Available at www.ComputerScienceWeb.com
POWERED BY SCIENCE DIRECT®

Robotics and Autonomous Systems 43 (2003) 215–229

Robotics and Autonomous Systems

www.elsevier.com/locate/robot

Rover navigation using stereo ego-motion

Clark F. Olson [a,*], Larry H. Matthies [b], Marcel Schoppers [b], Mark W. Maimone [b]

[a] *Computing and Software Systems, University of Washington, 18115 Campus Way NE, Box 358534, Bothell, WA 98011-8246, USA*
[b] *Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109-8099, USA*

Received 30 January 2002; received in revised form 15 December 2002

Abstract

Robust navigation for mobile robots over long distances requires an accurate method for tracking the robot position in the environment. Promising techniques for position estimation by determining the camera ego-motion from monocular or stereo sequences have been previously described. However, long-distance navigation requires both a high level of robustness and a low rate of error growth. In this paper, we describe a methodology for long-distance rover navigation that meets these goals using robust estimation of ego-motion. The basic method is a maximum-likelihood ego-motion algorithm that models the error in stereo matching as a normal distribution elongated along the (parallel) camera viewing axes. Several mechanisms are described for improving navigation robustness in the context of this methodology. In addition, 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. This method has been implemented to run on-board a prototype Mars rover.

Keywords: Robot navigation; Motion estimation; Stereo vision; Mars rovers

1. Introduction

The most common method for estimating the position of mobile robots is through dead-reckoning. This technique integrates the velocity history of the robot, 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, [1,7,10,16,19]). 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, for most sensor combinations, ego-motion estimation yields considerably more accurate position estimation.

Visual motion estimation can be viewed as a middle ground between dead-reckoning and global localization. We demonstrate that, when combined with an orientation sensor, this technique is able to reduce the expected growth rate of the error to a small fraction of the distance traveled. While the use of such an orientation sensor can also be used with dead-reckoning to achieve (on average) a linear rate of error growth, the overall rate of growth is typically much greater with this method. This technique is promising for improving the position estimation capability of a mobile

* Corresponding author.

0921-8890/03/$ – see front matter
doi:10.1016/S0921-8890(03)00004-6

robot owing to this reduced error in comparison to dead-reckoning.

With visual motion estimation, landmarks are tracked in an image sequence and the change in camera position is determined for each frame by estimating the relative movement of the tracked landmarks in the camera frame of reference. Several methods for the computation of ego-motion have been proposed using monocular sequences [2,3,8,9,20,24,26] and stereo sequences [11–13,22,25,27]. However, in order for these techniques to be effective in long-distance navigation of a robot, the techniques must be highly robust to problems such as poor odometry, inaccurate feature matching, and outliers.

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. On the other hand, if localization errors during traverses can be minimized, additional scientific activity is allowed.

We have developed a method that is capable of achieving accurate navigation over long distances using incremental stereo ego-motion. 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 [13,14] that models error in the landmark positions more accurately than a least-squares formulation, and, thus, yields more accurate results. Robustness issues are further addressed through optimized feature selection, improved motion prediction, and multiple outlier rejection mechanisms. We show that the reuse of landmarks between frames significantly improves the overall accuracy since the errors at successive estimation steps become negatively correlated.

For long-range navigation, it is important to consider the rate of error growth as the robot travels. Even a robust system based solely on ego-motion will accumulate errors that grow super-linearly (on average) with the distance traveled, if the absolute orientation is not corrected periodically. However, the incorporation of an orientation sensor, such as a compass or sun sensor [21,23], can greatly improve the long-range performance, reducing the expected 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. We have further evaluated the robustness of these techniques using real rover images captured in rocky terrain, similar to the terrain that a rover would encounter on Mars. Experiments on hundreds of real images have achieved errors of slightly above 1% of the distance traveled.

An alternative to the method that we present here is the SLAM (simultaneous localization and mapping) methodology [4,5]. While both methods detect landmarks in order to update the position estimate of a mobile robot, the methods are somewhat different. In contrast to the SLAM method, we do not maintain a map of the environment. Landmark identification is performed in our method by matching image features within stereo pairs and between successive stereo pairs, making this task simple and efficient. Hundreds of landmarks can be identified at each position of the robot with this technique.

2. Motion estimation

Our motion estimation method is based upon the maximum-likelihood ego-motion formulation of Matthies and Shafer [13,14]. This method determines the observer motion between two (or more) pairs of stereo images captured by calibrated cameras. The basic elements of the method are as follows:

- *Feature selection*: The first step is to select landmarks for which the 3D position can be precisely measured in successive stereo pairs. The initial landmarks are selected by finding easily trackable features in the left image of the first stereo pair. We select the features using a variation of the Förstner interest operator [6], with an additional constraint such that no two features that are selected have a pairwise distance that is below a selected threshold.
- *Stereo matching (1)*: An estimate of the 3D position of the landmarks is obtained by performing stereo matching in the initial stereo pair. This

procedure searches the right image of the stereo to find the corresponding location for each of the selected landmarks. A multi-resolution pyramid is used to for efficiency and the best match is selected using normalized correlation. Triangulation is then performed using the known relative position between the cameras to determine the position of the landmark with respect to the camera frame. This step also provides a covariance matrix that models the error in the position estimate.

- *Feature tracking*: Landmarks are located in subsequent stereo pairs using a method similar to the search performed in stereo matching. However, in this case the relative position between the cameras is not known precisely. This has two implications. First, there is more uncertainty in the match location. Second, the match is not constrained to lie on a one-dimensional epipolar line. Together these make the search for correct match more time-consuming than stereo matching and more likely to fail. We use prior knowledge of the approximate robot motion to select the search space for the feature tracking.
- *Stereo matching (2)*: A second stereo matching step is performed to estimate the 3D positions of the landmarks with respect to the new camera frame. As in the previous steps, this uses a correlation-based search and triangulation is performed to estimate the position.
- *Motion estimation*: Motion estimation is performed using Gaussian error distributions for the landmark positions [13]. This yields better robustness than weighted least-squares minimization, which implies a rotationally symmetric error distribution. Maximum-likelihood estimation of the new robot

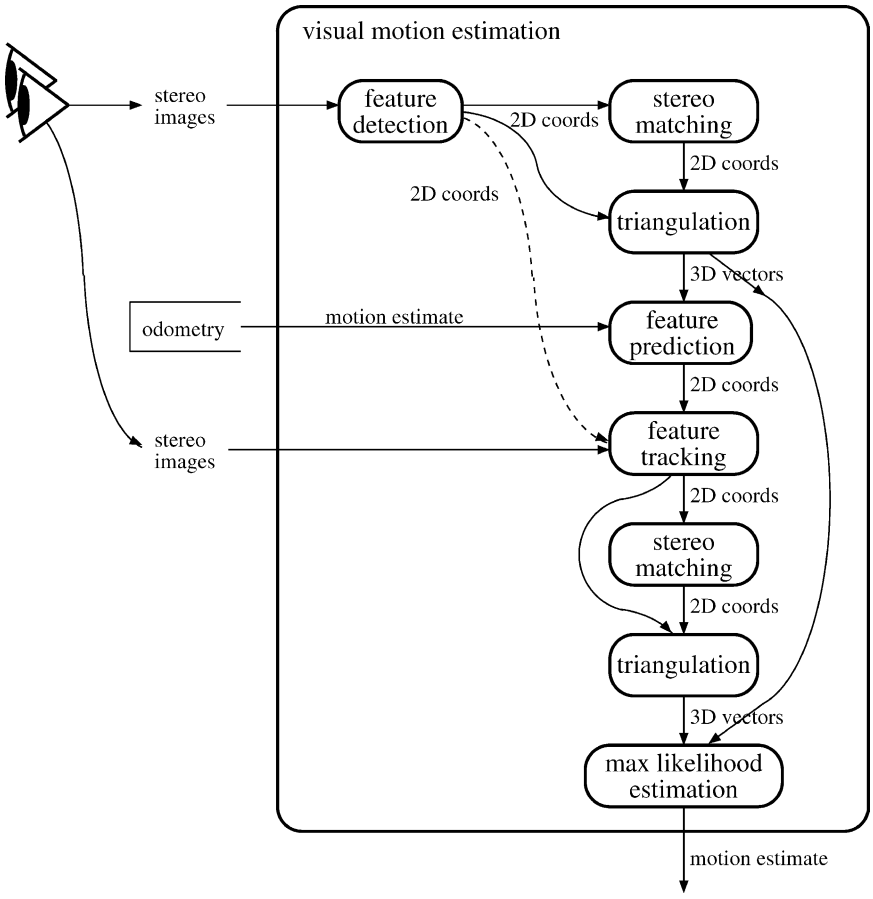

Fig. 1. Steps performed for motion estimation.

position requires an iterative optimization. However, convergence is fast and this step requires negligible computation time compared to the previous steps.

These steps are performed for each pair of consecutive stereo frames, retaining the same set of landmarks, but replenishing those that were not found or discarded. The overall motion estimate is determined as the combination of motions from each pair of frames. Fig. 1 shows the steps in the process to estimate the motion between two frames.

3. Maximum-likelihood ego-motion

Given the noisy landmark positions from stereo data, we use a maximum-likelihood formulation for motion estimation. An early version of this method was given in [13]. Further details can be found in [14].

Let L^{b} and L^{a} be $3 \times n$ matrices of the n observed landmark positions before and after a robot motion. The three-dimensional position of each landmark is estimated using stereo triangulation. For each landmark we have

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

where R and T are the rotation and translation of the robot and e combines the errors in the observed positions of the landmarks at both locations. Matthies and Shafer [13] has found that stereo errors are well approximated by a two-dimensional Gaussian distribution with the major axis aligned with the camera axis. We will 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:

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

where $r_i = L_i^{\mathrm{a}} - RL_i^{\mathrm{b}} - T$ 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 $\sum_{i=0}^{n} r_i^{\mathrm{T}} W_i r_i$. 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

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

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

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

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

where $H_i = [J_i I]$ and $L_i = L_i^{\mathrm{a}} - R_0 L_i^{\mathrm{b}} + J_i \Theta_0$. The covariance matrix is given by

$$\Sigma = \left[\sum_{i} H_i^{\mathrm{T}} W_i H_i\right]^{-1}. \tag{5}$$

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 [14].

4. 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.

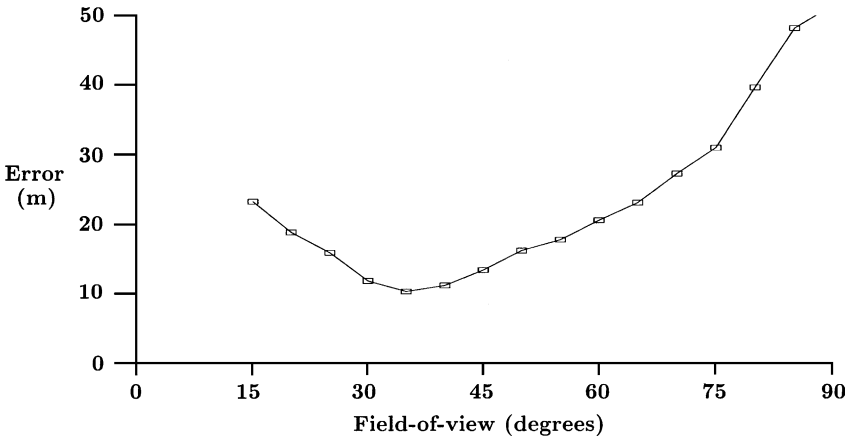

Fig. 2. The ego-motion error varies as a function of the camera field-of-view.

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 ($\sigma = 0.3$ pixels).

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.

4.1. 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. Fig. 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 downward tilt of 30°. The rover moved 50 cm between each ego-motion calculation. The total course length was 500 m for this experiment and no orientation information from other sensors was incorporated. With these parameters, the optimal camera field-of-view is 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 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.

Experiments with different distances between image pairs over the same distance course indicated that a step size of at least 50 cm may be desirable in order to limit the number of steps that introduce error. However, these experiments do not take into account the increased difficulty of tracking the features over longer distances, so it is likely that these experiments

significantly overestimate the optimal distance between image pairs. Further testing using real images is expected to resolve this issue.

4.2. 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 m 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). Fig. 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 the integral of the orientation error. We, thus, expect a super-linear contribution from this term, which has $O(d^{3/2})$ asymptotic growth, where d is the distance traveled. The contribution from the orientation error dominates the overall position error for long-range navigation.

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. Fig. 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.

From these experiments, we conclude that an absolute orientation sensor is critical for navigation over long distances, unless some other means is used to periodically update the robot position. If no orientation

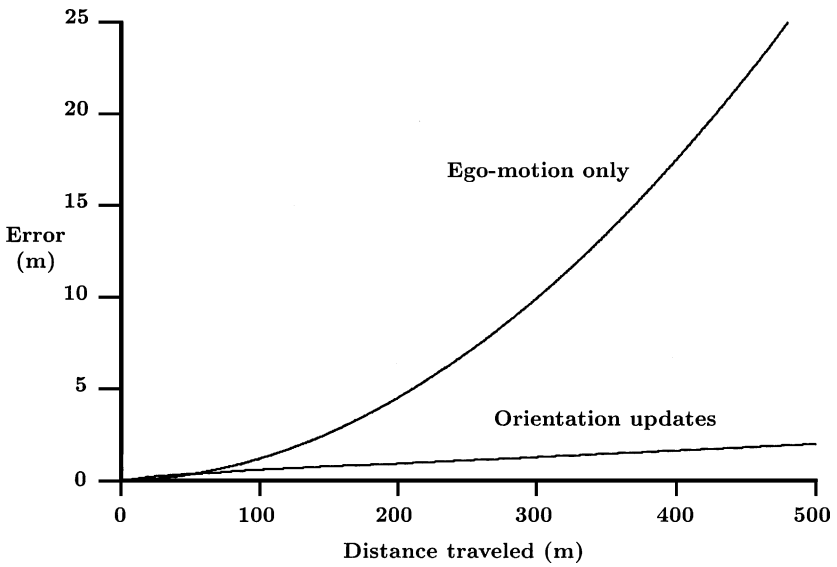

Fig. 3. Expected position error as a function of distance traveled.

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.

5. Robust estimation

In order to achieve accurate navigation over long distances, errors in the landmark position estimation and matching process must have a very small effect on each computed motion estimate. Landmarks must be chosen such that they are easy to track and yield little stereo error. Tracking must be performed such that mismatches are rare. When mismatches occur, there must be mechanisms for detecting and discarding them. We describe techniques for performing these steps here, while managing the overall error buildup over time and dealing with camera roll as the robot moves.

5.1. Optimized feature selection

Intuitively, one would expect for errors in stereo matching to produce larger errors in the motion estimate than errors in the landmark tracking. (Here we refer to the subpixel localization errors rather than mismatches.) The reason for this is that stereo error produces a larger effect in the estimation of each landmark position than error in feature tracking. A stereo mismatch by one pixel can yield a large change in the estimated position of the landmark, while a feature tracking error of one pixel usually results in a small change in the estimated position.

Our simulations have verified this effect. Fig. 4 shows the variation in the motion error over long distances as the stereo and feature tracking errors vary. For each plot, the error standard deviation for one of the matching steps was held constant at 0.3 pixels, while the other was varied. It can be observed that the navigation error varies much faster as the stereo error is changed than as the tracking error is changed.

While it is important to minimize both the stereo error and the tracking error, we conclude that navigation error is improved by performing landmark selection such that the localization precision along the x-axis has more weight than localization precision along the y-axis, since error in the y-direction has a lesser effect on the stereo error.

This has been implemented using a variation of the Förstner interest operator [6]. A feature is selected if the covariance ellipse of the feature localization is not highly elliptical, the precision of the feature localization is strong (with higher weighting on the horizontal

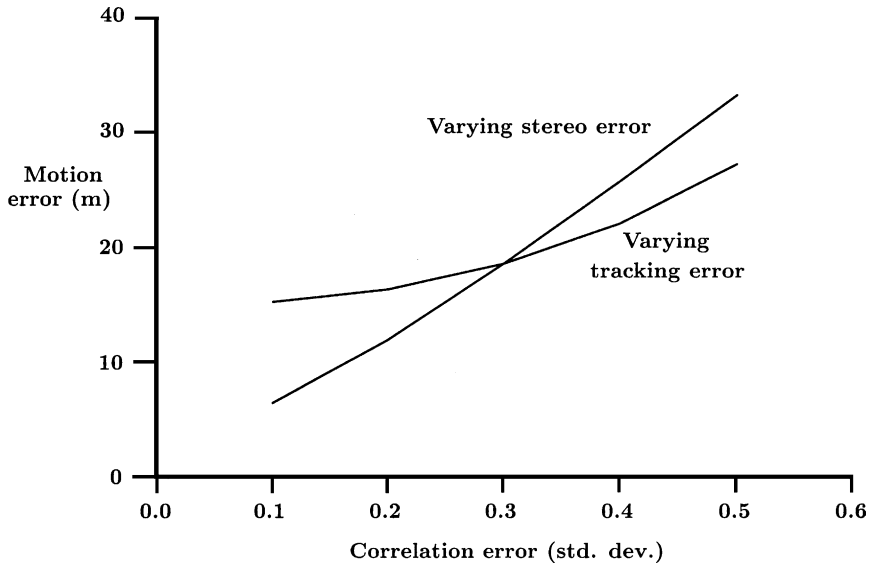

Fig. 4. Comparison of the effect of variation in stereo correlation error versus tracking correlation error.

precision), and there is no better feature within some bounded distance.

Further improvements can be achieved through the use of feature tracking measure that is more robust than the SSD measure [15]. However, we have not, yet, incorporated this measure into our robot implementation.

5.2. Improved feature tracking

In many environments, it is common for the landmarks that are selected to look somewhat similar to each other and other image locations. If a large search space is necessary for each feature, incorrect matches occur frequently, since the difference in the appearance of the landmarks after the camera motion may be greater than the difference in appearance between the landmark and other image locations. For this reason, it is important to limit the search space over which we search for landmarks. Of course, we cannot limit the search space to be so small that it does not contain the correct match.

An a priori estimate of each landmark position is obtained using the robot odometry estimate. However, errors in the odometry incur the need for a large search window. In order to decrease the size of this search window, we estimate the robot pitch and yaw errors by first detecting a landmark near the top of the image (and thus relatively far away for our applications) using a large template window. In this case, we use a large search window, but since the landmark is also large, we are able to avoid mismatches in the image. After correcting the robot pitch and yaw estimates such that the initial landmark match is correct, we can reduce the search windows for the later correlation steps, thereby reducing the chance of a false positive.

Within the reduced search windows, our experiments have indicated that correlation using a two-resolution pyramid with decimation by a factor of 4 provides the best combination of speed and tracking performance.

5.3. 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. This helps to filter out cases where a landmark is not present in the new image and cases where the change in appearance is so large that correct matching is not possible.

For each stereo match, the rays from the cameras through the image features are computed to determine if they are 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.) If this gap is not in front of the cameras, or if the projection of the gap into the image is larger than a pixel or two, the match can be rejected, since it is not geometrically feasible.

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.

Finally, outlier rejection is performed within the maximum-likelihood motion estimation procedure. After computing a motion estimate, the residual error for each landmark is determined. Once again, the worst matching landmarks are rejected if they have a residual greater than some threshold and the estimation is continued.

5.4. Multi-frame tracking

Matthies [14] has shown that the errors between successive motions are negatively correlated if the same landmarks are tracked through the images. We thus expect to have lower error when the same landmarks are tracked, rather than selecting new landmarks at each step. Of course, some landmarks must be replenished at each step, since some will move out of the field-of-view and some will be rejected as outliers. Our simulator experiments indicate that this effect is significant, even when there is only partial overlap between the landmark sets. The experiments showed a 27.7% reduction in navigation error when multi-frame tracking is used, rather than considering each pair of frames separately. This effect is thus useful in maintaining accurate navigation over long distances.

5.5. Camera roll

Camera roll (rotation about the viewing axis) 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 some 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.

6. Results

These techniques have been tested on hundreds of real stereo pairs in outdoor terrain with the robot undergoing six degree-of-freedom motion. Fig. 5 shows one complete cycle of the motion estimation process for a simple example of forward motion. Landmarks were selected automatically in the left image of the initial stereo pair. The selected landmarks appear to be well distributed in the area of the image expected to be seen after movement, although relatively few landmarks are selected close to the robot. The matching locations were then detected in the corresponding right image using stereo matching. Few, if any, of landmarks were discarded at this step through examination of the correlation score and the gap between the rays from the cameras. Next, the locations of the landmarks were predicted in the next image of the sequence. This step used an estimate of the camera motion and the estimated positions of the landmarks (from the stereo matching) in the prediction of the new image locations.

After correcting for pitch and yaw error, the actual locations of the landmarks were detected in the left and right images of the new stereo pair using the prediction positions to limit the search area for each landmark. Several landmarks were eliminated at this stage using the rigidity constraint. The remaining landmarks were used to determine the motion of the robot using the maximum-likelihood method described above. Finally, the landmark set was reduced by eliminating those features that were expected to move out of the field-of-view in the next step and replenished with new landmarks.

Fig. 6 shows landmark tracking for six consecutive frames of forward motion in rocky terrain. (Fig. 5 corresponds to the third step in this sequence.) Despite errors in the nominal camera movements and features that occur on occluding boundaries (making them 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 an extended sequence, we have applied them to images from a rover traverse consisting of 210 stereo pairs (among others). This traverse was performed with a small rover and a wide field-of-view, so the cameras were close to the ground and there was considerable distortion in the appearance of close-range locations. Fig. 7 shows an example of consecutive stereo pairs with 320×240 resolution. The rover traversed approximately 20 m, taking images about every 10 cm. 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. Fig. 8 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%.

7. Summary

We have examined techniques to perform stereo ego-motion robustly for long-distance robot navigation. Techniques for performing robust feature selection and tracking with outlier rejection have been

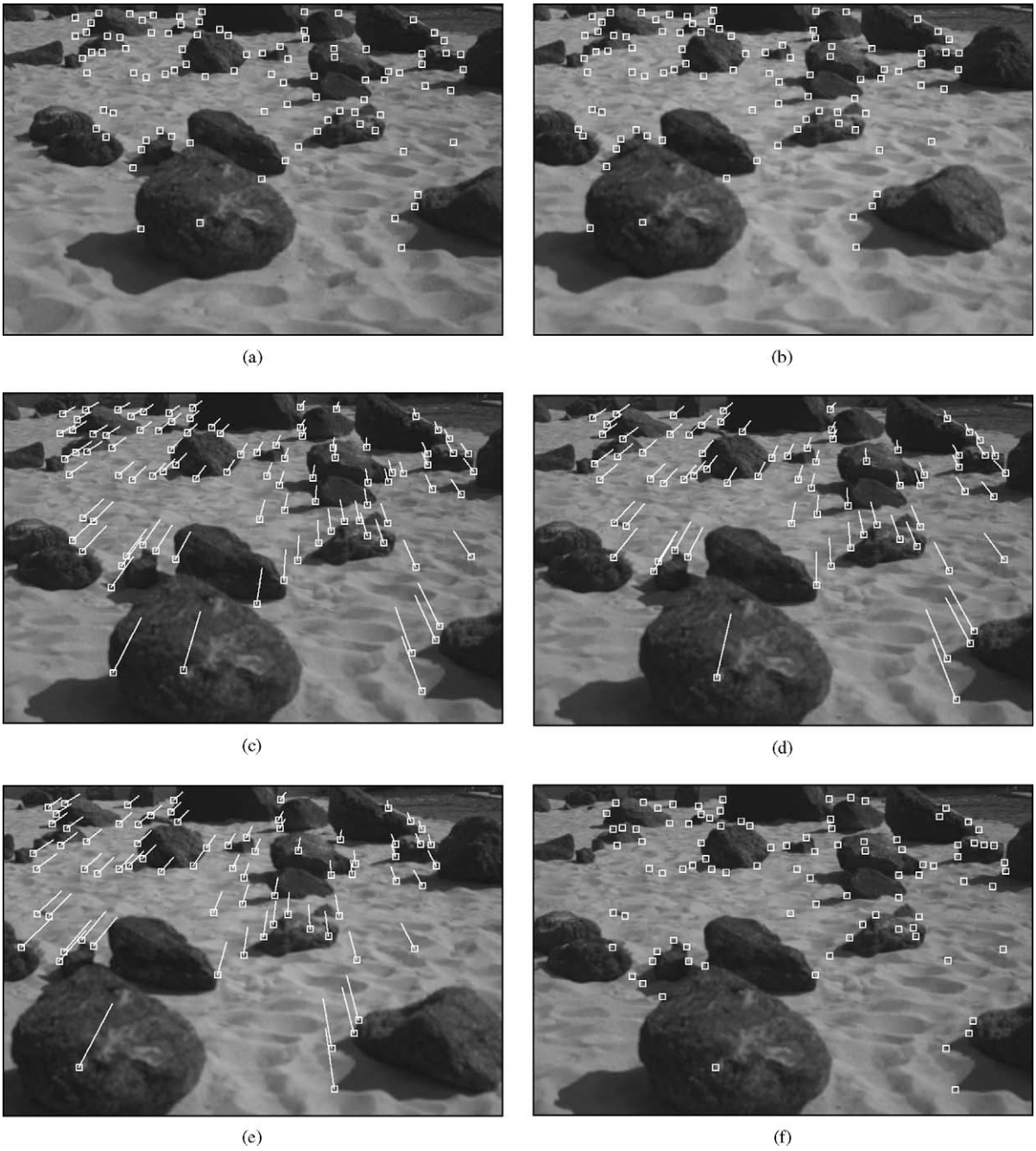

Fig. 5. One cycle of robust feature matching: (a) landmarks selected; (b) landmarks matched in right image; (c) predicted positions in next image; (d) matched positions in left image; (e) matched positions in right image; (f) landmarks after replenishment.

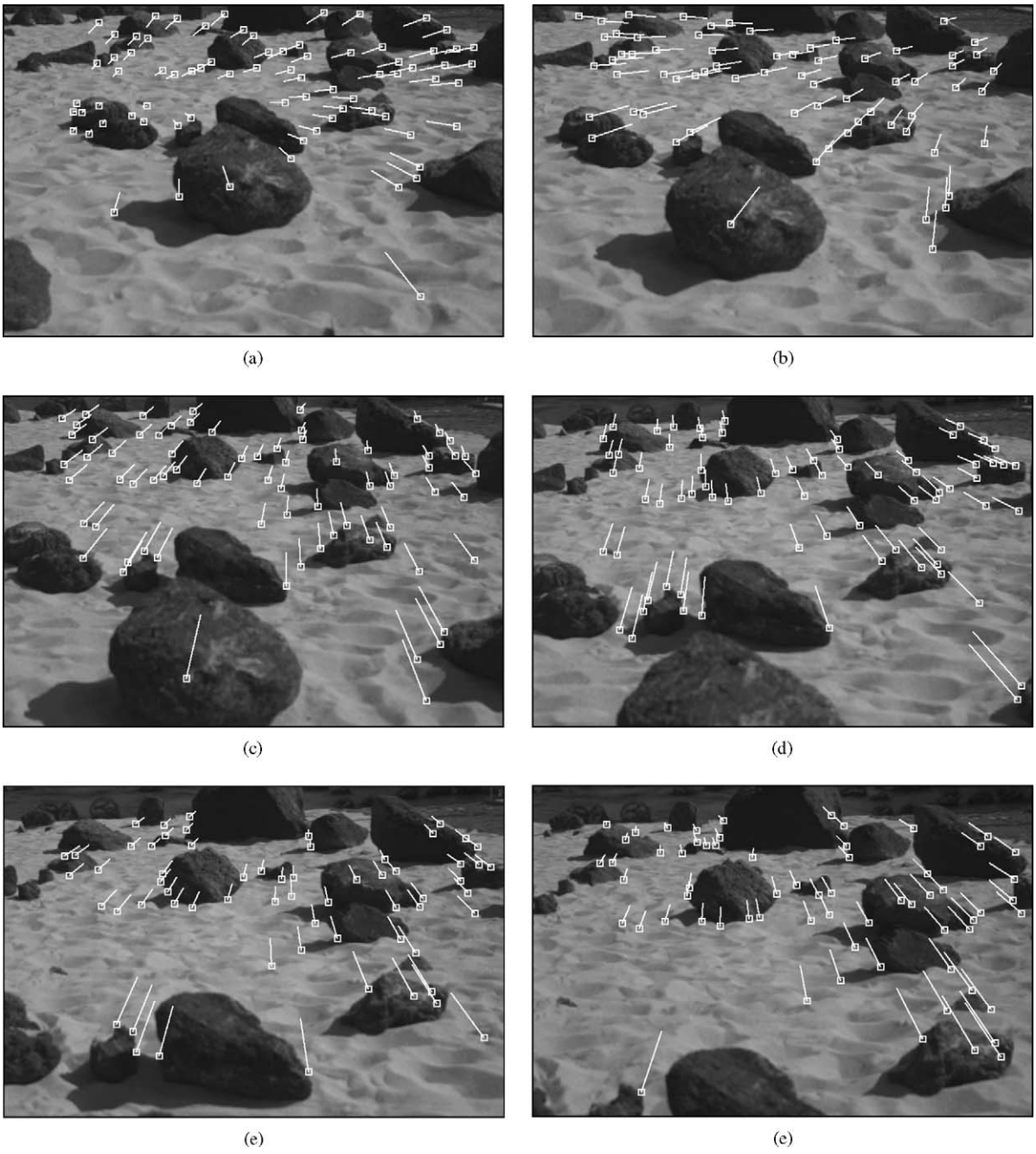

Fig. 6. Several cycles of robust feature matching for ego-motion. The squares indicate the tracked landmarks and the lines show the motion of the landmark from the previous frame.

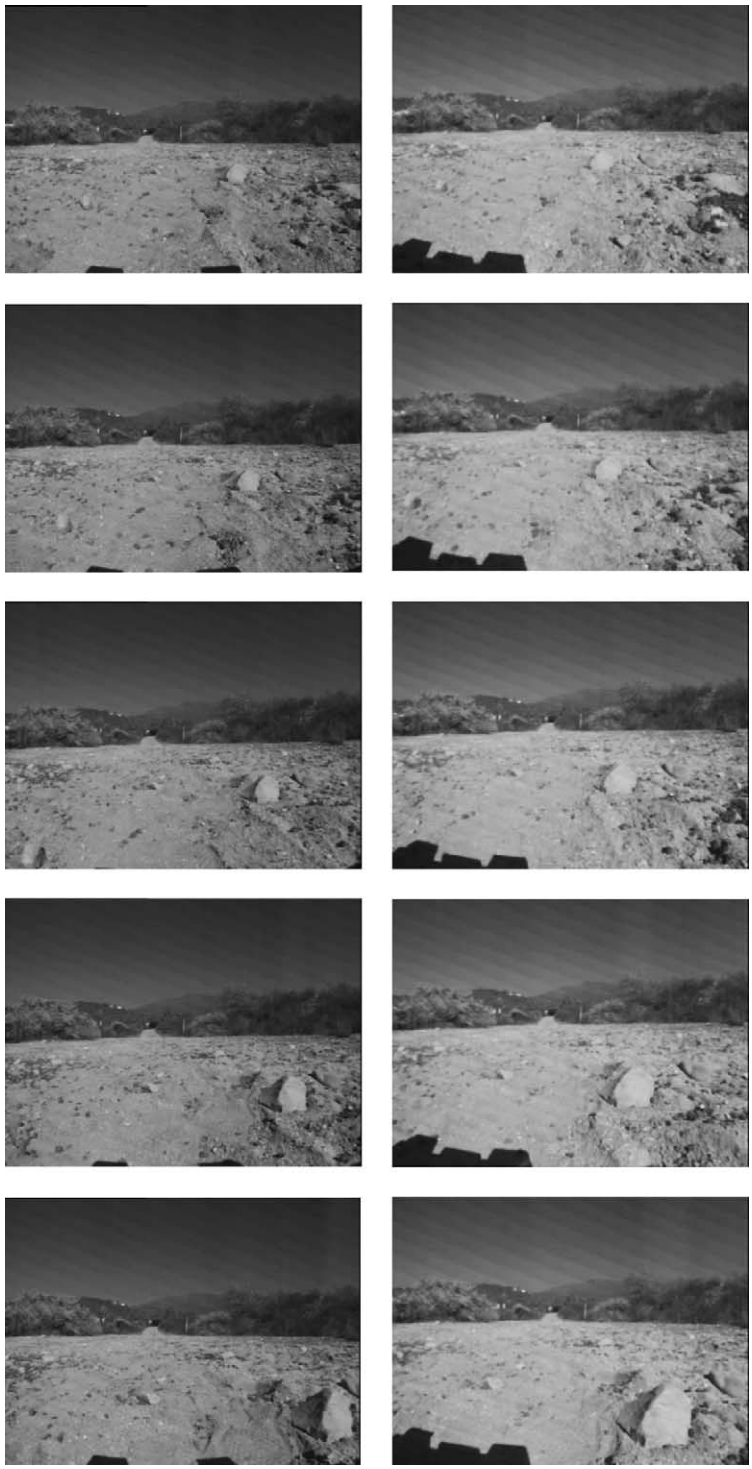

Fig. 7. Consecutive stereo pairs from a rover traverse sequence.

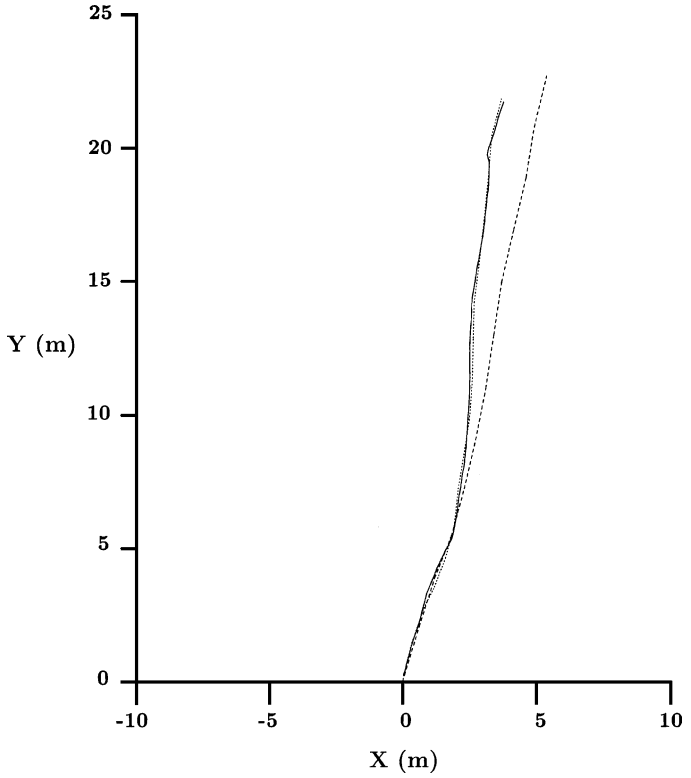

Fig. 8. An extended run consisting of 210 stereo pairs. The solid line is the GPS position of the rover. The dotted line is the ego-motion estimate. The dashed line is the odometry estimate.

developed in order to ensure accurate motion estimation at each step. An 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. The use of stereo data was also critical to elimination of outliers and accurate motion estimation. 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 in part at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. This paper is an expanded version of previous work on stereo ego-motion that has appeared in the IEEE Computer Society Conference on Computer Vision and Pattern Recognition [17] and the IEEE International Conference on Robotics and Automation [18].

References

[1] M. Betke, L. Gurvits, Mobile robot localization using landmarks, IEEE Transactions on Robotics and Automation 13 (2) (1997) 251–263.

[2] A.R. Bruss, B.K.P. Horn, Passive navigation, Computer Vision, Graphics, and Image Processing 21 (1983) 3–20.

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

[4] G. Dissanayake, P. Newman, H.F. Durrant-Whyte, S. Clark, M. Csobra, A solution to the simultaneous localisation and

map building (slam) problem, IEEE Transactions on Robotics and Automation 17 (3) (2001) 229–241.
[5] H.F. Durrant-Whyte, S. Majumder, M. de Battista, S. Thrun, S. Scheding, A Bayesian algorithm for simultaneous localisation and map building, in: Proceedings of the International Symposium on Robotics Research, 2001.
[6] W. Förstner, E. Gülch, A fast operator for detection and precise locations of distinct points, corners, and centres of circular features, in: Proceedings of the Intercommission Conference on Fast Processing of Photogrammetric Data, 1987, pp. 281–305.
[7] D. Fox, W. Burgard, S. Thrun, Active Markov localization for mobile robots, Robotics and Autonomous Systems 25 (3–4) (1998) 195–207.
[8] D.J. Heeger, A.D. Jepson, Subspace methods for recovering rigid motion. I. Algorithm and implementation, International Journal of Computer Vision 7 (2) (1992) 95–117.
[9] K. Kanatani, 3D interpretation of optical flow by renormalization, International Journal of Computer Vision 11 (3) (1993) 267–282.
[10] J.J. Leonard, H.F. Durrant-Whyte, Mobile robot localization by tracking geometric beacons, IEEE Transactions on Robotics and Automation 7 (3) (1991) 376–382.
[11] A. Mallet, S. Lacroix, L. Gallo, Position estimation in outdoor environments using pixel tracking and stereovision, in: Proceedings of the IEEE Conference on Robotics and Automation, vol. 4, 2000, pp. 3519–3524.
[12] R. Mandelbaum, G. Salgian, H. Sawhney, Correlation-based estimation of ego-motion and structure from motion and stereo, in: Proceedings of the International Conference on Computer Vision, vol. 1, 1999, pp. 544–550.
[13] L. Matthies, S.A. Shafer, Error modeling in stereo navigation, IEEE Transactions on Robotics and Automation 3 (3) (1987) 239–248.
[14] L.H. Matthies, Dynamic stereo vision, Ph.D. Thesis, Carnegie Mellon University, October 1989.
[15] C.F. Olson, Maximum-likelihood template matching, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, 2000, pp. 52–57.
[16] C.F. Olson, Probabilistic self-localization for mobile robots, IEEE Transactions on Robotics and Automation 16 (1) (2000) 55–66.
[17] C.F. Olson, L.H. Matthies, M. Schoppers, M.W. Maimone, Robust stereo ego-motion for long distance navigation, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, 2000, pp. 453–458.
[18] C.F. Olson, L.H. Matthies, M. Schoppers, M.W. Maimone, Stereo ego-motion improvements for robust rover navigation, in: Proceedings of the International Conference on Robotics and Automation, 2001, pp. 1099–1104.
[19] R. Talluri, J.K. Aggarwal, Mobile robot self-location using model-image feature correspondence, IEEE Transactions on Robotics and Automation 12 (1) (1996) 63–77.
[20] C. Tomasi, J. Shi, Direction of heading from image deformations, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1993, pp. 422–427.
[21] A. Trebi-Ollennu, T. Huntsberger, Y. Cheng, E.T. Baumgartner, B. Kennedy, P. Schenker, Design and analysis of a sun sensor for planetary rover absolute heading detection, IEEE Transactions on Robotics and Automation 17 (16) (2001) 939–947.
[22] R. Vidal, Y. Ma, S. Hsu, S. Sastry, Optimal motion estimation from multiview normalized epipolar constraints, in: Proceedings of the International Conference on Computer Vision, 2001, pp. 34–41.
[23] R. Volpe, Mars rover navigation results using sun sensor heading determination, in: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, vol. 1, 1999, pp. 460–467.
[24] R. Wagner, F. Liu, K. Donner, Robust motion estimation for calibrated cameras from monocular image sequences, Computer Vision and Image Understanding 73 (2) (1999) 258–268.
[25] J. Weng, P. Cohen, N. Rebibo, Motion and structure estimation from stereo image sequences, IEEE Transactions on Pattern Analysis and Machine Intelligence 8 (3) (1992) 362–382.
[26] T. Zhang, C. Tomasi, Fast, robust, and consistent camera motion estimation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, 1999, pp. 164–170.
[27] Z. Zhang, O.D. Faugeras, Estimation of displacements from two 3D frames obtained from stereo, IEEE Transactions on Pattern Analysis and Machine Intelligence 14 (12) (1992) 1141–1156.

Clark F. Olson received his B.S. degree in computer engineering in 1989 and M.S. degree in electrical engineering in 1990, both from the University of Washington, Seattle. He received his Ph.D. degree in computer science in 1994 from the University of California, Berkeley. After spending 2 years doing research at Cornell University, he moved to the Jet Propulsion Laboratory, where he spent 5 years working on computer vision techniques for Mars rovers and other applications. Dr. Olson joined the faculty at the University of Washington, Bothell, in 2001. His research interests include computer vision and mobile robotics. He teaches classes on the mathematical principles of computing and database systems, and he continues to work with NASA/JPL on mapping and localization techniques for Mars rovers.

Larry H. Matthies received his Ph.D. in computer science from Carnegie Mellon University in 1989 and has been at the Jet Propulsion Laboratory since then, where he is now supervisor of the Machine Vision Group. His research interests are in perception for autonomous navigation of robotic ground and air vehicles. At CMU, he developed the first vision system able to accurately estimate the ego-motion of

a mobile robot by tracking scene features with stereo cameras on the robot. At JPL, he developed the first successful real-time stereo vision system for autonomous, cross-country obstacle avoidance by mobile robots, in work under the NASA Mars Rover program in 1990; a refined version of this algorithm will fly on the Mars Exploration Rover (MER) mission to be launched in June 2003. His group pursues space-oriented research on recognizing craters as landmarks for orbit estimation and precision landing in missions to Mars and to asteroids, hazard detection and relative velocity estimation for safe landing on Mars, and Mars rover obstacle detection and localization. He also conducts research for earth-based robots on autonomous navigation in indoor/outdoor urban areas and on obstacle detection for off-road, day/night, all-weather autonomous navigation. He is an Adjunct Professor in the Computer Science Department at the University of Southern California and a member of the editorial board of the Autonomous Robots Journal.

Marcel Schoppers obtained his M.Sc. in 1982 from the Australian National University in Canberra, Australia, and his Ph.D. in 1989 from the University of Illinois at Urbana–Champaign. Along the way he worked at XEROX PARC, SRI International, and Advanced Decision Systems. From 1992 to 1997 he was CEO, Supreme Scientist and Benevolent Dictator of Robotics Research Harvesting, a small defense contracting company. Since 1997 he has been at the NASA Jet Propulsion Laboratory, usually designing and implementing flight software for Mars rovers.

Mark W. Maimone is a Machine Vision researcher at the Jet Propulsion Laboratory. In 1989 he completed the International Space University's summer program, in which he participated in the design of a lunar polar orbiter. He earned his Ph.D. in computer science from the Computer Science Department of Carnegie Mellon University in 1996, and was then a post-doctoral research associate at Carnegie Mellon's Robotics Institute, serving as Navigation and Software Lead for the 1997 Atacama Desert Trek. Since starting at JPL in 1997, he has worked on the several Mars Rover research projects, and a vision system for inspection of the Chornobyl reactor. Mark is currently a member of the 2003 Mars Exploration Rover flight software team, and is developing the vision and navigation subsystems for the robots that NASA will send to Mars in 2003. His research interests include robotic navigational autonomy, stereo vision, camera calibration, and software environments.