Visual Localization Methods for Mars Rovers using Lander, Rover, and Descent Imagery

Larry H. Matthies, Clark F. Olson, Greg Tharp, and Sharon Laubach
Jet Propulsion Laboratory, California Institute of Technology
4800 Oak Grove Drive, MS 107-102, Pasadena, CA 91109

Abstract

Mars rover sample return missions will require rover localization on scales ranging from under a meter near landers and sample caches, to 10's of meters for local exploration, to 100's of meters for long range exploration. A variety of technologies may contribute to rover localization, but at present there is no concretely defined localization architecture that covers all of these scales. This paper surveys approaches to visual localization using overhead and ground-level imagery and discusses particular capabilities under development at JPL. We conclude that visual localization is (1) reasonably well understood at the smallest scale, where mapping is done with mast-mounted cameras on the lander or rover, (2) reasonably promising at intermediate scales, where mapping is done with descent imagery, and (3) of uncertain potential for the largest scales, where mapping is done from orbit, due to lack of orbital imagery with adequate resolution. This suggests that the largest scales be addressed either with other sensors or by planning for additional, high resolution imaging, especially in stereo, in the area of traverse.

1 Introduction

Mars sample return missions are now being contemplated in which a rover would traverse on the order of 100 kilometers (km), periodically doing detailed explorations of regions a few 10's of meters in diameter [1]. Samples acquired during such a traverse would be cached for pick-up by a subsequent mission, which would land very near the sample cache and would carry a second rover to pick up the cache or perform other contingency operations. Such a mission scenario requires rover localization on scales ranging from under a meter near landers and sample caches, to 10's of meters to explore and map regions, to much larger scales to follow an overall plan for a 100 km mission.

A variety of technologies may contribute to rover localization in the above scenario, but at present there is no concretely defined localization architecture that covers all scales of such a mission. This paper surveys approaches to visual localization with overhead and ground-level imagery and discusses particular capabilities under development at JPL. We start with the smallest scale, by examining the case when the rover stays within view of a lander (section 2), or else operates within a small disk that was first mapped by mast-mounted cameras on the rover (section 3). We review past and present work at JPL and elsewhere that has developed algorithms for these contexts. Although many details remain to be resolved, visual localization on this scale seems to be quite feasible.

We then consider a scenario for localization in a region of several kilometers around a lander, which might be mapped by using a sequence of descent imagery acquired by the lander (section 4). As a specific case study, we look at the descent imaging experiment planned for the Mars Surveyor'98 lander. This experiment will use a camera with 1000 x 1000 pixels to acquire about 10 images during descent, scaled in 2:1 size ratios from roughly 8 x 8 km at 8 m/pixel down to 9 x 9 m at 9 mm/Pixel. Such imagery may help both rover localization and terrain traversibility analysis. Localization could use distinct albedo features, such as dark rock outcropping against lighter sand or dust; however, localization would certainly benefit from any topographic information that could be inferred from descent imagery. Therefore, we present a basic sensitivity analysis that relates the resolution of triangulation-based elevation mapping to the downward-looking, downward-moving trajectory of the camera and the sequence of progressively finer-scale images that will be acquired. It appears quite likely that elevation maps can be computed from descent imagery with enough precision to be useful for rover localization.

Finally, we consider visual localization at large distances from a lander (section 5). We survey the resolution and area coverage that is available, or is expected to be available, from the Viking, Mars Global Surveyor (MGS), and Mars Surveyor'98 orbiters. We also survey the digital terrain map (DTM) products available from Viking and expected from MGS. In general, it appears unlikely that accurate, semi-automatic rover localization will be achievable with available elevation data, for example by matching horizon features seen by the rover to features in a DTM. An alternative is to attempt to use ground-level imagery from the rover to recognize features from monocular orbiter imagery. Both of these approaches require further research into algorithms and achievable localization performance. Uncertainty about the potential performance of such methods suggests that other alternatives should be sought for navigation over 10's to 100's of kilometers, such as GPS-like navigation aids from orbit [2], or additional orbital imaging campaigns that would obtain resolutions on the order of 10 m/pixel over the area of traverse, preferably in stereo.

2 Localization with lander imagery

When a rover is within view from a lander, conceivable methods for visually estimating the position of the rover include:

a. recognizing the rover in imagery taken by the lander;

b. recognizing the lander in imagery taken by the rover;

c. mapping the area within view from the lander, using stereo cameras on the lander, then recognizing landmarks in the map in imagery taken by the rover.

Each approach has strengths and weaknesses. A strength of the first approach is that it can be very easy to implement; however, it requires involvement of the lander, which imposes potentially undesirable operational overhead on use of the lander cameras. The second approach avoids such imposition on the lander, but has other problems, like potential operational overhead on the rover (ie. turning to look at the lander) or occlusion of the lander by rocks of medium height. The third approach has the disadvantage that it may not work well if the landing site is very sparsely populated with features that can serve as landmarks. However, it has the significant advantage that it can also be used when the rover is exploring small areas very far from the lander, by first mapping each area with stereo cameras on a rover mast. This section will elaborate on work to date on approach (a); section 3 discusses work to date on approach (c). No development has taken place to date on approach (b).

To recognize the rover in lander imagery, we assume the lander has a stereo pair of cameras on a pan/tilt platform and has enough approximate knowledge of the rover location to aim the cameras at the rover. The problem is then to detect where the rover is in the images and to estimate the position of the rover. Initial discrimination of the rover from the background could be achieved by (1) using known color or geometric features of the rover, or (2) by taking images before and after the rover moves a small amount, then detecting the moving patch in the difference image. Although the latter method may require extra motions by the rover, it is completely independent of the appearance of the rover and places no constraints on the design of the rover shape or coloration. On this basis, we have chosen to implement this approach for the Rocky 7 Mars rover research vehicle [3].

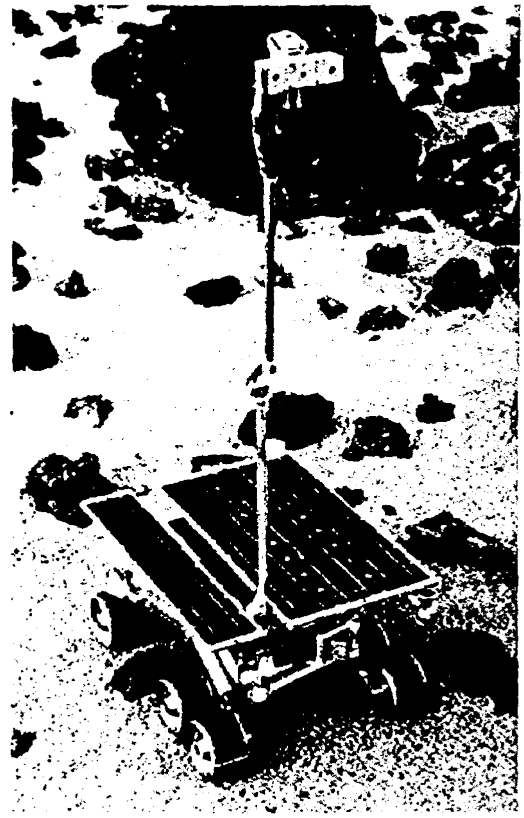

Figure 1: Rocky 7 Mars rover testbed with mast deployed. The vehicle is about 70 cm long. The mast is about 1.5 m high; it holds stereo cameras that can be panned to map the area around the rover. Additional stereo camera pairs are mounted under the solar panel on the front and rear of the vehicle for obstacle detection.

The rover motion could be driving forward some distance or turning in place through a small angle. Presently, we use an in-place turn, because the angle can be independent of the distance from lander, whereas the amount of forward motion would have to take into account both the distance from the lander and the direction of the motion relative to the direction to the lander. Images are taken with one camera before and after the turn; the absolute difference of these images is thresholded and binarized, then the center of the bounding box of the largest connected region in the binary image is taken to be the image coordinates of the rover. Cross correlation of an image patch around these coordinates gives the location of the rover in the other image of a stereo pair; finally, the position of the rover relative to the lander is computed by stereo triangulation. Rocky 7 estimates its own orientation with a sun sensor onboard the rover, so we do not need to estimate orientation visually.

This method has been tested in 150 trials performed by imaging the rover at distances of between

2 and 10 m from a mock-up lander. In each trial, the rover turned in place by 0.3 radians. Over all trials, the mean downrange distance estimate was 8.32% less than the measured distance, with a standard deviation of 6.66%. The mean absolute error in the cross-range distance was 3.30% of the downrange distance, with a standard deviation of 4.07%. The apparent bias in the downrange distance could be due to inaccurate camera calibration. These results show that the method is fairly successful for near-lander localization; future work will evaluate this performance in greater depth and compare it to theoretical predictions derived in [4].

3 Localization with panoramic imagery from mast-mounted rover cameras

Approach (c) above is applicable both for near-lander operations and for very distant operations, if the rover has stereo cameras on a mast that function as surrogates for lander cameras (figure 1). In the distant scenario, during a long distance traverse the rover would periodically stop and use its mast cameras to acquire panoramic stereo imagery of its environment. This imagery would be used to map roughly a 10 m radius disk around the rover. Scientists would designate several places to take measurements within this disk, and the rover would use the map to keep track of its position as it moves from place to place within the disk. The localization problem in this scenario is how to represent the map and how to use it to estimate the rover's position.

Two standard approaches to problems of this type are:

- To segment out discrete features in the map, such as rocks and gullies, and represent them with geometric primitives like polygons and curves. As the rover drives around, position estimation is achieved by matching primitives extracted from new imagery to the primitives in the map.

- To represent the map as an elevation grid, Position estimation is achieved by creating local elevation grids from new stereo imagery, acquired as the rover drives around, and correlating the "local" grids with the more global reference grid [5].

The former approach is susceptible to segmentation errors and limitations in modeling power of the geometric primitives; therefore, we are currently building a system based on the latter approach [6]. For this system, the reference grid may correspond to a previous rover position, an image panorama from the lander or rover, or descent imagery from the lander. The system determines the optimal relative position between the maps with respect to an iconic matching formulation through efficient search techniques. This section describes this system in more detail.

3.1 Computing elevation maps

In order to compute the local elevation map, we compute a range image at the local rover position using passive stereo vision [7], then convert the range image into a grid-based map representation. This is accomplished by, first, rotating the range data such that it has the same relative orientation as the global map we are comparing it to. Here we operate under the assumption that the orientation of the rover is known through sensors other than vision (for example, a sun sensor, accelerometer, and rate gyro have been incorporated into Rocky 7).

Next, the range points are binned in a two-dimensional grid covering the xy-plane at some specified scale. The terrain is a approximated as a single-valued function of the position in the xy-plane (i.e. $z = f(x, y)$). We thus take the average of the heights of the range points that fall into each of the bins as the height of the surface at this location. Figure 2 shows some intermediate results of this process for an example image.

3.2 Matching elevation maps

Once the elevation map haa been computed for the current position of the rover, we need to find the best relative position between this map and a map for which we know the frame of reference. To eliminate the need to search in the vertical direction, we first high-pass filter the elevation maps. We then use an image matching technique based on the Hausdorff distance [8]. This technique determines the relative position between the maps that maximizes the number of discrete locations in the local map that are "close" to a location in the global map, where "close" can be defined using any distance metric. To facilitate matching, we transform the two-dimensional map into a three-dimensional occupancy grid, where the z-axis is discretized at the same scale as the x- and y- axes.

To determine the position at which the best match occurs, we use a hierarchical cell decomposition approach. We first test the local position estimate obtained from dead-reckoning to determine an initial best known position for comparison. The space of possible relative positions is then divided into a set of rectilinear cells. Each cell is tested to determine whether it could contain a position that is a better match than the best known position so far. If it is determined that the cell cannot contain such a position, it is pruned. Otherwise, the cell is divided into subcells and the process is repeated recursively. At some level of fineness, the recursion is stopped and the position at the center of the cell is tested explicitly to determine if it is better than the best known posit ion.

The key to this method of searching the parameter space is a quick method to conservatively test whether a cell can contain a position that is better

(a) (b) (c)

Figure 2: Range maps are computed using stereo vision. (a) Left image of a stereo pair. (b) Height of pixels determined using stereo triangulation. (Black pixels indicate no data.) (c) Surface extracted.

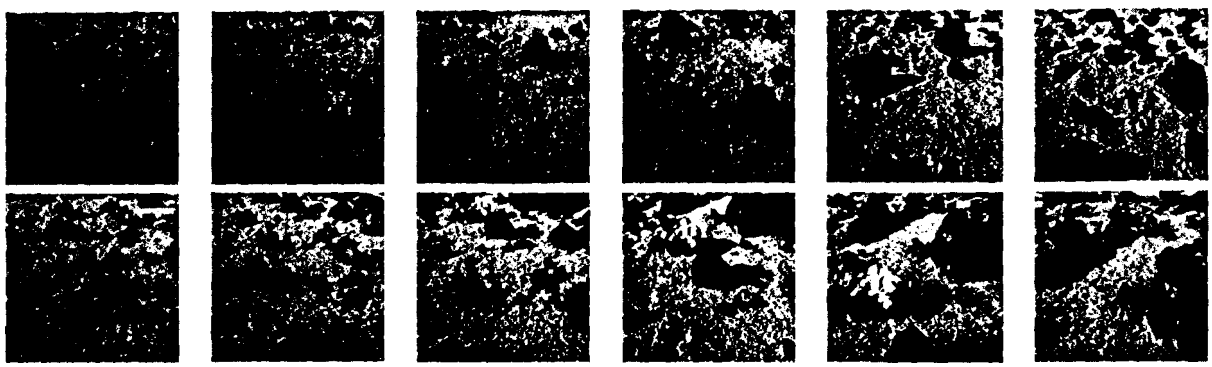

Figure 3: A sequence of images used for testing the localization techniques.

than the best known position. The test can fail to rule out a cell that does not contain such a position, but it should never rule out a cell that does contain one. This is performed by examining the distance transform of the occupancy map. See [6] for additional details.

3.3 Results

We have tested these techniques using images taken in the JPL Mars Yard with a stereo pair of cameras mounted on a tripod at approximately the Rocky 7 mast height. Figure 3 shows an example sequence of images that was used for testing the on-board localization techniques. This sequence consists of 12 images acquired at positions one meter apart in a straight line with approximately the same heading.

In this test sequence, the techniques find the qualitatively correct position between each pair of consecutive images, using the first as the global map and the second as the local map. The average absolute error in the localization steps is 0.0317 meters in the downrange direct ion and 0.0366 meters in the cross-range direction from the position measured by hand. Further accuracy can probably be achieved by using this position as the starting position for an iterative hill-climbing procedure. These results are quite encouraging. Ongoing work will test the approach with imagery from the Rocky 7 mast and obstacle detection cameras.

4 Localization with descent imagery

Descent imagery will be part of the Mars Surveyor'98 mission and is likely to be part of later missions. The descent camera for the 1998 mission has a resolution of about 1000 x 1000 pixels and field of view of 73.4 degrees [9]. The nominal imaging plan is to acquire 10 images during descent, beginning at an altitude of about 6750 meters and taking successive images at each halving of altitude. This would produce a sequence of images with footprints scaled in 2:1 size ratios from roughly 8 x 8 km at 8 m/pixel down to 9 x 9 m at 9 mm/pixel. This is much higher resolution than most orbital imagery, which has $\geq$ 30 m/pixel for most of the planet (see section 5). Therefore, descent imagery could be of tremendous benefit for planning rover exploration within several kilometers of the lander. However, such exploration would be ineffective without good rover position estimates

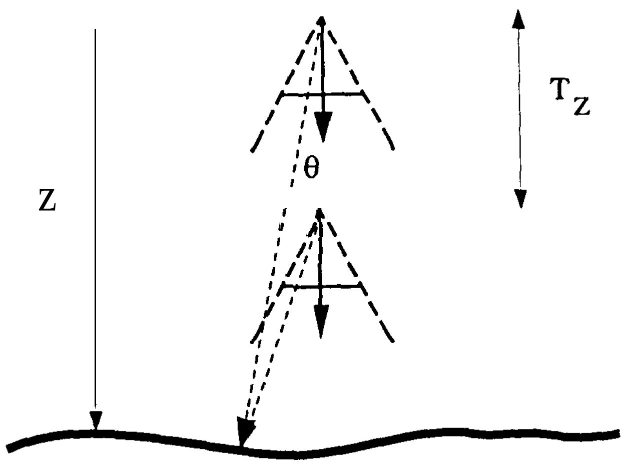

Figure 4: Elevation mapping with consecutive descent images: predicted precision of elevation estimates (l-sigma) versus distance from the landing site, assuming elevation data computed from consecutive members of a sequence of 10 descent images and image matching precision of 0.1 pixel.

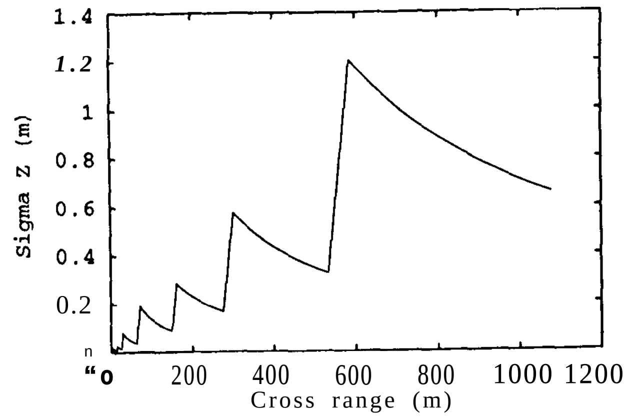

Figure 5: Elevation resolution for descent imagery.

relative to features seen in the descent imagery. This leads inevitably to asking how the descent imagery itself may be used to aid rover localization. Approaches to doing so include:

- using elevation maps computed from the descent sequence for terrain matching, analogous to the previous section, or for recognition of skyline features;
- inferring topographic structure from individual intensity images, such as identifying ridges, ravines, and large rocks from shading, then recognizing such structures aa the rover encounters them using onboard range data, odometry profiles, or rover intensity imagery.

In this paper, we explore the former approach.

Figure 4 illustrates the geometry of elevation mapping from two consecutive descent images, for the ideal case of purely vertical descent. θ represents the viewing angle from the top image to a specific point on the surface, measured with respect to the optical axis of the camera. Z represents the altitude for the first image and T_z the distance moved from the first to the second image. If the surface point can be found in the second image with an angular resolution of $\Delta\theta$ radians (ie. by cross-correlation), then it can be shown that the elevation of the surface point (relative to the first camera) can be estimated with a precision of

$$\Delta Z = \frac{(Z - T_Z)^2 \Delta\theta}{T_Z \tan\theta}$$

For the Mars'98 descent imagery, we expect that $T_Z \approx 2/2$. Clearly, the best precision is obtained at the edge of the image, and no elevation information can be obtained at the center of the image. Since successive images are taken at lower altitudes, the precision will improve progressively. Therefore, combining elevation information from all images of the sequence would produce a cross range resolution profile as shown in figure 5.

If the resolution predicted in figure 5 can be achieved in practice, elevations will be known with a standard deviation of one meter or better within a kilometer of the lander, and two meters within two kilometers of the lander. It is quite plausible that this is sufficient resolution to be useful for rover localization, terrain traversability analysis, and mission planning. Further work will attempt to validate the resolution predictions made here by determining how well the relative positions and orientations of the camera can be estimated for each descent image and by evaluating the precision with which features can be matched in consecutive imagery.

5 Localization with orbital imagery

Ultimately, rover missions will extend beyond the practical radius of descent imaging. What are the prospects for visual rover position estimation for such missions, using imagery acquired from orbit? Here, we survey the imagery and DTM products that are anticipated to be available from Viking, MGS, and Mars Surveyor'98, and use that to draw broad-brush conclusions about such prospects.

First, consider available imagery. Viking covered 100% of Mars at $\leq$ 260 m/pixel, but only 6% at 140 m/pixel and less than 1% at 30 m/pixel [10]. The highest resolution imagery is about 8 m/pixel, but very little of the planet was imaged at this resolution. The MGS orbiter will have selective coverage at 280 m/pixel and spot coverage at 1.4 m/pixel [11]. Adj scent ground tracks for MGS will be spaced by about 3 km at the equator. Since the MGS camera is nadir pointed, and the swath width of the spot coverage is less than 3 km, there is no possibility Of stereo imagery at the spot resolution; moreover, spacecraft navigation errors will limit the accuracy with which the spot imagery can be aimed at any particular point on the surface. The Mars Surveyor'98 mission will provide selective coverage at 40 m/pixel.

DTM's from Viking have a resolution of about 1 km/pixel. The laser altimeter on MGS has a vertical precision of 2 m and will give along-track sampling at 330 m/pixel with 3 km between-track spacing at the equator [12]. This data will be used to produce a global topographic map with resolution of 0.2 x 0.2 degrees/pixel (about 12 km/pixel) with 30 meter global accuracy in elevation.

At present, how well a rover could be localized with such DTM's is a matter of speculation. It will certainly depend on the topography of the specific region being traversed; as a worst case, for example, the Viking landing sites had no visible relief at kilometer scale. Prior research elsewhere [13] on using horizon features for localization has been tested with USGS elevation maps with 30 m/pixel resolution in mountainous regions of the United States, and has achieved localization errors of around 90 m, or 3 pixels. By crude analogy, localization to within 3 pixels in Viking DTM's would give a position error of around 3 km. This may be acceptable, if it can be achieved; however, this remains to be demonstrated, and may only pertain to very rough regions of Mars.

Localization with the aid of orbital intensity imagery is also speculative, but the possibility of selective coverage at 40 m/pixel makes it worth considering. The problem here is the same as with descent imagery: can we extract topographic features from the intensity imagery that could be recognized from the ground, using either rover intensity imagery, accumulated elevation maps, or odometry-based elevation profiles? Potentially yes, if sufficient topographic variation exists, but this remains to be demonstrated.

Much work remains to be done to determine how well rovers could be localized with orbital imagery. However, the prospects are sufficiently uncertain to make it attractive to examine alternatives, such as GPS-like navigation aids from orbit [2] or additional imaging campaigns that would obtain high resolution (order of 10 m/pixel or better) stereo coverage over the area of traverse).

Acknowledgements

The research described in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. The ideas and results presented here benefitted from discussions with Brian Wilcox, Samad Hayati, and Rick Welch, as well as from the efforts of Rich Volpe, Bob Balaram, and the rest of the the Rocky 7 team for design and development of the hardware and onboard soft ware.

References

[1] NASA web site. http://www.hq.nasa.gov/office/oss/research.html.

[2] K. M. Lau et al. An innovative deep space application of GPS technology for formation flying spacecraft. In *Proc. AIAA GN&C* Conference. American Institute of Aeronautics and Astronautics, July 1996.

[3] R. Volpe, J. Balaram, T. Ohm, and R. Ivlev. The Rocky 7 Mars rover prototype. In *IEEE/RSJ International* Conference *on* Intelligent *Robots and Systems (IROS)*, November 1996.

[4] L. Matthies, E. Gat, R. Harrison, B. Wilcox, R. Volpe, and T. Litwin. Mars microrover navigation: performance evaluation and enhancement. *Autonomous Robots*, 2(4):291–312, 1995.

[5] D. B. Gennery. Visual terrain matching for a Mars rover. In *Proceedings of the IEEE* Conference on *Computer* Vision *and Pattern* Recognition, pages 483-491. IEEE Computer Society Press, June 1989.

[6] C. F. Olson. Mobile robot self-localization by iconic matching of range maps. In *Proceedings of the 8th International Conference on Advanced Robotics (ICAR), 1997.*

[7] L. H. Matthies. Stereo vision for planetary rovers: stochastic modeling to near real-time implementation. International *Journal of Computer* Vision, 8(1):71–91, July 1992.

[8] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge. Comparing images using the Hausdorff dist ante. *IEEE Transactions on Pattern* Analysis and *Machine Intelligence,* 15(9):850–863, September 1993.

[9] Malin Space Science Systems web site. http://barsoom.msss.com/mars/surveyor/mardi.html.

[10] NASA/JPL Planetary Data System web site. http://pds.jpl.nasa.gov/cgi-bin/layout.pl?ds:dsOO0023:inv/ds/ds230 .1bl.

[11] Mars Global Surveyor web site. http://mgs-www.jpl.nasa.gov.

[12] Mars Orbiter Laser Alterimeter web site. http://ltpwww.gsfc.nasa.gov/tharsis/mola.html.

[13] F. Cozman and E. Krotkov. Position estimation from outdoor visual landmarks for teleoperation of lunar rovers. In *Proceedings of the 3rd IEEE Workshop on Applications of Computer Vision,* December 1996.