Global Rover Localization by Matching Lidar and Orbital 3D Maps

Patrick J.F. Carle and Timothy D. Barfoot
Institute for Aerospace Studies
University of Toronto
Toronto, Canada M3H 5T6
{pat.carle, tim.barfoot}@utoronto.ca

Abstract—Current rover localization techniques such as visual odometry have proven to be very effective on short to medium-length traverses (e.g., up to a few kilometres). This paper deals with the problem of long-range rover localization (e.g., 10km and up). An autonomous method to globally localize a rover is proposed by matching features detected from a 3D orbital elevation map and rover-based 3D lidar scans. The accuracy and efficiency of the algorithm is enhanced with visual odometry, and inclinometer/sun-sensor orientation measurements. The methodology was tested with real data, including 37 lidar scans of terrain from a Mars-Moon analogue site on Devon Island, Nunavut. When a scan contained a sufficient number of good topographic features, localization produced position errors of no more than 100m, and as low as a few metres in many cases. On a 10km traverse, the developed algorithm’s localization estimates were shown to significantly outperform visual odometry estimates. It is believed that this architecture could be used to accurately and autonomously localize a rover on long-range traverses.

I. INTRODUCTION

The ongoing Mars Exploration Rover (MER) missions have proven to be historic landmarks in space exploration. However, they are also humbling reminders of the challenges ahead. For example, the MER Opportunity has operated on Mars for over 5 years now, but has only driven a total of about 20km due to mechanical/energy limitations and a lack of autonomy [1]. An important goal for future generations of rovers will be to overcome these deficiencies to allow them to explore sites hundreds of kilometres away from their landers [2]. Rovers will consequently require an autonomous long-range localization system to aid them in their journey.

Currently, a rover employs a variety of techniques to determine its pose at any given time. The MERs were first localized with radio tracking [3], descent trajectory modeling, and by comparing orbital to ground camera imagery [1]. After leaving their landers, localization has been accomplished primarily with dead-reckoning techniques such as wheel odometry, visual odometry (VO) and local bundle adjustment (BA). Wheel odometry is not computationally intensive, but is vulnerable to sensor noise and mechanical disturbances (e.g., wheel slippage) [4]. Computer vision techniques, such as VO and BA, complement wheel odometry when needed.

VO is automated and can work in real-time, but is computationally very demanding. It has yielded impressive results in the past with error as low as 0.1% over a 10km traverse [5]. BA can offer further gains in accuracy [6], but efforts to automate the process are ongoing [7]. Despite significant advances in the technology, such dead-reckoning approaches are not suitable for long-range localization (e.g., more than 10km) since they will always exhibit unbounded error growth with distance traversed [8].

Global localization techniques can be used to correct dead-reckoning pose estimates once these become unreliable. On Earth, the Global Positioning System (GPS) is commonly used for this purpose. However, the satellite infrastructure required for such a system is not feasible for non-Earth applications. This paper proposes an alternate solution that aligns a rover-based three-dimensional (3D) local map to a satellite-based 3D global map. As shown in Figure 1.

In this research, the local map is a point cloud obtained from a time-of-flight lidar (Light Detection and Ranging). This instrument can measure distance to far-away objects by rapidly firing a laser and measuring the time for reflected beams to return. In a surveying configuration, a lidar can sample terrain with centimetre-accuracy at a range of up to 1.5km, making it a vital guidance and navigation sensor. Lidars have been tested in numerous applications on Earth [9], [10] and in space [11], [12].

The global map may be acquired from a satellite-based laser altimeter (e.g., LOLA, MOLA2), or by extracting 3D information from a stereo pair of high-resolution satellite images (e.g., LROC, HiRISE) [13]. Current satellites have extensive and accurate coverage of the Moon and Mars (see

Fig. 1. Matching rover’s 3D map (bottom) to a 3D orbital map (top).
Table I). Therefore, relevant map data could be loaded into a rover even before it begins its mission. This would allow the rover to autonomously localize without requiring any additional map data from Earth.

**Table I**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Moon</td>
<td>LOLA [14]</td>
<td>Total</td>
<td>50-100m</td>
<td>10cm</td>
</tr>
<tr>
<td></td>
<td>LROC [14]</td>
<td>&gt;10%</td>
<td>50cm</td>
<td>NA</td>
</tr>
<tr>
<td>Mars</td>
<td>MOLA2 [15]</td>
<td>Total</td>
<td>100m</td>
<td>1m</td>
</tr>
<tr>
<td></td>
<td>HiRISE [16], [17]</td>
<td>2%</td>
<td>&lt;1m</td>
<td>&lt;1m</td>
</tr>
</tbody>
</table>

Having obtained local and global maps, a means of matching them is then required. In feature-based matching, interest points are first extracted from the global and local maps, and then matched in search of global-local feature correspondences. If at least three, non-collinear feature correspondences are found, a rigid 3D transformation can be expressed that aligns the two maps [18].

The most prominent features visible to both a low-resolution orbital map and a high-resolution lidar map are topographic peaks. Since one side of a peak will always be hidden from the rover, features from the lidar map will only be partially describable as shown in Figure 1. It will therefore be difficult to compare features between the two maps if a descriptor-based approach is used, such as spin-image matching [19]. A more convenient alternative is to search for similar feature constellations, where the spacing between features now effectively acts as the descriptor. This method is known as the Data-Aligned Rigidity-Constrained Exhaustive Search (DARCES) [20].

Previous work has investigated these 3D matching techniques in the context of localization [10], [9]. Localization has been attempted with VO and BA on a network of 2D images and orbital data with some success [7]. However, automation is problematic. The VIPER algorithm [21] matches the skyline in a panoramic image to predicted skylines at various positions on a 3D global map. The final solution is not very accurate (>100m).

This paper investigates a novel global localization algorithm using 3D data. The overall architecture is outlined in Section II, followed by a description of the methodology in Sections III through V. Sections VI to VIII present results from field tests at a Mars/Moon analogue site in the Canadian Arctic.

### II. ARCHITECTURE

The architecture is developed for the general case where a rover traverses over some distance and occasionally stops to scan the terrain with the lidar as shown in Figure 2. The general formulation can also be simplified by omitting odometry to examine single-scan localization.

The goal is to determine the rover’s pose at each scan site with respect to the global map’s reference frame, \( \mathcal{F}_g \). The rover pose at scan site \( \ell \) is defined as a transformation, \( T_{\ell o} := \{ t_{\ell o}, S_{\ell o} \} \), from \( \mathcal{F}_o \) to the rover’s local frame \( \mathcal{F}_\ell \), where \( t_{\ell o} \) and \( S_{\ell o} \) are, respectively, the translation vector and rotation matrix from \( \mathcal{F}_o \) to \( \mathcal{F}_\ell \). An overview of the procedure is presented in Figure 3 and is summarized as follows:

(a) **Feature Detection**: Features are detected from the global and local maps.

(b) **Feature Matching**: DARCES obtains global-local feature correspondences and an initial estimate of the rover’s poses.

(c) **Pose Refinement**: Global-local feature correspondences are used in combination with orientation and odometry measurements to refine rover pose estimates.

The sections to follow discuss these in further detail.

### III. FEATURE DETECTION

Features must first be detected (box (a) in Figure 3) from the global map (the global features) and from the local map (the local features). It is assumed a global elevation map is given with resolution \( L_{xy} \), as well as one or more local elevation maps. Generally, the most prominent features common to both maps are topographic peaks. These peaks are detected using a local maxima detector based on morphological dilation\(^*\) [22], [23]. Other feature detectors could also be used within the presented framework.

The scan is leveled using pitch and roll measurements (e.g., inclinometer). This ensures the \( +z \)-direction of the global and local maps roughly coincide. The local map is gridded to the global map resolution with nearest-neighbour interpolation. This speeds up the detection process and ensures the scale of global and local features is the same.

Morphological dilation replaces lower grid values with neighbouring higher grid values, effectively blurring out low elevations. Once dilation is completed for all points on the grid, the blurred map is compared to the original map. Cells with no change in value are interpreted as local maxima. The dilation window is chosen to be a pixelated circle (Figure 4) to make the window’s coverage more uniform in all directions. The radius of this circle limits the size of the detected features, as well as the distance between features. The minimum distance between features, \( D_{\text{detect}} \), depends on the global map resolution, \( L_{xy} \), such that \( D_{\text{detect}} := n \times L_{xy} \) where \( n \) is the circle’s cell radius.

\(^*\)This was inspired by code found on the Matlab Central repository as ‘localMaximum.m’ by Yonathan Nativ.
Features must then undergo final processing to eliminate false peaks that might be detected in flat areas. This is done by ensuring the minimum distance between any two features is $D_{\text{detect}}$ and by testing areas for flatness. Sample feature detection results are shown in Figure 5.

A good uncertainty model is necessary to correctly assess the quality of individual features and global-local matches. Global feature positional uncertainty is assumed equal to the position uncertainty of a measured 3D point in the global map. The positional uncertainty of a measured 3D point from the lidar is comparatively very small ($<1\text{m}$) even for measurements far from the lidar’s origin. However, the effect of occlusions will dominate the uncertainty in local feature measurements. Local feature uncertainty is therefore better estimated by the size of a local feature, $D_{\text{detect}}$.

**IV. FEATURE MATCHING**

The feature matching methodology (box (b) in Figure 3) is based on the DARCES algorithm [20]. Lidar scans are processed individually.

**A. Hypothesis Search**

Hypothesized correspondences must first be generated between global and local features. A hypothesis is defined as a group of possible correspondences between three unique local features, called control points, and three unique global features. A hypothesis is not guaranteed to be correct due to noise in feature position measurements. Therefore, many control point groups are tested to increase the chances of finding a valid hypothesis.

A hypothesis is generated if the distances between three global features are similar to the distances between three local features. This test can be quantified knowing each feature’s position uncertainty.

**B. Hypothesis Evaluation**

The validity of a hypothesis is evaluated based on the transformation it produces between the global and local frame. For a hypothesis $i$, the transformation from $\mathcal{F}_o$ to $\mathcal{F}_l$, $T_{i,o} := \{ t_{i,o}^l, S_{i,o}^l \}$, can be obtained using a least-squares point-alignment algorithm to align the three global and three local features that comprise a hypothesis [24].

To improve the efficiency and robustness of DARCES, hypotheses are first screened with a number of simple tests. A map-boundary test rejects hypotheses that estimate the rover’s position outside the global map. A $\gamma$-deviation test ensures that an estimated $z$-position is within $E_{\gamma\text{dev}}$ of the global map’s elevation at the estimated $xy$-position. Finally, orientation tests check that estimates and measurements (e.g., inclinometer, sun sensor) of roll, $\alpha$, pitch, $\beta$, and heading, $\gamma$, differ by no more than $E_\alpha$, $E_\beta$ and $E_\gamma$. These thresholds are set to three standard deviations of the involved measurement’s uncertainty to ensure that a high proportion of valid hypotheses fall within the thresholds. For example, $E_{\gamma\text{dev}} := 3\sigma_{Gz}$, where $\sigma_{Gz}$ is one standard deviation of the global map $z$-position uncertainty. These tests are not theoretically necessary, but are practically very helpful in speeding up the search for valid hypotheses.

A measure of fitness is then calculated for each hypothesis by examining how well the hypothesized transformation aligns lidar scan points to the global map. To more efficiently and robustly calculate this metric, the full lidar scan is decimated to half the global map resolution, $L_{x,y}/2$. These more evenly-spread points are called the reference points.

For a hypothesis $i$, the fitness metric, $f_i$, is the average absolute $z$-error between the reference points, transformed to the global frame with $T_{i,o}$, and the global map:

$$f_i := -\frac{1}{N_{\text{ref}}} \sum_{j=1}^{N_{\text{ref}}} \left| z_{i,j}^g - z_{i,j}^G \right|,$$

(1)

where $N_{\text{ref}}$ is the number of reference points, $z_{i,j}^r$ is the $z$-position of the transformed reference point $j$ in the global frame, and $z_{i,j}^G$ is the interpolated global map elevation at the $xy$-position of the transformed reference point $j$. The negative is applied so that a low error corresponds to a high fitness.

**C. Hypothesis Selection**

Once a fitness is associated with each hypothesis, a search is made for valid hypotheses, which are defined as a group of
high-fitness hypotheses that have similar position estimates. The hypothesis with the highest fitness in this valid group is then returned along with its associated transformation estimate and global-local feature correspondences. If no valid group is found after all combinations of control points have been exhausted, then DARCES returns no solution.

V. POSE REFINEMENT

In the pose refinement stage, outlier feature correspondences are first rejected with RANSAC (Random Sample Consensus) [25]. The remaining, inlier feature correspondences are then combined with odometry and orientation measurements into a Simultaneous Localization And Mapping (SLAM) problem to refine pose estimates (box c) in Figure 3). The Multi-frame Odometry-compensated Global Alignment (MOGA) algorithm, solves this problem by minimizing the errors between all available measurements and desired estimates. This is essentially a batch SLAM algorithm that fuses relative and absolute pose measurements over an entire rover traverse. MOGA will be summarized here, but full details can be found from [26].

The main goal is to estimate the transformations from the global frame, $\mathbb{F}_o$, to each of the $M$ local frames, $\mathbb{F}_i$. This $M$-frame alignment problem is depicted in Figure 6. Input measurements are assumed to be corrupted with white, zero-mean, Gaussian noise. The four types of measurement are global feature positions $(r_{f\ell}^i)$, local feature positions $(r_{f\ell}^i)$, odometry measurements of the rotation $(\rho_{O\ell+1\ell}^i Q_{O\ell+1\ell}^i)$, and translation $(\rho_{O\ell+1\ell}^i Q_{O\ell+1\ell}^i)$, associated with these measurements are, respectively, $R_{O\ell+1\ell}^i$, $Q_{O\ell+1\ell}^i$, $(C_{O\ell+1\ell}^i)$, and $J_{O\ell+1\ell}^i$, Orientation Measurements:

$$J_{O\ell+1\ell}^i := \sum_{\ell} J_{fj}^i(z) + \sum_{\ell} J_{O\ell+1\ell}^i(z) + \sum_{\ell} J_{O\ell+1\ell}^i(z),$$

where maximum likelihood is assured by weighting each term with the inverse covariance of the involved measurement. For example, the objective function contribution of global feature $F_j$ is $J_{fj}^i(z) := \frac{1}{2} e_{fj}(z)^T \mathbf{R}_j^{-1} e_{fj}(z)$, where $\mathbf{e}_{fj}(z)$ is the error between the estimated and globally-measured feature positions.

The optimal design parameter column, $z^*$, is sought by minimizing the objective function through unconstrained optimization: $z^* := \text{argmin}_z J(z)$. The Gauss-Newton algorithm [27] is used to solve this nonlinear least-squares problem. When optimizing with Gauss-Newton, the objective function must be quadratic, which is not immediately the case due to the rotations involved. Therefore, it is necessary to linearize the error terms of $J(z)$ at each iteration and solve for $z^*$, giving the direction to the minimum of the local quadratic approximation of the objective function. This solution is used to update the nominal design parameters. Convergence to a local minimum is achieved when the relative change in the objective function is less than $E_{\text{converge}}$.

Single-frame localization can be obtained by dropping the odometry terms, but otherwise proceeding in the same way.

VI. FIELD DATA

A realistic dataset was collected from a Mars/Moon analogue site on Devon Island, Nunavut [28] at 75°22’N and 89°41’W. In total, 37 lidar scans were used to test the single-scan localization performance of the algorithm. A long-range rover traverse was simulated by connecting 23 lidar scans with real odometry data collected by a pushcart outfitted with a variety of rover engineering sensors (sun sensor, inclinometer, stereo cameras). The resulting 10km path is comparable in length to the distance traversed by the MERs to date. The orientation of the lidar and cart were matched by imaging a reference target in both systems [26].

When localizing, the rover position was assumed to be within the bounds of a 100km² global map. The map’s x and y resolutions were, respectively, 13m and 24m. Global feature positional uncertainty was estimated to a radius of 23m for one standard deviation. The local maps were collected with an Optech ILRIS3D-ER lidar, which had a range of about 1.5km, and a positional measurement uncertainty of less than a metre. However, occlusions in the scan increased...
the uncertainty of a local feature’s position to a radius of about 45m for one standard deviation.

The VO algorithm was similar to sparse feature correspondence methods [5]. Accuracy was on par with [29].

Heading measurements for A01-A25 were obtained from a sun sensor [30]. Heading for A26-A37 were computed with similar accuracy knowing the GPS and local positions of a distant target [26]. Roll and pitch were zero since the lidar was leveled before each scan. The uncertainties in heading, roll and pitch were, respectively, $\sigma_\gamma = \sigma_\alpha = \sigma_\beta = 1^\circ$.

Ground-truth $xy$-position measurements were obtained from a consumer-grade GPS. It had positional uncertainty 10m after averaging measurements for about 30 minutes.

VII. SINGLE-FRAME RESULTS

In the single-frame configuration, local frames were localized individually. There was a stochastic aspect to the DARCES algorithm since control points were randomly selected. Therefore, robustness was demonstrated by running 100 randomly-seeded trials on each of the 37 scans. The time to localize a single scan on a Matlab implementation was about 20 min., most of which was spent in DARCES.

Figure 8 shows localization performance for $x$ and $y$ positions. Although the algorithm outputs six-degree-of-freedom localization, only the $x$ and $y$ position performance is discussed since the others are assumed to be well-known from measurements (e.g., heading from the sun sensor). Most trials showed position error well below 50m.

In the 3,700 trials run, the algorithm never produced error greater than 100m. When the algorithm was unconfident about the rover’s position, it produced no DARCES solution. This occurred when few features were within the lidar’s range (flat areas, A10, A13), or when features were too close to the lidar (canyons, A02, A20-A23). In canyon-like settings, nearby hills occlude much of the view, leaving few features for the rover to detect (see Figure 7). Furthermore, a nearby detected peak is more likely to be a poor representation of the true peak, which would be occluded at close range. Future work could investigate how to automatically detect featureless areas when planning a path.

VIII. MULTI-FRAME RESULTS

In this configuration, VO measurements between scans A01 to A23 were available. This allowed for a single, large MOGA optimization to be executed. DARCES transformation estimates served as initial guesses for MOGA. However, if a particular frame did not have a DARCES solution, its transformation was estimated using VO to the next-closest, solved frame. As shown in Figures 9 and 10, MOGA can now produce localization estimates for frames with no DARCES solution using information from all available measurements in the chain of frames. VO estimates are provided in the figures with and without the use of absolute heading measurements (i.e., sun sensor). It should also be made clear that the VO paths made use of ground-truth for frame A01, whereas MOGA used no ground-truth whatsoever.

Note the significant improvement in the estimates as more information is given to the estimator. With absolute heading measurements, VO reduces its error from up to several kilometres to a few hundred metres. With absolute position measurements provided by matched features in MOGA, the error can be further reduced. This improvement can be attributed not only to VO, but also to feature-sharing between frames. Sharing features allows a given frame to use local features from other nearby frames with common global correspondences to improve estimates.

Frames with no DARCES solution had higher error since estimates were more dependent on odometry measurements. Despite these cases, MOGA localized a long-range traverse with no dependence on distance traversed.

IX. CONCLUSION

This research has produced a number of novel contributions. A global localization technique was developed that matches rover-based lidar scans to an orbital elevation map.
ACKNOWLEDGMENTS

Thanks to Paul Furgale for the visual odometry implementation and to Joseph Bakambu for helpful discussions on feature matching. Also thanks to Optech Inc., the Mars Institute, the Canadian Space Agency, the Canada Foundation for Innovation, and the Haughton Mars Project for making our field experiments on Devon Island possible.

REFERENCES