Mobile Robot Self-Localization by Matching Range Maps Using a Hausdorff Measure

Clark F. Olson

Jet Propulsion Laboratory, California Institute of Technology

4800 Oak Grove Drive, MS 107-102, Pasadena, CA 91109

http://robotics.jpl.nasa.gov/people/olson/homepage.html

Abstract

This paper examines techniques for a mobile robot to perform self-localization in natural terrain by comparing a dense range map computed from stereo imagery to a range map in a known frame of reference. The range map is first processed to generate a three-dimensional occupancy map of the terrain. This occupancy map is then compared to a similar map from a previous robot position or a global occupancy map of the environment that has been previously generated. The best relative position between the maps according to a Hausdorff measure is determined using efficient search techniques. These techniques allow localization of mobile robots in natural terrain that is robust to noise, (?7'(11', scene clutter, and missing data.

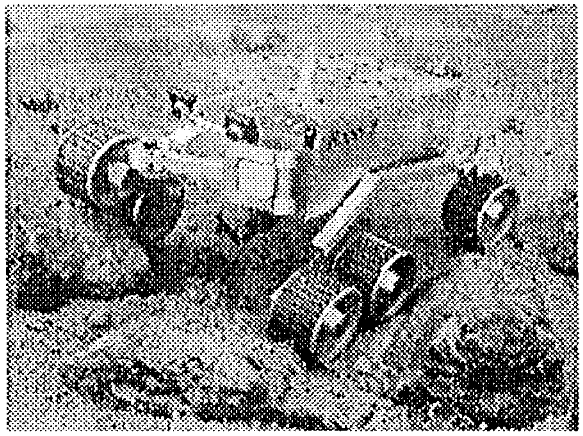


Figure 1: The Rocky 7 Mars rover prototype

1 Introduction

In this paper, we consider the problem of determining the position of a mobile robot with respect to a known frame of reference by comparing the range map computed from a stereo pair of images taken at the robot's current location to a range map from a previous location or to a composite range map of the environment that has been previously generated.

Our motivation for studying this problem is to increase the autonomy of the Rocky 7 Mars rover [19]. See Figure 1. While the position of the rover is continuously updated using dead-reckoning from wheel encoders and an angular-rate sensor, wheel slippage and sensor drift cause an accumulation of error in this estimated position [10]. It is thus desirable to have additional means for periodically localizing the rover to correct this accumulated error. Previous techniques that have been used to localize Rocky 7 have concentrated on imaging the rover from the lander that will carry the rover to the Mars surface [20], which limits the operable range of the rover to a small area around the lander. To enable operation of the rover at distances far from the lander, autonomous localization" procedures are necessary and we consider the use of stereo vision for this problem here.

A mast system has recently been integrated into Rocky 7 that allows (among other operations) stereo pairs to be taken from a height of approximately 1.5 meters above the ground, in addition to the stereo pairs that are taken from the navigation cameras a few centimeters above the ground. See Figure 2. Such stereo pairs allow the generation of a range map of the immediate surroundings of the robot. The premise of this work is that we can robustly determine the position of the robot in natural terrain by comparing the range map computed at the robot's local position with a range map encompassing the same terrain for which we know the frame of reference.

While previous work has also explored matching maps to perform localization (e.g. [3, 4, G, 17]), previous work has considered either man-made environments and/or used search techniques that required an

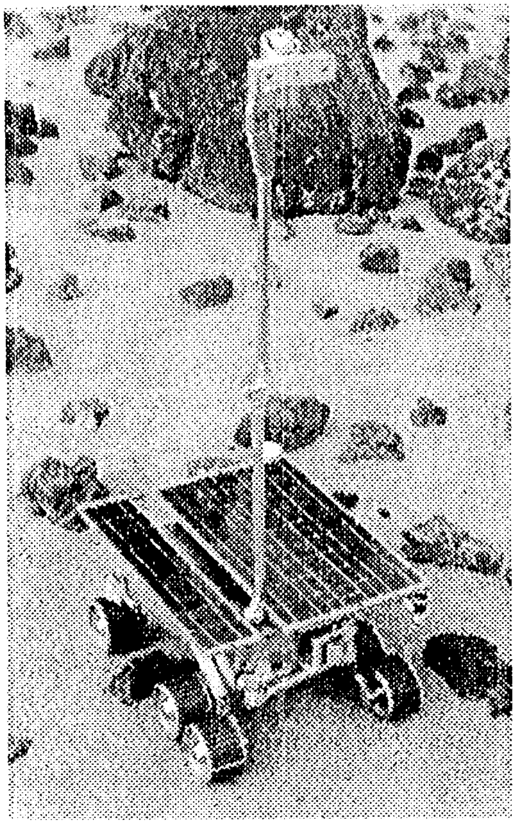


Figure 2: Rocky 7 with the mast deployed

initial estimate of the robot's position and could reach a locally optimal position estimate that was not globally optimal. A review of position estimation techniques for mobile robots can be found in [18]. The techniques that we describe here can operate in natural environments using a three-dimensional map and are guaranteed to find the globally optimal solution with respect to the matching measure that is used.

The balance of this paper explores these techniques in greater detail. Section 2 discusses the process by which the range maps of the robot's surroundings are computed and methods to transform these range maps into a voxel representation that allows robust and efficient matching. Section 3 discusses our use of Hausdorff matching techniques [7, 8] to find the relative position between the maps such that the maximum number of voxels match up to a given error bound. Section 4 gives an example where these techniques have been used to simulate determining the relative position of the Rocky 7 Mars rover in Mars-like terrain. Finally, Section 5 summarizes this work.

2 Computing range maps

We compute range maps from stereo pairs using passive stereo vision [9]. It is assumed that the robot cameras have been calibrated off-line. Rocky 7 uses a camera model that allows arbitrary affine transformations of the image planes [21] and that has been extended to include radial lens distortion [5]. The images are first warped to remove the lens distortion and the images are rectified so that the corresponding scan-lines yield corresponding epipolar lines in the image. Disparities are measured between the images for each pixel by minimizing the sum-of-squared-difference (SSD) measure of a window around the pixel over a finite disparity range. Sub pixel disparities are computed by fitting a parabola to the SSD values at the triple of disparities centered at the discrete pixel minimum. The parabola minimum is taken to be the sub-pixel disparity estimate. Smoothing is performed over a 3×3 window to reduce noise. Incorrect matches are filtered out in this process using both a left-right-line-of-sight consistency check and a process to remove small patches where the disparities do not agree with surrounding values. Given the disparities, the coordinates of each pixel are computed by triangulation. Details of these techniques can be found in [9, 11].

Once a range map has been computed from the stereo imagery, we convert it into a voxel-based map representation. We first rotate the data such that it has the same relative orientation as the map we are comparing it to. Here we operate under the assumption that the orientation of the robot is known through sensors other than vision (for example, a sun sensor, accelerometer, and gyrocompass have been incorporated into Rocky 7). For testing, and in case the accuracy of the sensors is lower than desired, we have used a simple technique for determining the orientation of the ground plane, assuming that ground is relatively flat. This technique simply determines the two principal components of the range points that are detected in the image and rotates them such that they are parallel to the xy-plane.

The next step is to bin the range points in a two-dimensional grid covering the xy-plane at some specified scale. We approximate the terrain 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. Now, we can eliminate the need to search over the possible translations of the robot in the z-direction by subtracting a local average of the terrain height from each cell (i.e. a high-pass filter). This step is not strictly necessary,

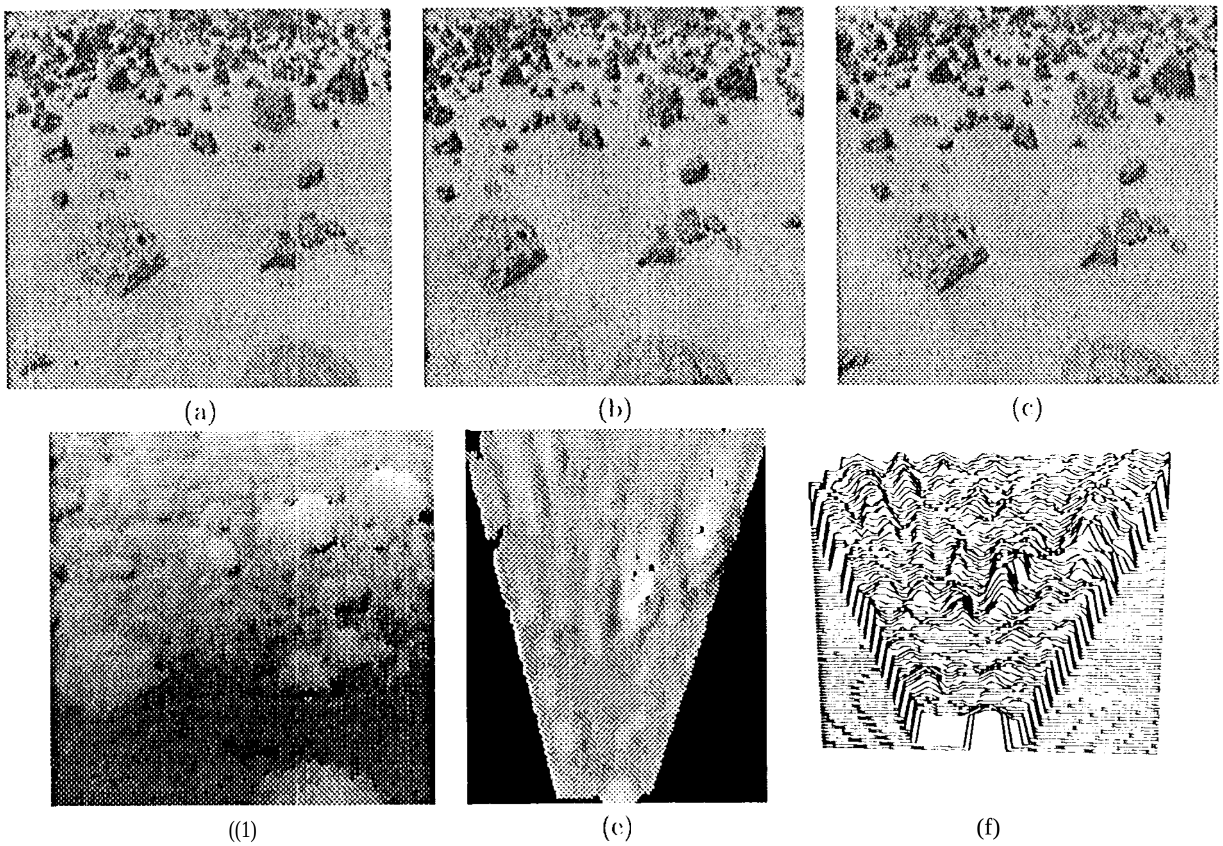


Figure 3: Range maps are computed using stereo vision. (a) Left image of a stereo pair. (b) Right image of a stereo pair (c) Left image after warping to remove radial lens distortion. (d) Height of pixels determined using stereo triangulation (e) Heights after binning in the xy-plane. (f) Surface extracted from the pixel heights.

and it reduces our ability to determine height changes in the position of the robot, but it also reduces the computation time that is required to perform localization. Finally, we perform smoothing on this two dimensional grid in such a way that bins that were not hit by any range pixel (e.g. due to sparseness of the range pixels) are given values, but otherwise do not contribute to the smoothing.

To facilitate matching using a Hausdorff measure, we transform this 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. For each column in the z-direction, the cell corresponding to the height of the surface at this location is said to be *occupied*, and the others are *unoccupied*.

Figure 3 shows several of the intermediate steps in this process.

3 Matching range maps

Once the occupancy map has been computed for the current position of the robot, we need to find the best relative position between this map and a map that was computed for a previous position of the robot, or a composite map that has been made of the robot's operating environment (possibly through combining maps taken from the robot's previous locations). We use an image matching technique based on the Hausdorff distance [7].

3.1 The Hausdorff distance

For two sets of points, A and B, the directed Hausdorff distance from A to B is:

$$h(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\|$$

where $\|\cdot\|$ is any norm. This yields the maximum distance of a point in set A from its nearest point in set B. For image matching, we wish to allow at least a small fraction of outliers that do not match well. We may thus use:

$$h_K(A,B) = K^{\text{th}}_{a \in A} \min_{b \in B} \|a - b\|$$

This determines the partial Hausdorff distance among the K points in A that best match points in B (we thus allow $|A| - K$ outliers in the set A). This measure is asymmetric, as it does not consider how well each of the points in B is fit by A. This allows matching to be performed against a large map, where the map generated at the local robot position is contained as a subset of this map.

A variation on this measure is to determine the maximum number of points in the local map such that the measure is below a given error threshold:

$$h_K(A,B) \leq \delta$$

This formulation is often easier to work with, since to compute this number, we must only count the number of points in A that match some point B up to the error, δ. We thus use this formulation in this work.

3.2 Searching for the best match

While previous Hausdorff matching methods [7, 8, 12, 16] have been applied to matching two dimensional image edge maps, we can apply essentially the same techniques to matching three-dimensional image surface maps.

In this method, the space of the possible relative positions between the maps is discretized. Each occupied cell in the maps is represented by the single point at the center of the cell. Since we search over only translation in the x- and y-directions, an obvious discretization exists such that each discrete position aligns the centers of various grid cells between the two maps. This discretization is guaranteed to find the optimal solution if we use the L_1 or L_∞ norm in our matching measure with an error, δ, that is an integral number of pixels.

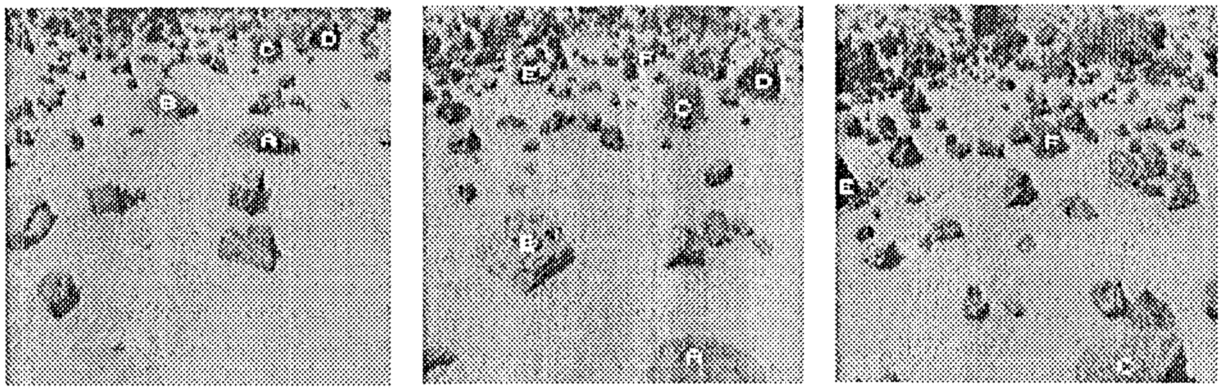
We could now examine each possible relative position between the maps in this discretization to determine which is optimal, but this method would be computationally expensive. We instead use a multiresolution search technique that has proven useful in object recognition and extraction of geometric primitives [2, 8, 12, 13, 16]. The basic idea is to consider the space of possible relative positions as a set of rectilinear cells, each of which covers many positions. Each cell is tested to determine whether it could contain a position that satisfies some matching criterion. If it is determined that the cell cannot contain such a position, then it is pruned. Otherwise, the cell is divided into subcells and the process is repeated recursively. When a cell is reached that contains only a single position in the discretization, this position is tested explicitly. Note that, since we are seeking the single best relative position between the maps, our matching criterion is adaptive. The criterion becomes stricter as we find positions of increasing quality in the search.

The key to this method of searching the parameter space is a quick method to conservatively test whether a cell can contain a position satisfying the matching criterion. 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. To accomplish this, we examine the distance transform [1, 14, 15] of the occupancy map for the known frame of reference.

First, the occupancy map is dilated by a box centered at the origin with $2\delta + 1$ pixels on each edge. This operation ensures that each cell within δ in each direction of an occupied cell in the original map is also occupied. Next, a distance transform of this map is computed. This distance transform measures the distance from each cell in the map to the closest occupied cell that lies in the same horizontal plane (since we search only in x and y). Now, a probe into this distance transform will yield 0 if the cell is within δ of an occupied cell in the undilated map, and otherwise yields the distance to the closest occupied cell in the dilated map.

Consider the set of distances that are obtained by probing the distance transform at the position of each of the occupied cells in our local map at the current robot position according to some relative position between the maps. If this set has K zero values, then at least K cells in the local map are within δ of occupied cells in the previous map. Otherwise, the K th largest value yields a bound on the closest distance to a position that could yield K zero values [7].

We can use these ideas to formulate an efficient test for a cell in the parameter space in the following manner. Let us say that the best position that has been found so far yields B cells in the local map that match the global map up to the allowed error (i.e. B probes into the distance transform for this position yield zero). To test a cell, we first determine the discrete position closest to the center of the parameter space cell. We then determine the distance between this position and the furthest corner of the cell. Denote this distance D_c. We now probe the distance



(a) (b) (c)

Figure 4: A test case in the JPL Mars yard for the localization techniques. The relative position between the first and second stereo pair is approximately 2.0 meters forward. The relative position between the second and third stereo pair is approximately 2.5 meters forward. Some rocks have been labeled with letters to assist the reader in making correspondences. (a) Left image of the first stereo pair. (b) Left image of the second stereo pair. (c) Left image of the third stereo pair.

transform at the locations of the local map with respect to the relative position at the center of the cell. If these probes yield less than B values that are not greater than D_c, we can prune this cell of the parameter space, since it cannot yield a position at which B cells in the current map match occupied cells in the previous map up to an error of δ.

For any cell that cannot be pruned, we divide the cell in both x and y, and repeat the process recursively on the subcells. When we reach a cell that contains a single position in the discretized pose space, we test this position explicitly. If the position yields more than B matches, then we store this position as the best found so far, increase B to the new value, and continue the search. This continues until all of the cells have been exhausted, at which point we are guaranteed to have found the best relative position between the maps according to variant of the Hausdorff measure we use.

4 Results

We have tested these techniques with images taken in the JPL Mars yard[1] using cameras mounted on a tripod at approximately the Rocky 7 mast height. Figure 4 shows an example test case that simulates approximate forward motion of the rover. The second stereo pair in this test case was taken approximately 2.0 meters forward from the first stereo pair. The third stereo pair was taken approximately an additional 2.5 meters forward from the second stereo pair. Note, in particular, the difficulty of performing localization of the third stereo pair in terms of the second, since few significant landmarks are present in both images.

Our localization techniques have yielded accurate results in these tests. In this example, the estimated position change in the first case is 2.037 meters forward and the estimated position change in the second case 2.517 meters forward, which are close to the measured positions of approximately 2.0 meters forward and 2.5 meters forward, respectively.

While the current implementation of these techniques runs on a workstation, they are presently being ported to the Rocky 7 hardware for full testing. The workstation implementation requires only a few seconds to perform all of the computation, including warping the images, computing the stereo range map, determining the ground plane, building the occupancy map, computing the distance transform of the map in the known frame of reference, and performing the search for the best relative position between the maps.

5 Summary

This paper has considered self-localization techniques for a mobile robot in natural terrain through the use of stereo vision. The robot's position is determined by comparing a terrain map computed at the robot's current location to a terrain map in a known frame of reference. We first generate a dense range map from stereo imagery and then process this data to create an occupancy map of the terrain surface. The

[1] See http://robotics.jpl.nasa.gov/tasks/scirover/marsyard

best relative position between this occupancy map and the occupancy map in the know n fi ame of referei ice i s determined with respect to a Hausdorff measure, which yields robustness to noise, error, scene clutter, and missing data. The optimal position is found using a sear a I strategy that recursively divide s and prunes the space of possible relative positions. An impoi tant featui e of this search strategy is that it can determine the optimal position without requiring an initial esti - mate of the position of the robot.

Acknowledgments

The author thanks Bob Balaram and Larry Matthies for insightful discussions on this topic. Todd 1 ,itwin developed the camera calibration and binocular stereo code.

The research described in this papei w a s carried out by the Jet Propulsion I ,aboratoi y, Califoi nia Institute of Technology, under a conti act with the National Aei onautics and Space Administi ation.

References

[1] G. Borgefors. Distance transform ations in digital images. *Computei Vision, Graphics, and Image Processing*, 34:344 371, 1986.

[2] '1'. M. Breuel. Fast recognition using adaptive subdivisions of trans formation space. In *Proceedings of the IEEE Conference on Computei Vision and Pattern Recognition*, pages 445 451, 1992.

[3] A . Elfes. Sonar-based1 real-world mapping and navig at ion. *IEEE Journal of Rob otics and Automat ion*, 3(3):249 265, June 1987.

[4] D. B. Gennery. Visual terrain matching for a mars rover. In *Proceedings of the IEEE Conference on Computei Vision and Pattern Recognition*, pa ges 483 491, 1989.

[5] 1). B. Gennery. Camera calibration includi ng lens dis tortion. JPL internal report 1)-8580, Jet Propulsion Laborat ory, California Institut e of Technology, 1991.

[6] M. Hebert, C. Caillas , E. Krotkob, 1. S. Kweon, an d T. Kanac le. Terrain mapping for a roving planetary e xplorer. In *Proceedings of the IEEE Conference o71 Robotics and Automation*, pages 997 1002," 1989.

[7] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge. Comparing ii nages using the Hau sdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(9):850 863, Sej)tember 1993.

[8] 1). P. Huttenlocher and w. J. Rucklidge. A multiresolution technique for comparing images using the Hausdorff distance. In *Proceedings of th e IEEE Con - ference on Computei Vision ai id Pattern R ecognition*, pages 705 706, 1993.

[9] L. Matthies. Stei eo vision for planetary rovers: Stoch astic modeling to ne ar real-time implementation. *International Journal of Computei Vision*, 8(1):71 91, July 1992.

[10] L. Matthies, E. Gat, R. Harrison, B. Wilcox, R . Volpe, and T. Litwi n. Mars microrover navigation: Performance evaluation and enhanc ement. *Autonomous Robots*, 2(4):291- 312, 1995.

[11] 1,. Matthies, A. Kelly, and T. Litwin. Obstacle detection for unmanned ground vehicles: A progress report. In *Proceedings of the International ,%J1(I]1OS%IL1(1 on Robotics Reseai ch*, 1995.

[12] C. F. Olson and I), 1', Huttenlocher. Automatic target recognition by matching oriented edge pixels. *IEEE Transactions o71 Image Processing*, 6(1), Jan - uary 1997.

[13] C. F. Olson and 1,. H. Matthies. Locating geometic primitives by pruning the parameter space. Submitted to CVPR'97.

[14] 1). W. Paglieroni. Distance transforms: Properties and machine vision applications. *CVGIP: Graphical Models and Image Processing*, 54(1):56 74, January 1992.

[15] A . Rose nfeld and J. Pfaltz. Sequential operations in dig ital picture processing. *Journal of the Association foi Computing Machinery*, 13 :471- 494, 1966.

[16] W. J. Rucklidge. Locat ing objects usii ig the Hausdorff distance. In *Proceedings of the Int ernational Confei - ence o71 Computei Vision*, pa ges 457 464, 1995.

[17] R. Szeliski. Estimating motion from sparse range data without coi respondence. In *Proceedings of the International Conference on Computei Vision*, pages 207 216, 1988.

[18] R. Talluri and J. K. Aggarwal. Position estimation techniques f o r an autonoi nous mobile robot A review. In C. H. Chei 1, 1,. F. Pau, and P. S. P. Wang, editors, *Handbook of Pattern Recognition and Com - putei Vision*, chapter 4.4, pages 769 801. World Scientific, 1993.

[19] R. Volpe, J. Balaram, T. 011111, and R. Ivlev. T he Rocky 7 Mars rover prototype. Ii I *Proceedings of the IEEE/RSJ Int ernational Conference on Int elligent Robots and Syst ems*, 1996.

[20] R. Volpe, '1'. Litwin, and 1,. Matthies. Mol)ile robot loc alization by remote viewing of a colored cylinder. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 1995.

[21] Y. Yakimovsky and R. Cunningham. A system f o r extracting three-dimensional measu rements from a stereo pair of tv cameras. *Computei Vision, Graphics, and Image Processing*, 7:195 210, 1978.