Stereo Ego-motion Improvements for Robust Rover Navigation

Clark F. Olson, Larry H. Matthies, Marcel Schoppers, and Mark W. Maimone
Jet Propulsion Laboratory, California Institute of Technology
4800 Oak Grove Drive, Pasadena CA 91109-8099

Abstract

Robust navigation for mobile robots over long distances requires an accurate method for tracking the robot position in the environment. Techniques for position estimation by determining the camera ego-motion from monocular or stereo sequences have been previously described. However, long-distance navigation requires a very high level of robustness and a very low rate of error growth. In this paper, we describe a methodology for long-distance rover navigation that meets these goals using robust estimation. We show that a system based on only camera ego-motion estimates will accumulate errors with super-linear growth in the distance traveled, owing to increasing orientation errors. When an absolute orientation sensor is incorporated, the error growth can be reduced to a linear function of the distance traveled. We have tested these techniques using both extensive simulation and hundreds of real rover images and have achieved a low, linear rate of error growth.

1 Introduction

Our goal is to perform robust and accurate rover navigation autonomously over long distances in order to reach terrain landmarks with known locations, but that are not within sight. This is motivated by the high desirability for Mars rovers to autonomously navigate to science targets observed in orbital or descent imagery. Since communication with such rovers usually occurs only once per day, navigation errors can result in the loss of an entire day of scientific activity.

The most common method for the position estimation of mobile robots is through dead-reckoning. This technique integrates the velocity history, using the estimated speed and direction of travel, to determine the change in position from the starting location. Unfortunately, pure dead-reckoning methods are prone to errors that grow without bound over time, so some additional method is necessary to periodically update the robot position. This can be performed through global localization of the robot (see, for example, [7]). In this paper, we concentrate on a different method called ego-motion (or visual odometry). Like dead-reckoning, this method accumulates error as the robot moves, so that some periodic update is beneficial. However, we demonstrate that, when combined with an orientation sensor, this technique is able to reduce the growth rate of the error to a small fraction of the distance traveled.

Several methods for the estimation of ego-motion have been proposed using monocular sequences [1, 2, 3] and stereo sequences [4, 5, 9, 10]. In order for such techniques to be effective for long-distance rover navigation, the techniques must be highly robust to problems such as poor odometry, inaccurate feature matching, and outliers. We have developed a method that is capable of accurate navigation over long distances using incremental stereo ego-motion [8]. The use of stereo information in this method has been crucial in both outlier rejection and reducing random errors that occur due to feature localization and drift in each frame. We use a maximum-likelihood formulation of motion estimation that models the error in the positions more accurately than a least-squares formulation and, thus, yields better results.

For long-range navigation, we must examine the rate of error growth as the robot navigates the environment. Even a robust incremental method (such as ego-motion) accumulates errors that grow super-linearly with the distance traveled, if the absolute orientation is not corrected periodically. We demonstrate that incorporation of an orientation sensor, such as a compass or sun sensor, can greatly improve the long-range performance, reducing the accumulated error to a linear function of the distance traveled.

We have constructed a simulator in order to evaluate changes in the ego-motion methodology with respect to navigation performance. The simulator indicates that, with our improvements, ego-motion performance with error below 0.5% of the distance traveled is potentially feasible. Experiments on hundreds of real images have achieved errors of approximately 1% of the distance traveled.

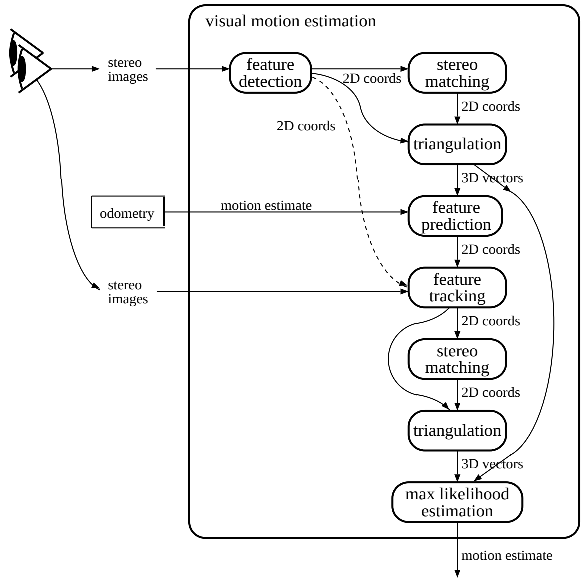

Figure 1: Steps performed for motion estimation.

2 Motion estimation

We use a variation of the maximum-likelihood ego-motion formulation originally developed by Matthies [5, 6]. This method determines the change in position for calibrated cameras using two (or more) pairs of stereo images. The basic elements of the method are shown in Fig. 1.

First, landmarks that can be easily tracked are selected in an initial image. Stereo matching is used to find the landmark in the corresponding stereo image and the 3D landmark positions are estimated through triangulation. Next, the landmarks are tracked into a subsequent stereo pair with correlation-based search using prior knowledge of the approximate robot motion. Stereo matching is then performed in the subsequent stereo pair. Finally, the motion estimate is computed using maximum-likelihood estimation as described below.

Each of these steps is performed for each pair of adjacent frames in the stereo sequence. At each iteration, the set of landmarks is retained from the previous step, with new landmarks added for those that were not successfully tracked. The overall motion estimate is computed as the product (in homogeneous coordinates) of each of the incremental motions over the image sequence.

The maximum-likelihood motion estimation step takes the landmark positions that have been estimated from consecutive stereo images and determines the change in the rover positions. We summarize the details of this method here. Let L^b and L^a be $3 \times n$ matrices of the observed landmark positions before and after a robot motion. For each landmark we have:

$$L_i^a = R L_i^b + T + e_i, \tag{1}$$

where R and T are the rotation and translation of the robot and e_i combines the errors in the observed positions of the ith landmark at both locations. Assume, for the moment, that the pre-move landmark positions are errorless and the post-move landmark positions are corrupted by Gaussian noise. In this case, the joint conditional probability density of the observed post-move landmark positions, given R and T, is Gaussian:

$$f(L_1^a, ..., L_n^a | R, T) \propto e^{-\frac{1}{2}\sum_{i=0}^{n} r_i^T W_i r_i}, \tag{2}$$

where $r_i = L_i^a - R L_i^b - T$ and W_i is the inverse covariance matrix of e_i. The maximum-likelihood estimate for R and T is given by minimizing the exponent $\sum_{i=0}^{n} r_i^T W_i r_i$. Note that this reduces to the least-squares solution if we let $W_i = w_i I$.

Solving for the maximum-likelihood motion estimate is a nonlinear minimization problem, which we solve through linearization and iteration. We linearize the problem by taking the first-order expansion with respect to the rotation angles. Let Θ_0 be the initial angle estimates and R_0 be the corresponding rotation matrix. The first-order expansion is:

$$L_i^a \approx R_0 L_i^b + J_i(\Theta - \Theta_0) + T + e_i, \tag{3}$$

where J_i is the Jacobian for the ith landmark and e_i is a Gaussian noise vector with covariance $\Sigma_i = \Sigma_i^a + R_0 \Sigma_i^b R_0^T$.

We can now determine a maximum-likelihood estimate for Θ and T using $r_i = L_i^a - R_0 L_i^b - J_i(\Theta - \Theta_0) - T$ and $W_i = (\Sigma_i^a + R_0 \Sigma_i^b R_0^T)^{-1}$. Differentiating the objective function with respect to Θ and T and setting the derivatives to zero yields:

$$\left[\sum_{i=0}^{n} H_i^T W_i H_i\right] \begin{bmatrix} \Theta \\ T \end{bmatrix} = \left[\sum_{i=0}^{n} H_i^T W_i L_i\right], \tag{4}$$

where $H_i = [J_i \; I]$ and $L_i = L_i^a - R_0 L_i^b + J_i \Theta_0$.

After solving (4), the new motion estimate is used as an initial estimate for the next step and the process is iterated until convergence. Further details, and a technique to estimate only Θ without T, so that estimation of T can be removed from the iteration, can be found in [6].

3 Simulator experiments

One of the goals of our work has been to study the long-range performance of ego-motion techniques under controlled conditions. To this end, we have developed a simulator that tracks randomly generated landmarks for motion estimation. The initial landmarks are generated by selecting random image locations in the left image of the first (pre-move) stereo pair. The positions of the landmarks are backprojected into 3D using a random (uniformly distributed) height. Each landmark is then reprojected into the right image of the stereo pair with Gaussian noise ($\sigma = 0.3$ pixels) added in order to simulate feature matching error.

A second (post-move) stereo pair is generated using the same set of landmarks, but using camera models translated and rotated to a new position (simulating robot motion). The left image of the pair is generated by projecting the landmarks according to the new camera model and adding more Gaussian noise ($\sigma = 0.5$ pixels) in order to simulate the feature tracking error. The new image features are again backprojected into 3D (with the same heights) and reprojected into the right image of the post-move stereo pair with additional noise.

The incremental robot motion estimate is computed using the maximum-likelihood ego-motion method described above. Long-distance navigation is simulated by chaining many of the incremental moves together. At each step, the second set of landmark positions is saved for use as the initial set in the next step and new landmark positions are generated as above. When landmarks move out of the robot field of view, they are replenished with randomly positioned landmarks within the field of view.

3.1 Optimal field-of-view

We have used the simulator to perform an experiment determining the effect of changing the camera field-of-view on the ego-motion performance. Our expectation was that error in the ego-motion performance would be better for smaller field-of-view camera, if the other parameters remained the same, due to the improved angular resolution of the camera. Of course, at some point, this must break down due to the field of view becoming too small to track the features effectively. Figure 2 shows the result of an experiment where the camera field of view was varied from 15° to 90°. The baseline of the stereo pair was maintained at 10 cm with a camera height of 1.4 m and a tilt of 30°. The rover moved 50 cm between each ego-motion calculation. In this case, the optimal camera field-of-view appears to be approximately 35°. The optimal field-of-view changes when other parameters of the system change, but not by a large amount. When the rover movement was varied between 30 and 70 cm between ego-motion calculations, the optimal field of view remained between 30° and 40°. Similar results were also obtained with a varying baseline and camera elevation. Our conclusion is that decreasing the field-of-view helps up to a point, but when the field-of-view becomes less than 30° the improvement is negated or reversed by other effects. In particular, the limited field-of-view over which landmarks can be tracked results in poor sensitivity with respect to the orientation of the cameras.

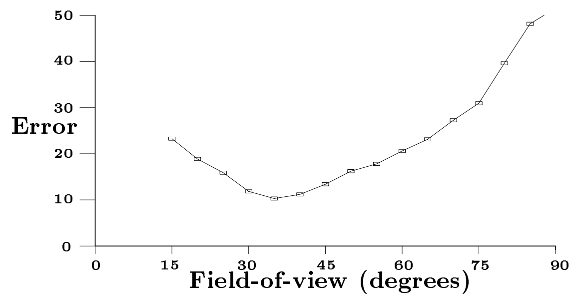

Figure 2: The ego-motion error varies as a function of the camera field-of-view.

3.2 Long-range error growth

Since we are interested in long-range navigation for Mars rovers, we have performed experiments examining the error growth of the stereo ego-motion techniques by applying them to a long sequence of simulated data. Our goal here is to understand the asymptotic growth of the error over long distances.

We performed an experiment with a 500 meter traverse. Ego-motion estimates were computed every 50 cm using cameras with a 45° field-of-view and 512×480 pixels (corresponding to the values on our research prototype rover). Figure 3 shows the error growth in the robot position for this experiment. It can be observed that the growth in the error is greater than linear in the distance traveled. The explanation for this is that the expected error in the orientation parameters grows approximately proportional to the square root of the distance traveled (since the overall variance is the sum of the individual variances). The overall position error grows as the sum of two terms. First, the individual position errors contribute a term that is expected to grow with the square root of the distance traveled. Second, the accumulating orientation errors contribute a term that grows with

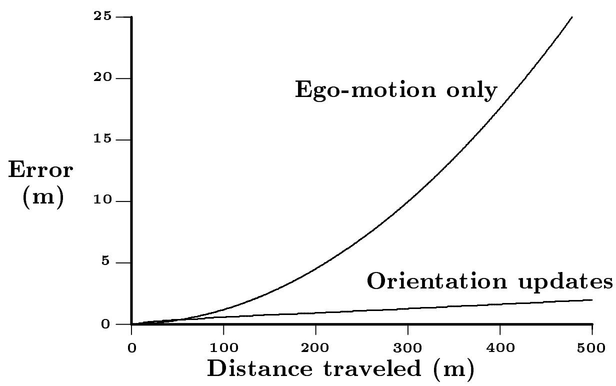

Figure 3: Expected position error as a function of distance traveled.

the integral of the orientation error. We, thus, expect a super-linear contribution from this term, which has $O(d^{\frac{3}{2}})$ asymptotic growth, where d is the distance traveled. The contribution from the orientation error thus dominates the overall position error in this case.

In order to eliminate the super-linear error growth, we have examined the use of an absolute orientation sensor to provide periodic updates to the orientation estimate. For example, accelerometers can be used to provide roll and pitch information, while a compass, sun sensor, or even a panoramic camera could be used to determine the robot yaw. We have simulated such sensors as providing periodic orientation updates with Gaussian noise having zero mean and 1° standard deviation. Figure 3 shows that this results in linear error growth in the distance traveled when the orientation updates are used and, in general, the growth is much slower than when only the ego-motion estimates are used. In this experiment, the simulations indicate that error less than 1% of the distance traveled is achievable with the error variances described above.

An absolute orientation sensor appears to be critical for navigation over long distances, unless some other means is used to periodically update the robot position. If no orientation sensor is used, the robot may navigate safely over short distances. However, over long distances the increasing orientation errors will build until the position estimate is useless.

4 Robust estimation

In order to achieve accurate navigation over long distances, errors in the matching process and in the position estimation of the landmarks must have a very small effect on each computed motion estimate. Tracking must be performed such that mismatches are rare. When mismatches occur, there must be mechanisms for detecting and discarding them. This section reviews some improvements we have developed for performing these steps [8], while reducing the overall error growth in the rover position for improved navigation.

4.1 Improved feature tracking

In many environments, including Martian terrain, the landmarks that are selected for tracking appear similar to each other and other image locations. We don't want to search for each feature over a large portion of the image, because incorrect matches will occur frequently in this case. However, error in the *a priori* estimate for each landmark position (obtained using dead-reckoning) usually requires the use of a large search window.

In order to decrease the size of this search window, we first estimate the errors in the robot pitch and yaw by searching for a large, distant landmark in the image. The use of a large landmark allows us to avoid mismatches, even searching over a large portion of the image for this landmark. The robot pitch and yaw estimates are then corrected such that they accurately predict the position of the landmark in the new image. Once this correction has been performed, the search for the smaller landmarks in the image can be performed over a small area, and so the possibility of an incorrect match is reduced.

4.2 Outlier rejection

We use several methods to reject outliers in the motion estimation process. Initially, matches in both the stereo matching and feature tracking steps are eliminated if the correlation score is too low. In addition, for each stereo match, the rays from the cameras through the image features are computed to determine if they consistent. The consistency is measured by the distance between the rays at the location of smallest separation. (If there was no error, the rays would intersect.) Finally, after all of the matches have been found and tracked in both stereo pairs, a rigidity test is applied to prevent gross errors. Here, we use a constraint that the landmarks must be stationary. If a landmark moves between stereo frames, the landmark is not useful for determining the robot motion. This test repeatedly rejects the landmark that appears to have moved the most, by examining the pairwise distances between the landmarks before and after the robot motion. Landmarks are rejected until all remaining deviations are small enough to be considered noise.

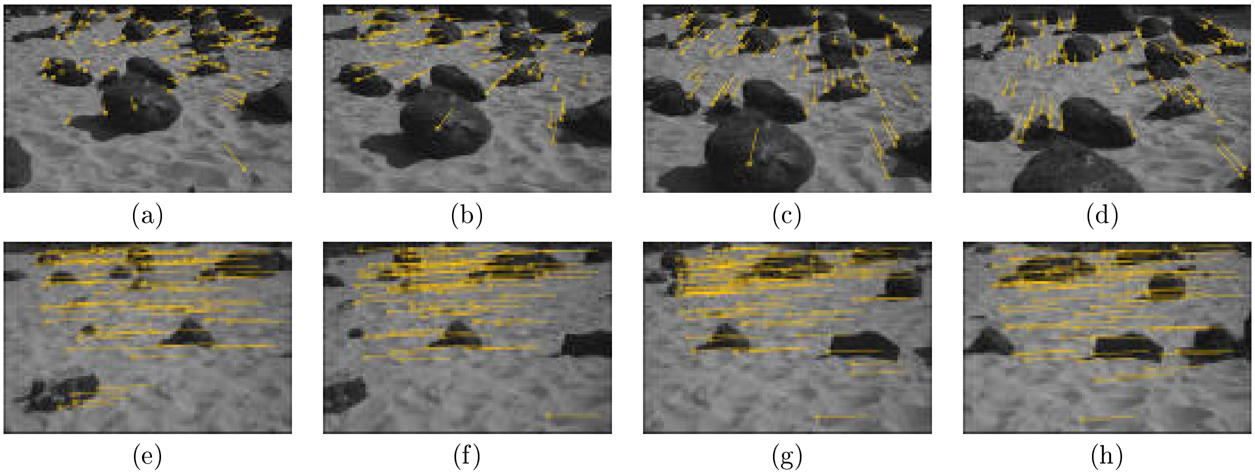

Figure 4: Image sequences demonstrating robust feature tracking. The top row is largely forward motion. The bottom row is largely rotation. The squares indicate the tracked landmarks and the lines show the motion of the landmark from the previous frame.

4.3 Camera roll

Camera roll due to traversing rough terrain is a significant problem for robots that operate outdoors. While pitch and yaw are reasonably approximated by translation of the features in the image, roll causes the features to be rotated and makes tracking significantly more difficult. Our experiments indicate that correlation scores degrade approximately linearly with the camera roll. In most terrains, camera roll of less than $10°$ can be tolerated without difficulty to the feature tracking.

Clearly, a robust motion estimation system for outdoor navigation must consider the effects of camera roll. The simplest solution to this problem is to ensure that image pairs are captured frequently enough that the robot does not roll by more than $10°$ between frames. For many systems, this solution is adequate. An alternative, for cases where large amounts of camera roll are possible, is the use of an orientation sensor, such as a gyro or accelerometer. If the approximate roll of the camera is known, then the correlation window for each landmark can be rotated to the appropriate orientation for tracking.

5 Results

These techniques have been tested on hundreds of stereo pairs, including outdoor terrain, with the robot undergoing six degree-of-freedom motion. Figure 4 shows landmark tracking for several frames of mostly forward motion (top row) and mostly rotational motion (bottom row) in rocky terrain. Despite errors in the nominal camera movements and features occurring on occluding boundaries that are difficult to track, it can be observed that the final tracking is highly robust, with no outliers in the tracking process. For this data set, the overall error was 1.3% of the distance traveled.

In order to test the performance of these techniques on extended sequences, we have applied them to imagery from a rover traverse consisting of 210 stereo pairs. This traverse was performed with a small rover and a wide field-of-view, so the cameras were close to the ground (10 cm) and there was considerable distortion in the appearance of close-range locations. Figure 5 shows an example of consecutive stereo pairs with 320×240 resolution. The rover traversed approximately 20 meters, taking images about every 10 centimeters. For cameras with a higher viewpoint and narrower field-of-view, the techniques could be executed less frequently. However, for this rover, small motions between stereo pairs are necessary to track the foreground landmarks. Figure 6 shows the results for this traverse. It can be observed that the ego-motion track closely follows the ground-truth from GPS, while the odometry estimate diverges from the true position. The error in this run was approximately 1.2%.

6 Summary

We have discussed techniques for improving long-range rover navigation using stereo ego-motion. An

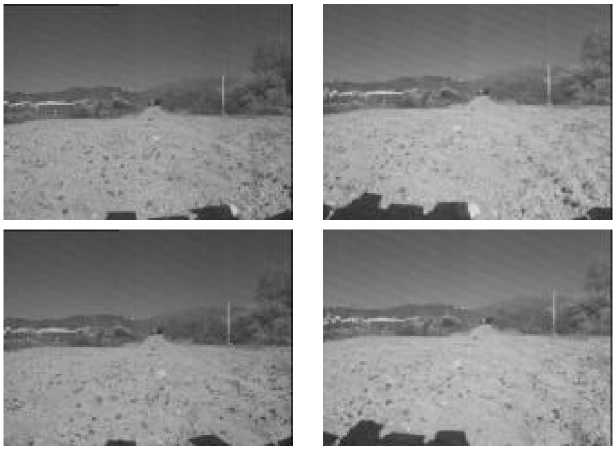

Figure 5: Stereo pairs from rover traverse sequence.

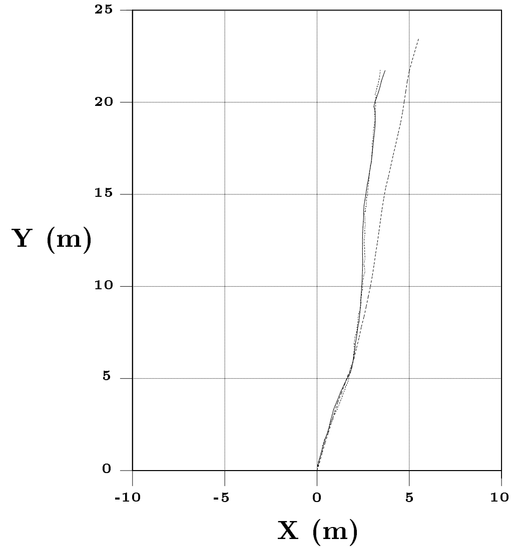

Figure 6: An extended run consisting of 210 stereo pairs. The solid line is the GPS position of the rover. The dotted line is the ego-motion estimate. The dashed line is the odometry estimate.

important result of our investigation is that an absolute orientation sensor is necessary to perform accurate navigation over long distances, since estimation based on ego-motion alone has error that grows super-linearly with the distance traveled. The use of an orientation sensor reduces the error growth to linear in the distance traveled and results in a much lower error in practice. Stereo data was also critical to elimination of outliers and accurate motion estimation. Techniques for performing robust feature selection and tracking with outlier rejection have been developed in order to ensure accurate motion estimation at each step. We believe that this combination of techniques results in a method with greater robustness than previous techniques and that is capable of accurate motion estimation for long-distance navigation.

Acknowledgements

The research described in this paper was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

References

[1] S. Chaudhuri, S. Sharma, and S. Chatterjee. Recursive estimation of motion parameters. *Computer Vision and Image Understanding*, 64(3):434–442, November 1996.

[2] D. J. Heeger and A. D. Jepson. Subspace methods for recognition rigid motion. I. algorithm and implementation. *International Journal of Computer Vision*, 7(2):95–117, January 1992.

[3] K. Kanatani. 3-d interpretation of optical flow by renormalization. *International Journal of Computer Vision*, 11(3):267–282, December 1993.

[4] A. Mallet, S. Lacroix, and L. Gallo. Position estimation in outdoor environments using pixel tracking and stereovision. In *Proceedings of the IEEE Conference on Robotics and Automation*, volume 4, pages 3519–3524, 2000.

[5] L. Matthies and S. A. Shafer. Error modeling in stereo navigation. *IEEE Transactions on Robotics and Automation*, 3(3):239–248, June 1987.

[6] L. H. Matthies. *Dynamic Stereo Vision*. PhD thesis, Carnegie Mellon University, October 1989.

[7] C. F. Olson. Probabilistic self-localization for mobile robots. *IEEE Transactions on Robotics and Automation*, 16(1):55–66, February 2000.

[8] C. F. Olson, L. H. Matthies, M. Schoppers, and M. W. Maimone. Robust stereo ego-motion for long distance navigation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 453–458, 2000.

[9] J. Weng, P. Cohen, and N. Rebibo. Motion and structure estimation from stereo image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(3):362–382, June 1992.

[10] Z. Zhang and O. D. Faugeras. Estimation of displacements from two 3-d frames obtained from stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(12):1141–1156, December 1992.