\left(\mathbf{1} - \boldsymbol{\theta}_{j,j+1}^{\times}\right)\overline{\mathbf{C}}_{j,j+1} \\
&\approx \overline{\mathbf{C}}_{1,j}\overline{\mathbf{C}}_{j,j+1} - \boldsymbol{\theta}_{1,j}^{\times}\overline{\mathbf{C}}_{1,j}\overline{\mathbf{C}}_{j,j+1} - \overline{\mathbf{C}}_{1,j}\boldsymbol{\theta}_{j,j+1}^{\times}\overline{\mathbf{C}}_{j,j+1} \\
&= \overline{\mathbf{C}}_{1,j+1} - \boldsymbol{\theta}_{1,j}^{\times}\overline{\mathbf{C}}_{1,j+1} - \overline{\mathbf{C}}_{1,j}\boldsymbol{\theta}_{j,j+1}^{\times}\overline{\mathbf{C}}_{1,j}^{T}\overline{\mathbf{C}}_{1,j}\overline{\mathbf{C}}_{j,j+1} \\
&= \overline{\mathbf{C}}_{1,j+1} - \boldsymbol{\theta}_{1,j}^{\times}\overline{\mathbf{C}}_{1,j+1} - \left(\overline{\mathbf{C}}_{1,j}\boldsymbol{\theta}_{j,j+1}\right)^{\times}\overline{\mathbf{C}}_{1,j+1} \\
&= \left(\mathbf{1} - \left(\boldsymbol{\theta}_{1,j} + \overline{\mathbf{C}}_{1,j}\boldsymbol{\theta}_{j,j+1}\right)^{\times}\right)\overline{\mathbf{C}}_{1,j+1},
\end{aligned}$$

where we have used the identity $\mathbf{A}\mathbf{b}^{\times}\mathbf{A}^{T} = (\mathbf{A}\mathbf{b})^{\times}$ (for a rotation matrix, $\mathbf{A}$, and a 3×1 column, $\mathbf{b}$), and dropped the

products of small terms. Removing the nominal value from both sides gives us an equation for $\boldsymbol{\theta}_{1,j+1}$

$$\boldsymbol{\theta}_{1,j+1} = \boldsymbol{\theta}_{1,j} + \overline{\mathbf{C}}_{1,j}\boldsymbol{\theta}_{j,j+1}. \qquad (21)$$

Following the same process, we expand (16)

$$
\begin{aligned}
\overline{\boldsymbol{\rho}}_1^{j+1,\,1} + \delta\boldsymbol{\rho}_1^{j+1,\,1} &= \overline{\boldsymbol{\rho}}_1^{j,\,1} + \delta\boldsymbol{\rho}_1^{j,\,1} \\
&\quad + \left(\mathbf{1} - \boldsymbol{\theta}_{1,j}^{\times}\right)\overline{\mathbf{C}}_{1,j}\left(\overline{\boldsymbol{\rho}}_j^{j+1,\,j} + \delta\boldsymbol{\rho}_j^{j+1,\,j}\right) \\
&\approx \overline{\boldsymbol{\rho}}_1^{j,\,1} + \overline{\mathbf{C}}_{1,j}\overline{\boldsymbol{\rho}}_j^{j+1,\,j} + \delta\boldsymbol{\rho}_1^{j,\,1} - \boldsymbol{\theta}_{1,j}^{\times}\overline{\mathbf{C}}_{1,j}\overline{\boldsymbol{\rho}}_j^{j+1,\,j} \\
&\quad + \overline{\mathbf{C}}_{1,j}\delta\boldsymbol{\rho}_j^{j+1,\,j} \\
&= \overline{\boldsymbol{\rho}}_1^{j,\,1} + \overline{\mathbf{C}}_{1,j}\overline{\boldsymbol{\rho}}_j^{j,\,j+1} + \delta\boldsymbol{\rho}_1^{j,\,1} \\
&\quad + \left(\overline{\mathbf{C}}_{1,j}\overline{\boldsymbol{\rho}}_j^{j+1,\,j}\right)^{\times}\boldsymbol{\theta}_{1,j} + \overline{\mathbf{C}}_{1,j}\delta\boldsymbol{\rho}_j^{j+1,\,j}
\end{aligned}
$$

where we have dropped the products of small terms. Subtracting the nominal value from both sides gives us an equation for $\delta\boldsymbol{\rho}_1^{j+1,\,1}$,

$$\delta\boldsymbol{\rho}_1^{j+1,\,1} = \delta\boldsymbol{\rho}_1^{j,\,1} + \left(\overline{\mathbf{C}}_{1,j}\overline{\boldsymbol{\rho}}_j^{j+1,\,j}\right)^{\times}\boldsymbol{\theta}_{1,j} + \overline{\mathbf{C}}_{1,j}\delta\boldsymbol{\rho}_j^{j+1,\,j}. \quad (22)$$

Stacking (21) and (22) gives us the combined update equation

$$\mathbf{x}_{1,j+1} = \underbrace{\begin{bmatrix} \mathbf{1} & \mathbf{0} \\ (\overline{\mathbf{C}}_{1,j}\overline{\boldsymbol{\rho}}_j^{j+1,\,j})^{\times} & \mathbf{1} \end{bmatrix}}_{=:\mathbf{U}}\mathbf{x}_{1,j} + \underbrace{\begin{bmatrix} \overline{\mathbf{C}}_{1,j} & \mathbf{0} \\ \mathbf{0} & \overline{\mathbf{C}}_{1,j} \end{bmatrix}}_{=:\mathbf{V}}\mathbf{x}_{j,j+1}. \quad (23)$$

Using $E\left(\mathbf{x}_{1,j}\mathbf{x}_{j,j+1}^{\mathrm{T}}\right) = \mathbf{0}$ (i.e. the current estimate is independent of any previous estimates), the recursive uncertainty update is

$$
\begin{aligned}
\mathbf{D}_{1,j+1} &:= E\left(\mathbf{x}_{1,j+1}\mathbf{x}_{1,j+1}^{\mathrm{T}}\right) \\
&\approx \mathbf{U}\mathbf{D}_{1,j}\mathbf{U}^{\mathrm{T}} + \mathbf{V}\mathbf{D}_{j,j+1}\mathbf{V}^{\mathrm{T}}.
\end{aligned}
$$

3.2. Three-dimensional Surface/Subsurface Modeling

3.2.1. Surface Mesh Generation Our three-dimensional, photorealistic terrain models are generated by passing our motion estimate into the instant scene modeler (ISM) developed at MDA Space Missions (Se and Jasiobedzki 2008; Barfoot et al. 2006). At time j we use the sum-of-absolute-differences, correlation-based, dense-stereo algorithm to generate a set of Q_j scene points in $\underset{\rightarrow}{\mathcal{F}}_j$

$$\left\{ \mathbf{q}_j^{i,j} \,\middle|\, 1 \le i \le Q_j \right\}. \qquad (24)$$

The motion estimate from (15) and (16) is then used to transform the points from all frames into $\underset{\rightarrow}{\mathcal{F}}_1$

$$\mathbf{q}_1^{i,1} = \mathbf{C}_{1,j}\mathbf{q}_j^{i,j} + \boldsymbol{\rho}_1^{j,\,1}. \qquad (25)$$

Aligning dense point clouds from multiple views fills in holes in the data and provides us with the rich volumetric information we need to create a surface mesh. Dense-stereo methods are often used for terrain assessment (Lacroix et al. 2001; Biesiadecki and Maimone 2006) and so the point clouds may be available onboard at no extra computational cost. However, using every point obtained from stereo processing in the terrain model is not efficient. There are many redundant measurements and, due to incorrect matches, occlusions, or lack of texture, the data may contain noise and missing regions. A triangular mesh representation reduces the amount of data when multiple point sets are combined and thus also reduces the amount of bandwidth needed to send the resulting models offboard (e.g. to Earth). Creating a surface mesh fills up small holes and eliminates outliers, resulting in smoother and more realistic reconstructions.

To generate triangular meshes from the merged point set, we employ a voxel-based method (Roth and Wibowoo 1997), which accumulates three-dimensional points with their associated normals. The points are accumulated into voxels at each frame. Outliers are filtered out using their local orientation and by selecting the threshold of range measurements required per voxel for a valid mesh vertex. It takes a few seconds to construct the triangular mesh, which is dependent on the data size and the voxel resolution.

Photo-realistic appearance of the reconstructed scene is created by texture mapping the triangular mesh using the original images. Such surfaces are more visually appealing and easier to interpret as they provide additional surface details. Color images from the stereo camera are used for texture mapping. The vertices of each triangle in the mesh are projected into the images using (1) as shown in Figure 4. As each triangle may be observed in multiple images, one image must be selected to provide the texture for each triangle. The algorithm selects the smallest set of images that will cover all of the triangles. This decreases the amount of bandwidth required to transmit the models, reduces the appearance of texture discontinuities between adjacent triangles, and makes the scene easier to load and view.

Scenes produced by ISM are exported as virtual reality modeling language (VRML) files and may be viewed in any VRML viewer. The meshes are exported in metric scale based on the motion estimate and the points derived from stereo processing. As such, the resulting three-dimensional model provides a rich record of the site and may be used for scientific interpretation. Osinski et al. (2010) explore the use of ISM as a scientific tool.

3.2.2. GPR Mesh Generation After estimating the rover's motion and building a terrain model, we couple the subsurface data to the surface model. The data from the traces along a rover traverse are associated with the rover's motion estimate through common time-stamps. With this
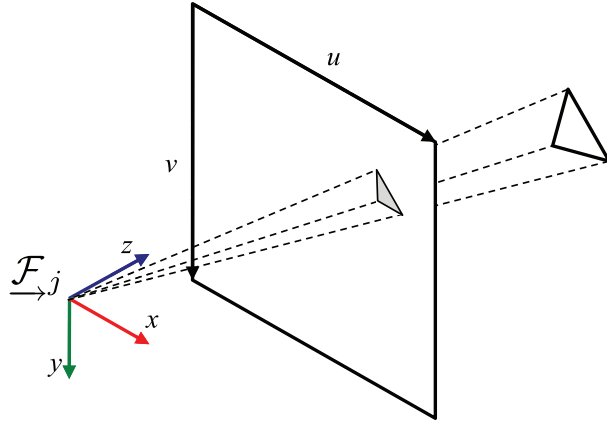
Fig. 4. The photorealism of the surface model is achieved by mapping the original color input images on to the surface mesh. Texture coordinates (u, v) for each triangle in the terrain model are found using (1) to project the triangle's vertices into the color images captured by the rover.
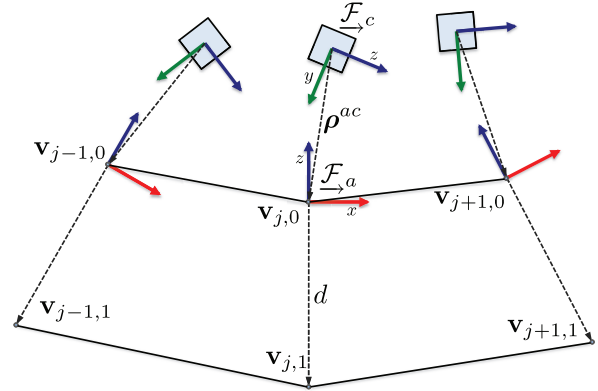

Fig. 5. The subsurface model is built by reconstructing the trajectory of the GPR antenna over the course of the traverse. From the six-degree-of-freedom motion estimate, the calibration parameters, $\{\mathbf{C}_{ca}, \boldsymbol{\rho}_c^{ac}\}$, and the scan depth, d, we can build a three-dimensional polygonal mesh, which is texture-mapped with a radargram of the input GPR data. The resulting model goes beyond a typical topographic correction as the traces are corrected for wheel slip, tilt, and topography.
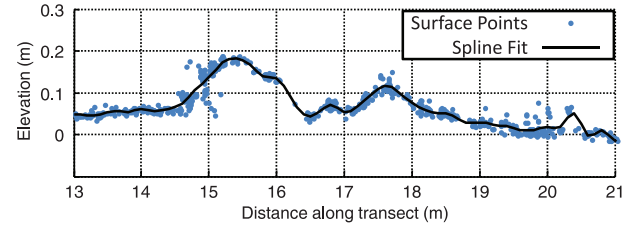
information, we can build a three-dimensional model of the GPR data.

To display the subsurface data with the surface model, we build a mesh for the GPR data. In a static configuration, we define the left camera frame (discussed in Section 3.1) as $\underrightarrow{\mathcal{F}}_c$ and the frame attached to the GPR antenna as $\underrightarrow{\mathcal{F}}_a$. Calibration of the rover determines the static transformation between $\underrightarrow{\mathcal{F}}_c$ and $\underrightarrow{\mathcal{F}}_a$, $\{\mathbf{C}_{ca}, \boldsymbol{\rho}_c^{ac}\}$. The depth of each trace, d, is calculated from the measured signal return time and the estimated velocity of the radar pulse through the subsurface (Hinkel et al. 2001). Electromagnetic wave velocity typical for ice-rich substrates in permafrost is approximately 120 m/μs (Fortier and Allard 2004). The mesh created in Section 3.2 is represented in frame $\underrightarrow{\mathcal{F}}_1$ and so we create the GPR mesh in the same frame. For each time j in the motion estimate, we create two vertices,

$$\mathbf{v}_{j,0} = \mathbf{C}_{1,j} \boldsymbol{\rho}_c^{ac} + \boldsymbol{\rho}_1^{j,1} \tag{26}$$

$$\mathbf{v}_{j,1} = \mathbf{C}_{1,j} \left(\boldsymbol{\rho}_c^{ac} + \mathbf{C}_{ca} \begin{bmatrix} 0 \\ 0 \\ -d \end{bmatrix} \right) + \boldsymbol{\rho}_1^{j,1}, \tag{27}$$

representing the antenna position ($\mathbf{v}_{j,0}$) and the position at the bottom of the trace ($\mathbf{v}_{j,1}$). For each time $2 \leq j \leq M$, the polygon, $\{\mathbf{v}_{j-1,0}, \mathbf{v}_{j-1,1}, \mathbf{v}_{j,1}, \mathbf{v}_{j,0}\}$, is added to the mesh. The geometry of this polygon construction is shown in Figure 5. Using the VO-derived motion estimate to build both the surface model and the subsurface mesh results in a coupled model by construction; all points in both meshes are expressed in the same coordinate frame, $\underrightarrow{\mathcal{F}}_1$.

Texture coordinates are assigned to the vertices based on the time-stamp association between the rover poses and the traces in the GPR radargram. Texture mapping the GPR data on to this mesh goes one step beyond standard topographic correction. Because the full six-degree-of-freedom pose of the rover is used to generate the mesh, the

Fig. 6. Fitting a spline to the surface points (reconstructed from the stereo images) along the GPR antenna's path produces a high-resolution surface profile. This captures narrow terrain features that may not be present in the motion estimate (due to the antenna's size having a smoothing effect) and provides further information about surface roughness along the transect.

texture-mapped traces are corrected for tilt, topography, and wheel slip. Future work could involve using the samples from collocated scans to perform automated dense GPR reconstruction as in Lehmann and Green (1999).

3.3. Two-dimensional Topographic Correction and Surface Profile

When the GPR transect of interest lies along sloping or undulating terrain, it is typical to perform a topographic survey using standard survey instruments such as DGPS (Lunt and Bridge 2004) or an engineer's level (Busby and Merritt 1999). The motion estimate from Section 3.1 can be used for the same purpose. The height of the antenna at each trace is interpolated from the motion estimate and this is used to apply a topographic correction to the transect.
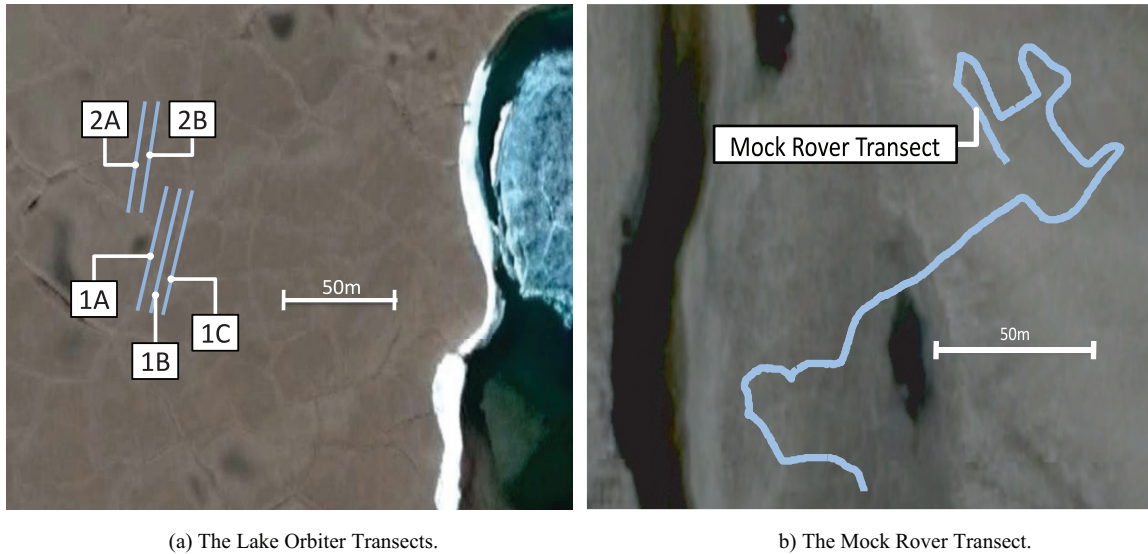
(a) The Lake Orbiter Transects.

(b) The Mock Rover Transect.

Fig. 7. Locations and transects on Devon Island, Nunavut, Canada used for field testing our integrated surface/subsurface modeling technique.

Further, given only the sparse feature points used to produce the motion estimate, we can plot the points along the antenna path and fit a spline that describes the surface. As only this sparse set of features is used, this technique avoids having to run a dense-stereo algorithm on the original images. The surface spline improves on the topographic correction as it is able to capture narrow features over which the antenna slides. This extra data allows the user to account for the random clutter field that surface roughness produces in the GPR data (Daniels 2004). A sample surface profile is shown in Figure 6.

4. Field Testing

4.1. Test Site

The experiments described in this paper were conducted on Devon Island in the Canadian High Arctic, as part of the Haughton-Mars Project (Lee et al. 2007). The HMP-RS is situated just outside the northwest area of the Haughton impact crater, which is located at 75° 22′ N latitude and 89° 41′ W longitude. Haughton presents unique qualities for planetary analogue studies because it offers an unusually wide variety of geological features and microbiological attributes of strong planetary analogue value or potential. Haughton is also in a polar desert environment, which presents real challenges to field exploration that are analogous in fundamental ways to those expected in planetary exploration. This site has been used for rover testing in the past (Wettergreen et al. 2002, 2005; Fong et al. 2007, 2008).

Our experiments were conducted approximately 10 km northeast of HMP-RS near Lake Orbiter. This site was selected based on ongoing research into the polygonal terrain it hosts. Image sequences from the stereo camera and GPR data were logged at two sites:

1. The Lake Orbiter Transects: Five straight-line transects were taken at the Lake Orbiter site (Figure 7(a)). Each transect is approximately 60 m long.
2. The Mock Rover Transect: One approximately 357 m transect at a site that had not been previously studied (Figure 7(b)).

Polygonal terrain—a network of interconnected trough-like depressions in the ground—is a landform commonly found throughout the polar regions of both Earth (Lachenbruch 1962) and Mars (Kuzmin and Zabalueva 2003). In terrestrial environments, these features are formed by the response of an ice-bonded substrate to thermal forcing mechanisms induced by winter freezing and subsequent warming later in the season (Mackay 2000) and are often indicative of subsurface ice bodies termed "ice wedges" (Mackay 1999). On Mars, the recent Phoenix mission appears to have confirmed the presence of an ice-bonded substrate, but evidence of massive ice wedges has not yet been identified. However, previous studies on Mars polygonal terrain has suggested that such ice bodies may indeed exist beneath the Martian surface (e.g. Mellon (1997) and Mangold et al. (2004)). The two sites selected for investigation are comprised primarily of poorly sorted angular clasts ranging from centimeters to tens of centimeters in size. The polygonal shapes measure a few meters to tens of meters between subsequent troughs, with individual troughs averaging approximately 1–2 m in width and tens of centimeters in depth.

4.2. Rover

In our experiments, a rover was simulated using a pushcart equipped with a stereo camera, a ground-penetrating radar, an on-board computer, and two independent GPS systems (one Real-Time Kinematic) used for ground-truth
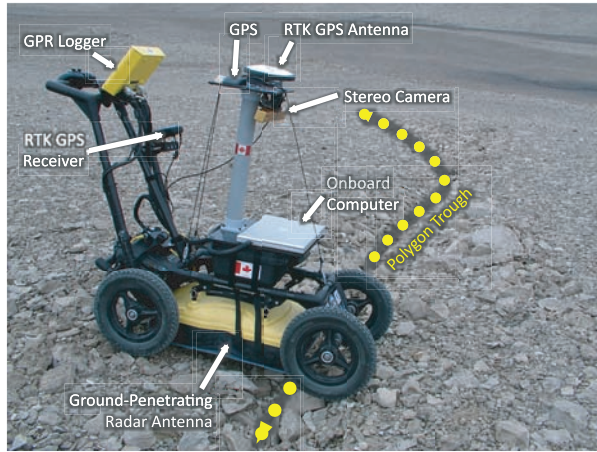
(a) Rover mockup.
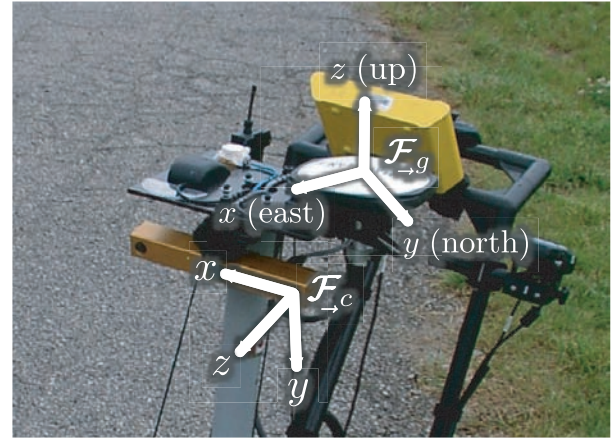
(b) Frames associated with the rover.

Fig. 8. The rover platform used for field testing.

positioning. See Figure 8(a). Although this was not an actuated rover, our focus in this work is on problems of estimation, and thus it was entirely sufficient as a means to gather data. The GPR (and cart) we used was a Sensors&Software Noggin 250 MHz system (Barfoot et al. 2003). Efforts were made to minimize the effect of the rover body on the GPR data quality (e.g. using plastic parts where possible). The stereo camera was a Point Gray Research Bumblebee XB3 with a 24 cm baseline and $70°$ field of view, mounted approximately 1 m above the surface pointing downward by approximately $20°$. Each image of the stereo pair was captured at 1280×960 pixel resolution. A video of the data collection process is shown in Extension 1.

5. Results

5.1. Motion Estimates

The VO algorithm described in Section 3.1 was used to process all data collected at the Lake Orbiter site. We used a real-time kinematic GPS unit as groundtruth for our motion estimate. The GPS frame is depicted in Figure 8(b). The x-axis is aligned with lines of longitude (east positive), the y-axis is aligned with the lines of latitude (north positive) and the z-axis points up. The inclinometers were used to determine the initial pitch and roll of the rover. However, as the data was collected near the north magnetic pole, we were unable to use a compass to determine the initial heading of the vehicle. We determine the initial heading through a least-squares fit of the estimated track to the GPS for a small number of poses at the start of the traverse. These poses are then discarded and are not used when evaluating the linear position error. This is similar to the method used by Nistér et al. (2004).

The results are shown in Table 1, which lists the distance traveled and errors for all datasets collected. On the short Lake Orbiter transects (50–60 m), position errors were

Table 1. VO Motion Estimate Results

Transect	Length (m)	Linear Error (m)	% Error	Images
Mock Rover	357.30	5.83	1.63%	4,818
poly-1AS-1	54.23	2.87	5.29%	333
poly-1BN-1	59.63	0.68	1.14%	316
poly-1BS-1	60.06	1.24	2.06%	317
poly-1CN-1	60.67	1.98	3.26%	327
poly-2AN-1	51.49	1.04	2.01%	270
poly-2AS-1	50.16	0.25	0.51%	263
poly-2BN-1	49.47	1.16	2.34%	260
poly-2BS-1	49.05	0.96	1.96%	258
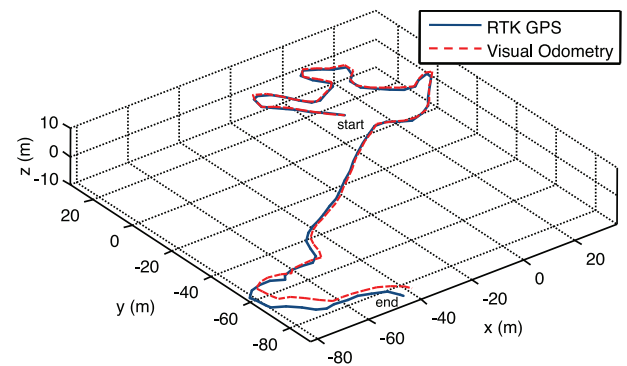
Fig. 9. Track plots of the GPS and VO estimate on the 357 m Mock Rover Transect.

no greater than 5.3% of distance traveled. The results of the estimation on the Mock Rover Transect are plotted in Figures 9 and 10. The estimate accumulated 1.63% position error over this 357.3 m traverse. Figure 10 also shows the 3σ confidence envelope of the translation estimate derived from the diagonals of the trajectory covariance matrices, $\mathbf{D}_{1,j}$. The bound makes it appear that the algorithm is becoming more confident at certain points during the
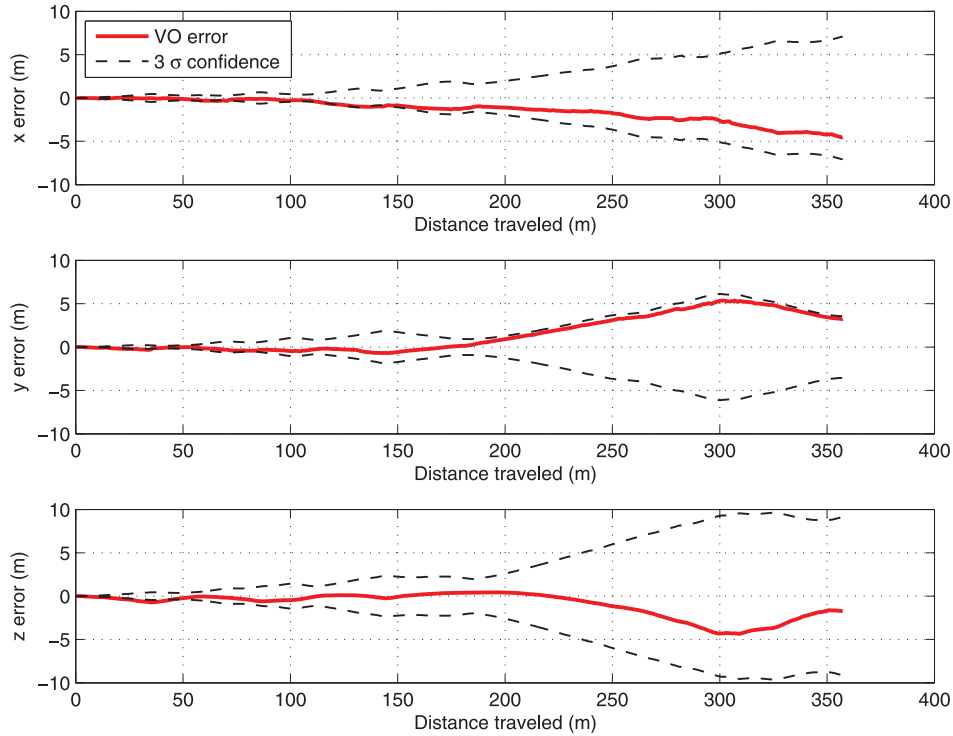
Fig. 10. Error plots for the visual odometry estimate of the 357 m Mock Rover Transect. The black dotted line is the 3σ confidence bound derived from the diagonals of the trajectory covariance matrix, $\mathbf{D}_{1,j}$, at each timestep.

traverse, but this is only the result of projecting the uncertainty onto the axes. Figure 11 shows that the total uncertainty is monotonically increasing. This is to be expected for any dead-reckoning method.
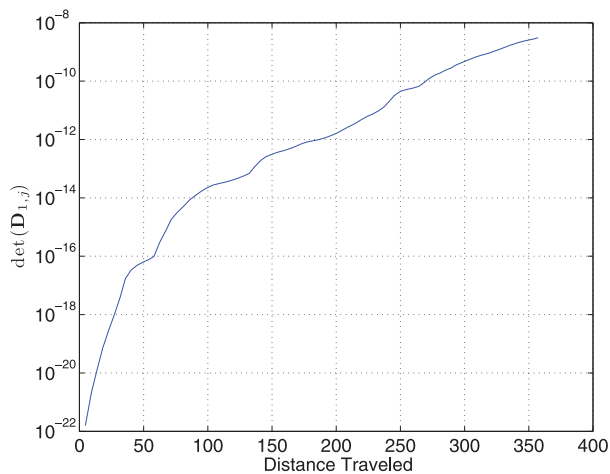

Fig. 11. The determinant of the covariance matrix, $\mathbf{D}_{1,j}$, can be used as a measure of the total uncertainty. Plotting this value over the Mock Rover Transect shows that the total uncertainty is monotonically increasing. This is expected for any dead-reckoning method.

We performed further tests of our VO algorithm on a dataset provided by the Stanford Research Institute. These further results are shown in Appendix B, and are comparable to those from the Lake Orbiter data.

5.2. Coupled Surface/Subsurface Models

The complete coupled surface/subsurface model of the Mock Rover Transect is shown in Figure 12. The texture-mapped triangle mesh of the surface is displayed above the ribbon of GPR data. The model may be inspected using a VRML viewer and rendered from any viewpoint using off-the-shelf computer components and OpenGL. A video flythrough of this model is included in Extension 1.

Figure 13 shows the entire corrected GPR radargram produced from data collected at the Mock Rover site. As is evident, the application of topographic corrections provides a more realistic representation of subsurface stratigraphy with respect to variations in surface elevation than would with an uncorrected model.

GPR is particularly useful for identifying subsurface ice because of the vast differences in dielectric properties displayed by ice, sediment, water, and air (Milsom 2003). In ice-rich sediments, for example, it is quite easy to distinguish frozen from unfrozen sediments (e.g. detecting the base of the seasonally thawed *active layer* (Moorman et al. 2003; dePascale et al. 2008)) because of the immediate increase in permittivity that occurs due to
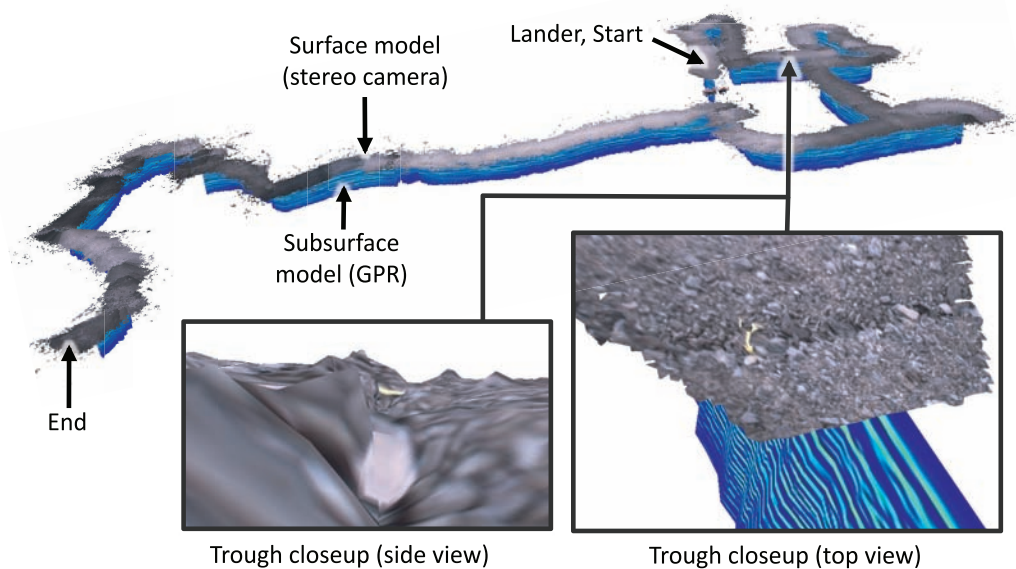
Fig. 12. Views of the key dataproduct in this paper: a coupled surface/subsurface model built from stereo imagery and a GPR, two sensors slated to fly on upcoming Mars missions. The texture-mapped triangle mesh of the surface is displayed above the ribbon of GPR data. The model may be rendered from any viewpoint using off-the-shelf computer components and OpenGL. The inset figures show possible views of one of the polygon troughs along the transect. The coupled model provides situational awareness, allowing operators to explore the interaction of surface topography and subsurface structure even though they are working remotely.

freezing at the top of the permafrost (Scott et al. 1990). In Figure 13, a noticeable continuous reflector virtually mirroring the surface topography is located approximately 1 m beneath the surface along the transect length. This depth corresponds well with the permafrost table located by augering at the highlighted section on Figure 13, and thus the reflection is interpreted as the base of the active layer.

In addition to detecting the frozen/unfrozen unconformity within the ground, a boundary between a discrete ground ice body and its enclosing sediments should produce a strong distinguishable reflector because of a sharp decrease in permittivity at the sediment/ice interface (Moorman and Michel 2000). However, although ice wedges are discrete ice bodies, imaging their structure properly using GPR is not straightforward.

As noted by Hinkel et al. (2001, p. 187), ice wedges "produce exceedingly complex, high amplitude hyperbolic reflections" due to the conical shape of the emitted GPR pulse. As a result, while ice wedges themselves are roughly triangular in shape—wider at the top and progressively narrowing with depth—their appearance on a radargram more resembles an inverted hyperbola (e.g. Hinkel et al. (2001) and Fortier and Allard (2004)).

Figure 13 illustrates an example of a hyperbolic subsurface reflection detected within the radargram. At this and other locations along the transect, the hyperbolic reflectors are found immediately beneath the troughs as indicated by small V-shaped depressions in the stereo camera surface profile. Because polygon troughs are the most obvious surface expression of ice wedge

locations (Mackay 1999), the successful coupling of our surface/subsurface model is further supported.

6. Conclusion and Future Work

We have presented a coupled surface/subsurface modeling method for planetary exploration. Our method uses only a stereo camera and a ground-penetrating radar unit to produce:

1. An estimate of the rover's trajectory over the course of the transect.
2. A photorealistic three-dimensional surface/subsurface model.
3. Topography-corrected GPR traces plotted with a two-dimensional profile of the surface along the transect.

The models and corrections allow operators to work remotely, surveying the data as if they were on site. Our approach automates GPR collection without the use of GPS, thereby enabling the technology to be used in a planetary exploration context. The data products produced allow subsurface stratigraphy to be examined in the context of the surface morphology, a key scientific technique used by field geologists to identify sites worthy of further study. We collected our data in a Mars analogue environment at sites of scientific interest. The coupled models were generated only using sensors that are slated to fly on future rover missions, such as ExoMars.
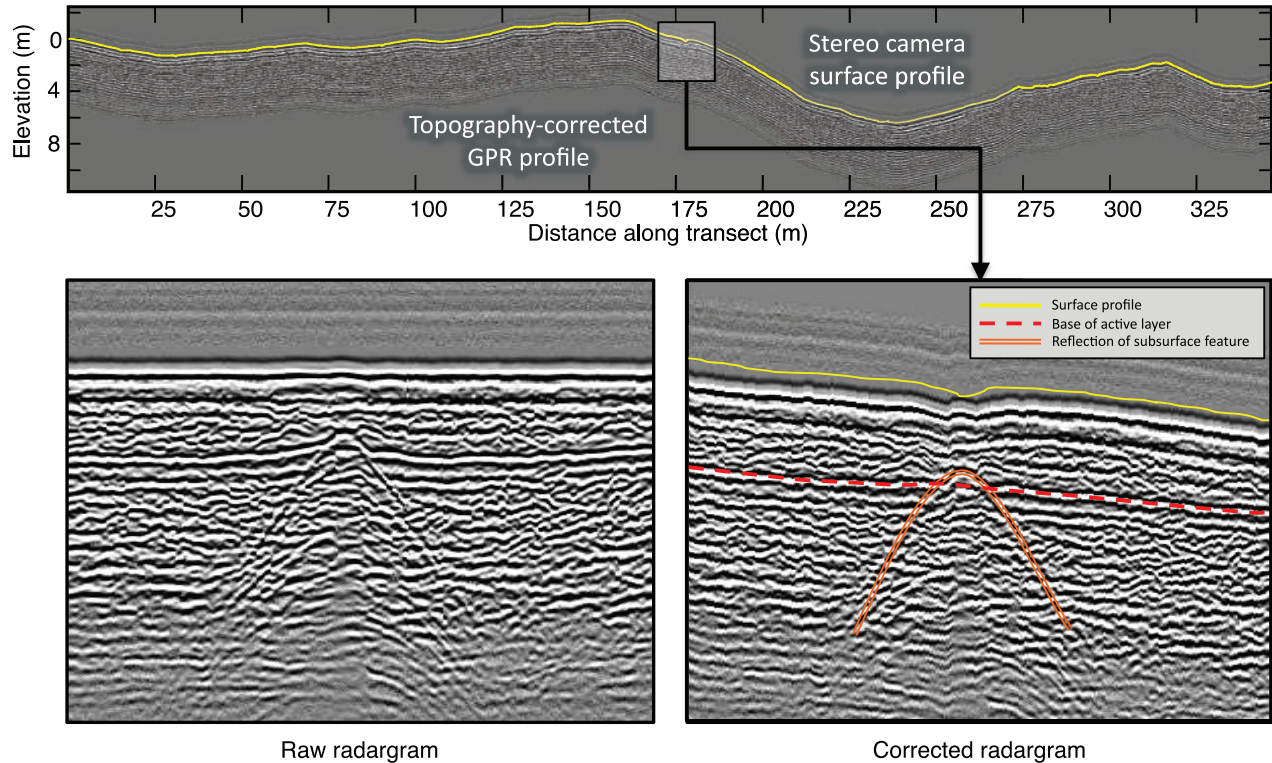
Fig. 13. The complete corrected GPR radargram of the Mock Rover site (above), produced using the process described in Section 3.3. The surface profile derived from stereo processing is plotted in solid/yellow. The bottom plots compare the difference between the corrected and uncorrected radargram for a portion of the data near a large polygon trough. The corrected radargram shows that the base of the seasonally-thawed *active layer* (dashed/red) is nearly flat and that a dip in the surface topography corresponds to a subsurface feature, shown here by a hyperbolic reflection pattern (doubleline/orange). As is evident, the topographic correction and surface profile provide further context to aid in the interpretation of the subsurface stratigraphy.

Fig. 14. Future work will include studying the signal corruptioniss ues when using GPR on an actuated rover.

Before a GPR is deployed on a flight platform, the significant issue of GPR/rover integration must be addressed (Barfoot et al. 2003). The proximity of the rover to the GPR will produce signal corruption due to the reflections from the metal chassis and electromagnetic signals generated by the actuators. Attempts to shield the antenna from these effects will increase the mass of the payload and may also produce interference at the receiving antenna (Annan 2009). In short, this is a major issue that must be addressed before rover-based GPR data collection is deployed in space. In the summer of 2009, we returned to Devon Island to investigate this issue using the actuated rover platform shown in Figure 14. Anecdotally, we found that the rover did not have much influence on the GPR data collected, but this issue requires further attention and depends heavily on the specific robotic platform.

While we are pleased by the results generated by this implementation of VO, they are not nearly as accurate as modern methods incorporating sparse bundle adjustment (Konolige et al. 2007). Unfortunately, the restrictive computational power and memory available on flight platforms may preclude the use of these advances in near-term missions. However, the super-linear growth of position error in VO estimates is due to the effect of orientation error (Olson et al. 2003). Future work in this area will focus on incorporating direct attitude measurements from
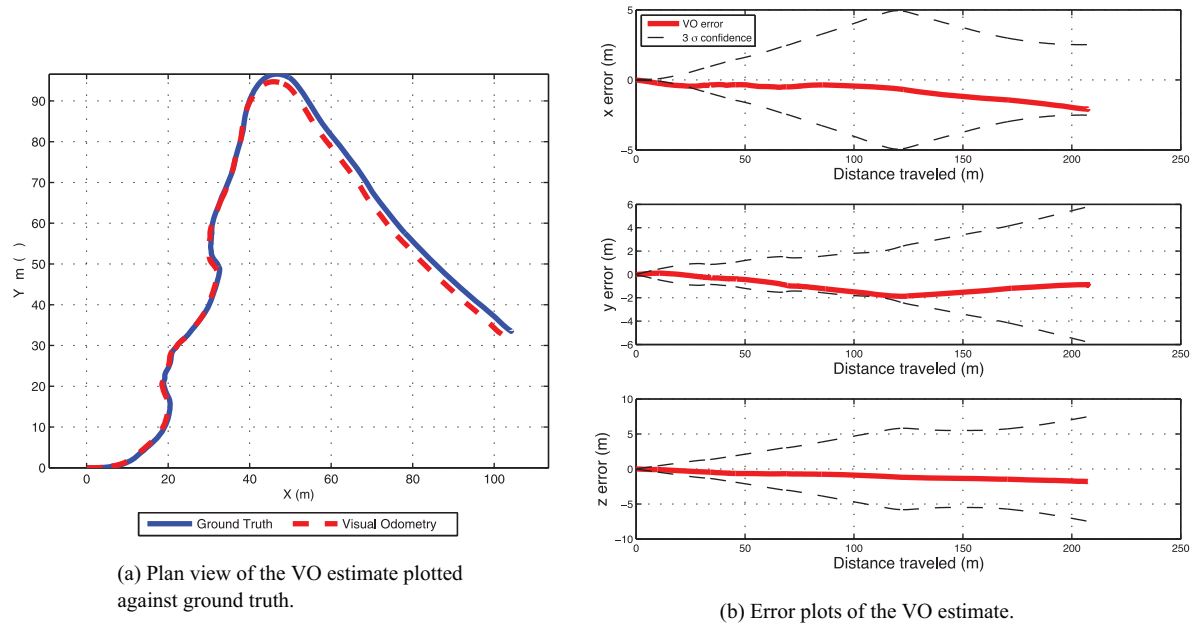
(a) Plan view of the VO estimate plotted against ground truth.

(b) Error plots of the VO estimate.

Fig. 15. VO results for sequence *sri-ss1*.

(a) Plan view of the VO estimate plotted vagainst ground truth.
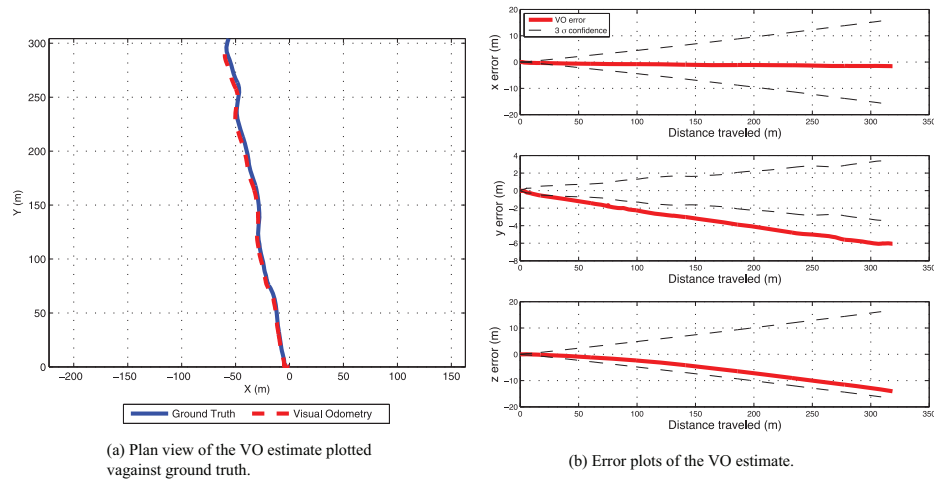
(b) Error plots of the VO estimate.

Fig. 16. VO results for sequence *sri-ss2*.

a sun sensor/inclinometer pair to increase the accuracy of VO without significantly increasing the amount of memory and computation required.

(a) Plan view of the VO estimate plotted against ground truth.
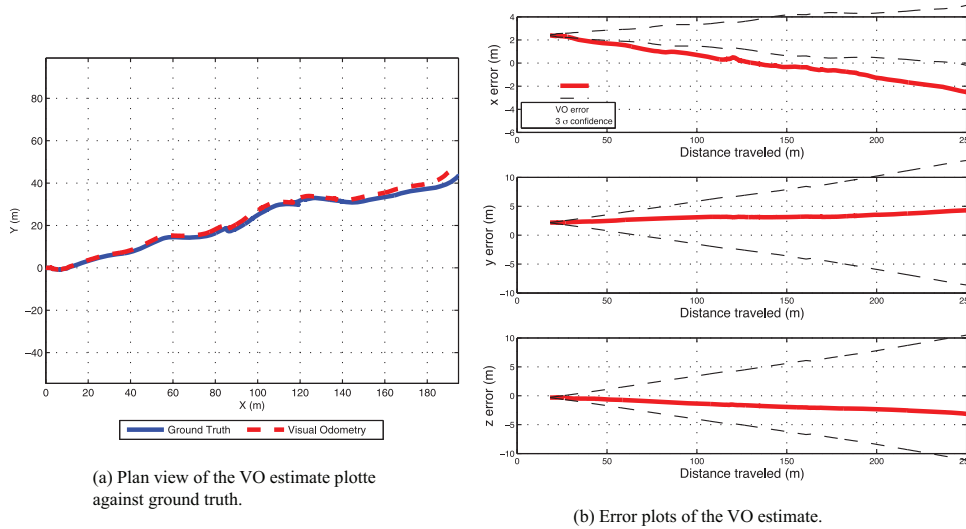
(b) Error plots of the VO estimate.

Fig. 17. VO results for sequence *sri-ss3*.

Institute provided the excellent dataset used in Appendix B. Our algorithm used the SURF library developed by Herbert Bay, Luc Van Gool, and Tinne Tuytelaars and available at http://www.vision.ee.ethz.ch/~surf/. Finally, at the University of Toronto Institute for Aerospace Studies, Konstantine Tsotsos and Peter Miras provided logistical support, and Rehman Merali pitched in during assembly of our pushcart rover.

A. Index to Multimedia Extensions

The multimedia extension page is found at http://www.ijrr.org

Table of Multimedia Extensions

Extension	Media Type	Description
1	Video	Overview of the project. Videos of GPR data collection, stereo data collection, raw stereo and GPR data, and data products produced.

B. Further Visual Odometry Results

We have performed further testing of our VO algorithm on the *Dunes* dataset, provided to us by Kurt Konolige of the Stanford Research Institute. The dataset consists of approximately 42, 000 rectified, grayscale, stereo image pairs at 512×384 pixels resolution. The camera had a baseline of 0.52 m and a horizontal field of view of $65°$.

As we are primarily interested in VO for planetary rovers, three sub-sequences that had a minimum amount of vegetation were selected for testing. The VO results from the three subsequences are plotted in Figures 15–17 and summarized in Table 2.

Table 2. VO Motion Estimate Results on the *Dunes* Dataset

Transect	Length (m)	Linear Error (m)	% Error	Start Frame	Images
sri-ss1	207.36	2.88	1.39%	182,385	848
sri-ss2	318.60	15.42	4.84%	185,558	1,105
sri-ss3	213.89	5.63	2.63%	184,325	814

References

Aggarwal, J. and Nandhakumar, N. (1988). On the computation of motion from sequences of images: A review. *Proceedings of the IEEE*, 76: 917–935.

Annan, A. (2009). Electromagnetic principles of ground penetrating radar. *Ground Penetrating Radar Theory and Applications*, Jol, H. M. (ed.). Elsevier, Amsterdam, pp. 1–40.

Annan, A. and Davis, J. (1976). Impulse radar sounding in permafrost. *Radio Science*, 11: 383–394.

Arcone, S., Lawson, D. and Delaney, A. (1995). Short-pulse radar wavelet recovery and resolution of dielectric constants within englacial and basal ice of Matanuska glacier, Alaska, USA. *Journal of Glaciology*, 41: 68–86.

Arcone, S., Prentice, M. and Delaney, A. (2002). Stratigraphic profiling with ground-penetrating radar in permafrost: A review of possible analogs for Mars. *Journal of Geophysical Research*, 107(E11): 5108.

Bakambu, J. N., Allard, P. and Dupuis, E. (2006). 3D terrain modeling for rover localization and navigation. *Computer and Robot Vision, Canadian Conference*, p. 61.

Barfoot, T., D'Eleuterio, G. and Annan, P. (2003). Subsurface surveying by a rover equipped with ground penetrating radar. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Las Vegas, NV, pp. 2541–2546.

Barfoot, T., Se, S. and Jasobedzki, P. (2006). Visual motion estimation and terrain modelling for planetary rovers. *Intelligence for Space Robotics*, Ch. 4. TSI Press, Albuquerquem, NM.

Barfoot, T. D. (2005). Online visual motion estimation using fastslam with sift features. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 579–585.

Barfoot, T. D., Furgale, P. T., Osinski, G. R., Ghafoor, N. and Williams, K. (2009). Field testing of robotic technologies to support ground ice prospecting in martian polygonal terrain. *Planetary and Space Science*. To appear in the special issue on "Exploring other worlds by exploring our own: The role of terrestrial analogue studies in planetary exploration".

Baumgartner, E. T., Leger, P. C., Schenker, P. S. and Huntsberger, T. L. (1998). Sensor-fused navigation and manipulation for a planetary rover. *Proceedings of SPIE* 3523.

Bay, H., Tuytelaars, T. and Van Gool, L. (2006). SURF: Speeded up robust features. *9th European Conference on Computer Vision*, Graz, Austria.

Benoit, S. and Ferrie, F. (2007). Towards direct recovery of shape and motion parameters from image sequences. *Computer Vision and Image Understanding*, 105(2): 145–165.

Biesiadecki, J. and Maimone, M. (2006). The Mars Exploration Rover surface mobility flight software: driving ambition. *2006 IEEE Aerospace Conference*, p. 15.

Blais, F. (2004). Review of 20 years of range sensor development. *Journal of Electronic Imaging*, 13: 231–243.

Busby, J. P. and Merritt, J. W. (1999). Quaternary deformation mapping with ground penetrating radar. *Journal of Applied Geophysics*, 41(1): 75–91.

Cassidy, N. J. (2009). Ground penetrating radar data processing, modelling and analysis. *Ground Penetrating Radar Theory and Applications*, Jol, H. M. (ed.). Elsevier, Amsterdam, pp. 141–176.

Comport, A. I., Malis, E. and Rives, P. (2010). Real-time quadrifocal visual odometry. *The International Journal of Robotics Research*.

Corke, P., Strelow, D. and Singh, S. (2004). Omnidirectional visual odometry for a planetary rover. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) 2004*, Vol. 4, pp. 4007–4012.

Daniels, D. (ed.) (2004). *Ground Penetrating Radar*. Peter Peregrinus Ltd.

Degenhardt, J. J. and Giardino, J. R. (2003). Subsurface investigation of a rock glacier using ground-penetrating radar: Implications for locating stored water on Mars. *Journal of Geophysical Research (Planets)*, 108: 8036.

dePascale, G., Pollard, W. and Williams, K. (2008). Geophysical mapping of ground ice using a combination of capacitive coupled resistivity and ground-penetrating radar, Northwest Territories, Canada. *Journal of Geophysical Research*, 113: F02S90.

Fiore, P. (2001). Efficient linear solution of exterior orientation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2): 140–148.

Fischler, M. A. and Bolles, R. C. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6): 381–395.

Folkesson, J. and Christensen, H. (2007). Sift based graphical slam on a packbot. *International Conference on Field and Service Robotics (FSR)*.

Fong, T., Allan, M., Bouyssounouse, X., Bualat, M., Deans, M., Edwards, L., Fluckiger, L., Keely, L., Lee, S., Lees, D., To, V. and Utz, H. (2008). Robotics site survey at Haughton Crater.

Proceedings of the 9th International Symposium on Artificial Intelligence, Robotics and Automation in Space (iSAIRAS), Los Angeles, CA.

Fong, T., Deans, M., Lee, P. and Bualat, M. (2007). Simulated lunar robotic survey at terrestrial analog sites. *Proceedings of the 38th Lunar and Planetary Science Conference*, League City, TX.

Förstner, W. and Gulch, E. (1987). A fast operator for detection and precise location of distinct points, corners and centres of circular features. *Proceedings of ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data*, Interlaken, Switzerland, pp. 281–305.

Fortier, D. and Allard, M. (2004). Late holocene syngentic ice-wedge polygons development, Bylot Island, Canadian Arctic Achipelago. *Canadian Journal of Earth Sciences*, 41: 997–1012.

Freese, M., Matsuzawa, T., Oishi, Y., Debenest, P., Takita, K., Fukushima, E. F. and Hirose, S. (2007). Robotics-assisted demining with gryphon. *Advanced Robotics*, 21(15): 1763–1768.

Furgale, P. and Barfoot, T. (2010a). Stereo mapping and localization for long-range path following on rough terrain. *IEEE International Conference on Robotics and Automation, ICRA*, Anchorage, Alaska, USA.

Furgale, P. and Barfoot, T. (2010b). Visual path following on a manifold in unstructured, three-dimensional terrain. *IEEE International Conference on Robotics and Automation, ICRA*, Anchorage, Alaska, USA.

Goldberg, S., Maimone, M. and Matthies, L. (2002). Stereo vision and rover navigation software for planetary exploration. *Aerospace Conference Proceedings, 2002*, Vol. 5, IEEE, pp. 5-2025–5-2036.

Grant, J. A., Schutz, A. E. and Campbell, B. A. (2003). Ground-penetrating radar as a tool for probing the shallow subsurface of Mars. *Journal of Geophysical Research*, 108(E4).

Harris, C. and Stephens, M. (1988). A combined corner and edge detector. *Proceedings of The Fourth Alvey Vision Conference*, pp. 147–151.

Hartley, R. I. and Zisserman, A. (2000). *Multiple View Geometry in Computer Vision*. Cambridge, Cambridge University Press.

Hinkel, K., Doolittle, J., Bockheim, J., Nelson, F., Paetzold, R., Kimble, J. and Travis, R. (2001). Detection of subsurface permafrost features with ground-penetrating radar, Barrow, Alaska. *Permafrost and Periglacial Processes*, 12: 179–190.

Howard, A. (2008). Real-time stereo visual odometry for autonomous ground vehicles. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) 2008*, pp. 3946–3952.

Johnson, A., Goldberg, S., Cheng, Y. and Matthies, L. (2008). Robust and efficient stereo feature tracking for visual odometry. *IEEE International Conference on Robotics and Automation (ICRA) 2008*, pp. 39–46.

Kim, S. S., Carnes, S., Haldemann, A., Ho Wah Ng, E., Ulmer, C. and Arcone, S. (2006). Miniature ground penetrating radar, crux GPR. *Proceedings 2002 IEEE Aerospace Conference*.

Konolige, K., Agrawal, M. and Solà, J. (2007). Large scale visual odometry for rough terrain. *Proceedings of the International Symposium on Research in Robotics (ISRR)*.

Kuzmin, R. and Zabalueva, E. (2003). Polygonal terrains on Mars: Preliminary results of global mapping of their spatial distribution. *Lunar and Planetary Science XXXIV*, Abstract 1912.

Lachenbruch, A. (1962). Mechanics of thermal contraction cracks and ice-wedge polygons in permafrost. *Special paper to the Geological Society of America*, 70: 69.

Lacroix, S., Mallet, A. and Bonnafous, D. (2001). Autonomous rover navigation on unknown terrains functions and integration. *Experimental Robotics VII* (*Lecture Notes in Control and Information Sciences*, Vol. 271). Springer, Berlin, pp. 501–510.

Lee, P., Braham, S., Boucher, M., Schutt, J., Glass, B., Gross, A., Hine, B., McKay, C., Hoffman, S., Jones, J., Berinstain, A., Comptois, J.-M., Hodgson, E. and Wilkinson, N. (2007). Haughton-Mars project: 10 years of science operations and exploration systems development at a Moon/Mars analog site on Devon Island, high arctic. *Proceedings of the 38th Lunar and Planetary Science Conference*, League City, TX, pp. 2426–2427.

Lehmann, F. and Green, A. G. (1999). Semiautomated georadar data acquisition in three dimensions. *Geophysics*, 64(3): 719–731.

Leuschen, C., Kanagaratnam, P., Yoshikawa, K., Arcone, S. and Gogineni, P. (2002). Field experiments of a surface-penetrating radar for Mars. *IEEE International Geoscience and Remote Sensing Symposium, (IGARSS'02)*, Vol. 6, pp. 3579–3581.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110.

Lunt, I. A. and Bridge, J. S. (2004). Evolution and deposits of a gravelly braid bar, Sagavanirktok River, Alaska. *Sedimentology*, 51(3): 415–432.

Mackay, J. (1999). Periglacial features developed on the exposed lake bottoms of seven lakes that drained rapidly after 1950, Tuktoyaktuk peninsula area, western Arctic coast, Canada. *Permafrost and Periglacial Processes*, 10: 39–63.

Mackay, J. (2000). Thermally induced movement in ice wedge polygons, western arctic coast: a long-term study. *Geographie Physique et Quaternaire*, 54(1): 41–68.

Maimone, M., Cheng, Y. and Matthies, L. (2007). Two years of visual odometry on the Mars exploration rovers. *Journal of Field Robotics*, 24(3): 169–186.

Maki, J. N., III, J. F. B., Herkenhoff, K. E., Squyres, S. W., Kiely, A., Klimesh, M., Schwochert, M., Litwin, T., Willson, R., Johnson, A., Maimone, M., Baumgartner, E., Collins, A., Wadsworth, M., Elliot, S. T., Dingizian, A., Brown, D., Hagerott, E. C., Scherr, L., Deen, R., Alexander, D. and Lorre, J. (2003). Mars exploration rover engineering cameras. *Journal of Geophysical Research*, 108(E12).

Malin, M. C., Bell, III, J. F., Cameron, J., Dietrich, W. E., Edgett, K. S., Hallet, B., Herkenhoff, K. E., Lemmon, M. T., Parker, T. J., Sullivan, R. J., Sumner, D. Y., Thomas, P. C., Wohl, E. E., Ravine, M. A., Caplinger, M. A. and Maki, J. N. (2005). The Mast Cameras and Mars Descent Imager (MARDI) for the 2009 Mars Science Laboratory, *Lunar and Planetary Institute Technical Report* Mackwell, S. and Stansbery, E. (eds), *36th Annual Lunar and Planetary Science Conference*.

Mallet, A., Lacroix, S. and Gallo, L. (2000,). Position estimation in outdoor environments using pixel tracking and stereovision. In *IEEE International Conference on Robotics and Automation*, Vol. 4, pp. 3519–3524.

Mangold, N., Maurice, S., Feldman, W., Costard, F. and Forget, F. (2004). Spatial relationships between patterned ground and ground ice detected by the neutron spectrometer on Mars. *Journal of Geophysical Research*, 109: E08001.

Matthies, L. (1989). *Dynamic stereo vision*. PhD Thesis, Carnegie Mellon University Computer Science Department.

Matthies, L., Maimone, M., Johnson, A., Cheng, Y., Willson, R., Villalpando, C., Goldberg, S., Huertas, A., Stein, A. and Angelova, A. (2007). Computer vision on Mars. *International Journal of Computer Vision*, 75(1): 67–92.

Matthies, L. and Shafer, S. (1987). Error modeling in stereo navigation. *IEEE Journal of Robotics and Automation*, 3(3): 239–248.

Mellon, M. (1997). Small-scale polygonal features on Mars: Seasonal thermal contraction cracks in permafrost. *Journal of Geophysical Research*, 102(E11): 617–25, 628.

Milsom, J. (2003). *Field Geophysics*, 3rd edn. Wiley, New York.

Moorman, B. and Michel, F. (2000). Glacial hydrological system characterization using ground-penetrating radar. *Hydrological Processes*, 14: 2645–2667.

Moorman, B. J., Robinson, S. D. and Burgess, M. M. (2003). Imaging periglacial conditions with ground-penetrating radar. *Permafrost and Periglacial Processes*, 14(4): 319–329.

Moravec, H. (1980). *Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover*. PhD Thesis (Technical Report Stanford AIM-340), Stanford University. Republished as Technical Report CS-80-813, Carnegie Mellon University Robotics Institute.

Nistér, D. (2003). An efficient solution to the five-point relative pose problem. *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '03)*, Vol. 2, p. 195.

Nistér, D., Naroditsky, O. and Bergen, J. (2004). Visual odometry. *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 1, pp. 652–659.

Nistér, D., Naroditsky, O. and Bergen, J. (2006). Visual odometry for ground vehicle applications. *Journal of Field Robotics*, 23(1): 3.

Nistér, D. and Stewenius, H. (2006). Scalable recognition with a vocabulary tree. *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '06)*, Washington, DC, USA, pp. 2161–2168.

Olson, C. F., Matthies, L. H., Schoppers, M. and Maimone, M. W. (2003). Rover navigation using stereo ego-motion. *Robotics and Autonomous Systems*, 43(4): 215–229.

Olson, C. F., Matthies, L. H., Wright, J. R., Li, R. and Di, K. (2007). Visual terrain mapping for Mars exploration. *Computer Vision and Image Understanding*, 105(1): 73–85.

Osinski, G. R., Barfoot, T. D., Ghafoor, N., Izawa, M., Bangerjee, N., Jasiobedzki, P., Tripp, J., Richards, R., Auclair, S., Sapers, H., Thomson, L. and Flemming, R. (2010). Lidar and the mobile scene modeler (mSM) as scientific tools for planetary exploration. *Planetary and Space Science*, 58(4 Special Issue): 691–700.

Pedersen, L., Smith, D., Deans, M., Sargent, R., Kunz, C., Lees, D. and Rajagopalan, S. (2005). Mission planning and target tracking for autonomous instrument placement. *2005 IEEE Aerospace Conference*, pp. 34–51.

Pettinelli, E., Burghignoli, P., Pisani, A. R., Ticconi, F., Galli, A., Vannaroni, G. and Bella, F. (2007). Electromagnetic propagation of GPR signals in martian subsurface scenarios including material losses and scattering. *IEEE Transactions on Geoscience and Remote Sensing*, 45(5): 1271–1281.

Picardi, G., Plaut, J. J., Biccari, D., Bombaci, O., Calabrese, D., Cartacci, M., Cicchetti, A., Clifford, S. M., Edenhofer, P., Farrell, W. M., Federico, C., Frigeri, A., Gurnett, D. A., Hagfors, T., Heggy, E., Herique, A., Huff, R. L., Ivanov, A. B., Johnson, W. T. K., Jordan, R. L., Kirchner, D. L., Kofman, W., Leuschen, C. J., Nielsen, E., Orosei, R., Pettinelli, E., Phillips, R. J., Plettemeier, D., Safaeinili, A., Seu, R., Stofan, E. R., Vannaroni, G., Watters, T. R. and Zampolini, E. (2005). Radar soundings of the subsurface of Mars. *Science*, 310(5756): 1925–1928.

Rial, F. I., Pereira, M., Lorenzo, H. and Arias, P. (2005). Acquisition and synchronism of GPR and GPS data: Application on road evaluation. *Image and Signal Processing for Remote Sensing XI*, Bruzzone, L. (ed.), *Proceeedings of SPIE*, 5982: 598219.1–598219.8.

Ross, N., Harris, C., Christansen, H. and Brabham, P. (2005). Ground penetrating radar investigations of open-system pingos, Adventdalen, Svalbard. *Norsk Geografisk Tidsskrift—Norwegian Journal of Geography*, 59: 129–138.

Roth, G. and Wibowoo, E. (1997). An efficient volumetric method for building closed triangular meshes from 3-D image and point data. *Proceedings of the Conference on Graphics Interface '97*. Toronto, ON, Canadian Information Processing Society, pp. 173–180.

Scott, W., Sellman, P. and Hunter, J. (1990). Geophysics in the study of permafrost, in geotechnical and environmental geophysics. *Geotechnical and Environmental Geophysics*, Ward, S. (ed.) (*Investigations in Geophysics*, Vol. 5). Society of Exploration Geophysicists.

Se, S. and Jasiobedzki, P. (2008). Stereo-vision based 3D modeling and localization for unmanned vehicles. *International Journal of Intelligent Control and Systems*, 13(1 Special Issue): 47–58.

Se, S., Lowe, D. G. and Little, J. J. (2002). Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks. *International Journal of Robotics Research*, 21(8): 735–758.

Seu, R., Phillips, R. J., Biccari, D., Orosei, R., Masdea, A., Picardi, G., Safaeinili, A., Campbell, B. A., Plaut, J. J., Marinangeli, L., Smrekar, S. E. and Nunes, D. C. (2007). Sharad sounding radar on the Mars reconnaissance orbiter. *Journal of Geophysics Research*, 112(E5).

Sibley, G., Matthies, L., and Sukhatme, G. (2008). A sliding window filter for incremental slam. *Unifying Perspectives in Computational and Robot Vision*, (*Lecture Notes in Electrical Engineering*, Vol. 8). New York, Springer, pp. 103–112.

Strelow, D. and Singh, S. (2004). Motion estimation from image and inertial measurements. *The International Journal of Robotics Research*, 23(12): 1157–1195.

Sünderhauf, N., Konolige, K., Lacroix, S. and Protzel, P. (2006). Visual odometry using sparse bundle adjustment on an autonomous outdoor vehicle. *Autonome Mobile Systeme 2005*. Berlin, Springer, pp. 157–163.

Thrun, S., Burgard, W. and Fox, D. (2001). *Probabilistic Robotics (Intelligent Robotics and Autonomous Agents)*. Cambridge, MA, The MIT Press.

Torr, P. H. S. (2002). Bayesian model estimation and selection for epipolar geometry and generic manifold fitting. *International Journal of Computer Vision*, 50(1): 35–61.

Torr, P. H. S. and Zisserman, A. (2000). MLESAC: A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding*, 78(1): 138–156.

Triggs, B., McLauchlan, P., Hartley, R. and Fitzgibbon, A. (2000). Bundle adjustment – A modern synthesis. *Vision Algorithms: Theory and Practice*, Triggs, W., Zisserman, A. and Szeliski, R., (eds) (*Lecture Notes in Computer Science*). Berlin, Springer, pp. 298–375.

Vago, J., Gardini, B., Kminek, G., Baglioni, P., Gianfiglio, G., Santovincenzo, A., Bayón, S. and van Winnendael, M. (2006). ExoMars - searching for life on the Red Planet. *ESA Bulletin*, 126: 16–23.

Wettergreen, D., Dias, M., Shamah, B., Teza, J., Tompkins, P., Urmson, C., Wagner, M. and Whittaker, W. (2002). First experiment in sun-synchronous exploration. *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Washington, DC, pp. 3501–3507.

Wettergreen, D., Tompkins, P., Urmson, C., Wagner, M. and Whittaker, W. (2005). Sun-synchronous robotic exploration: Technical description and field experimentation. *The International Journal of Robotics Research*, 24(1): 3–30.

Williams, K., Grant, J. and Schutz, A. (2005). Ground-penetrating radar in Mars analog terrains: Testing the strata instrument. *Workshop on Radar Investigations of Planetary and Terrestrial Environments*, Houston, TX.

Wright, J., Trebi-Ollennu, A., Hartman, F., Cooper, B., Maxwell, S., Yen, J. and Morrison, J. (2005). Terrain modelling for in-situ activity planning and rehearsal for the Mars exploration rovers. *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, vol. 2, pp. 1372–1377.

Zhang, Z. (1997). A stereovision system for a planetary rover: Calibration, correlation, registration, and fusion. *Machine Vision and Applications*, 10(1): 27–34.