Proceedings of the 12th International Conference on Advanced Robotics, pages 376-383, 2005.

Wide-Baseline Stereo Experiments in Natural Terrain

Clark F. Olson
University of Washington
Computing and Software Systems
Box 358534
18115 Campus Way NE
Bothell, WA 98011
cfolson@u.washington.edu

Habib Abi-Rached
University of Washington
Electrical Engineering
Box 352500
253 EE/CSE Building
Seattle, WA 98195
habib@hitl.washington.edu

Abstract

We have developed a wide-baseline stereo vision technique for Mars rovers in order to map terrain distant from the rover and allow localization and navigation over large areas. This technique uses two images captured by the same camera at different rover positions as a virtual stereo pair. The larger baseline yields better accuracy for distant terrain, but makes stereo matching more difficult. In addition, calibration of the relative camera positions in advance is no longer possible and odometry errors result in uncertainty in the camera positions in practice. Our methodology addresses these problems. We have tested these techniques in several locations containing natural terrain (mostly desert). This paper describes the experiments and discusses the successes and failures of the wide-baseline stereo technique.

1 Introduction

Autonomous navigation over large distances has been a longstanding goal for Mars rovers. A rover that is able to traverse to a distant science target in a single command cycle could greatly increase the amount of scientific data that the rover could obtain in a fixed lifespan. One of the barriers to this goal has been the inability to perform mapping and localization over significant distances, since the conventional stereo vision techniques used by current rovers have a limited range of usefulness.

We have studied several methods for improving rover localization and mapping capabilities [4, 5, 8, 9]. Recently, we have implemented a wide-baseline stereo vision technique for Mars rovers that combines motion estimation and robust image matching in order to map distant terrains [7].

Stereo error generally increases with the square of the distance to the terrain, but decreases with the baseline distance (the distance between the cameras). Wide-baseline stereo [11, 12, 13] is based on the idea that an arbitrarily large baseline can be achieved using two images from the same camera, but at different positions. While this improves the quality of the stereo range estimates for distant terrain, it introduces two new problems. Stereo calibration can no longer be performed in advance, since the relative positions and orientations of the cameras are not fixed. Furthermore, the poses where the images are taken are not known with high accuracy owing to odometry errors. In addition, performing stereo matching between the images is more difficult, since the terrain is seen from different viewpoints.

Our methodology is able to address these issues. To accomplish this, we perform the following steps:

1. **Motion refinement.** We use an estimate of the relative camera positions from odometry, but this estimate is refined using matches detected between the images. This process uses several substeps:
   (a) **Feature selection.** Features in one of the images are selected using the Förstner interest operator [1].
   (b) **Feature matching.** Matches for the selected features are detected using a hierarchical search over the entire image. Note that both images are high-pass filtered to remove some illumination effects.
   (c) **Outlier rejection.** Outliers are rejected by finding matches where the vertical and/or

horizontal disparity is not in agreement with nearby points.

(d) **Nonlinear motion estimation.** We optimize the motion parameters using the Levenberg-Marquardt method [10] with a robust objective function that minimizes the distances between the matched point locations and the locations backprojected using the estimated motion parameters.

2. **Image rectification.** The images are rectified so that the stereo match for each point lies on the same scanline [3]. This reduces the search space for each point to one dimension.

3. **Stereo matching.** Dense stereo matches are detected using a maximum-likelihood image matching formulation [6]. Efficient stereo search techniques are used to reduce the search time. Finally, subpixel disparities are computed and outliers are rejected using a technique that eliminates small regions of nonconformity.

4. **Triangulation.** The position of each pixel relative to the camera is computed from the disparity using standard triangulation techniques.

The following sections summarize the steps above and discuss experiments performed with this technique in natural, desert terrain.

2 Feature selection and matching

Motion estimation requires a set of features matches between the images. To accomplish this, we first select up to 256 features that appear to be matchable from one of the images using the Förstner interest operator [1]. For each image window, this operator considers both the strength and the isotropy of the gradients in the window. The features are selected in subregions of the image, to ensure that the features are well distributed in the image. They are also subject to a threshold, so that completely featureless regions do not contribute.

For each selected feature, we use a hierarchical multi-resolution search to locate the feature in the second image. This search first selects candidate matches at a low resolution. Each candidate match is then refined at a higher resolution. Optionally, affine matching is performed to further improve the quality of the matches. Finally, the best candidate is selected according to the scoring function (sum-of-absolute-differences in our implementation).

After matching has been performed, we prune some matches from consideration. Two tests are used. First, if there is another candidate with a similar score the top candidate, we prune the match, since either could be correct. Second, we estimate the standard deviation of the match position by examining how quickly the score changes as the match is moved. If the standard deviation is large, we prune the match. We also remove outliers by comparing the disparities of the nearby points.

3 Motion refinement

Once matches have been determined between the images, we can refine the estimated motion between the camera positions by enforcing geometrical constraints. We create a state vector consisting of the five recoverable motion parameters (the scale is not recoverable from this information) and the depth to each of the feature points with respect to the first camera position. Our objective function uses the distance from each matched feature position to the backprojected position using the motion parameters and feature depth. The distances are combined in an M-estimator, such as described by Forsyth and Ponce [2], although we have found that a larger estimate for the scale more often yields convergence to the correct result. Given the state vector and object function, we use a variation of Levenberg-Marquardt optimization in order to refine the input motion estimate.

4 Disparity estimation

Subsequent to the motion estimation, we rectify the images using the method of Fusiello *et al.* [3]. After rectification, the correspondences between the images will lie on the same scanline, so that the search for each match can be performed in one dimension. We find dense matches between the images using a robust template matching method [6]. This method has two advantages over typical matching methods, such as normalized correlation or the sum-of-squared-differences (SSD). Given a potentially matching location, many template matching methods assume that the mapping between pixels in the template and the search image is simply a translation, with no warping or nonlinearity, such as caused by occlusion. Our method is more general and allows pixels in the template to match the

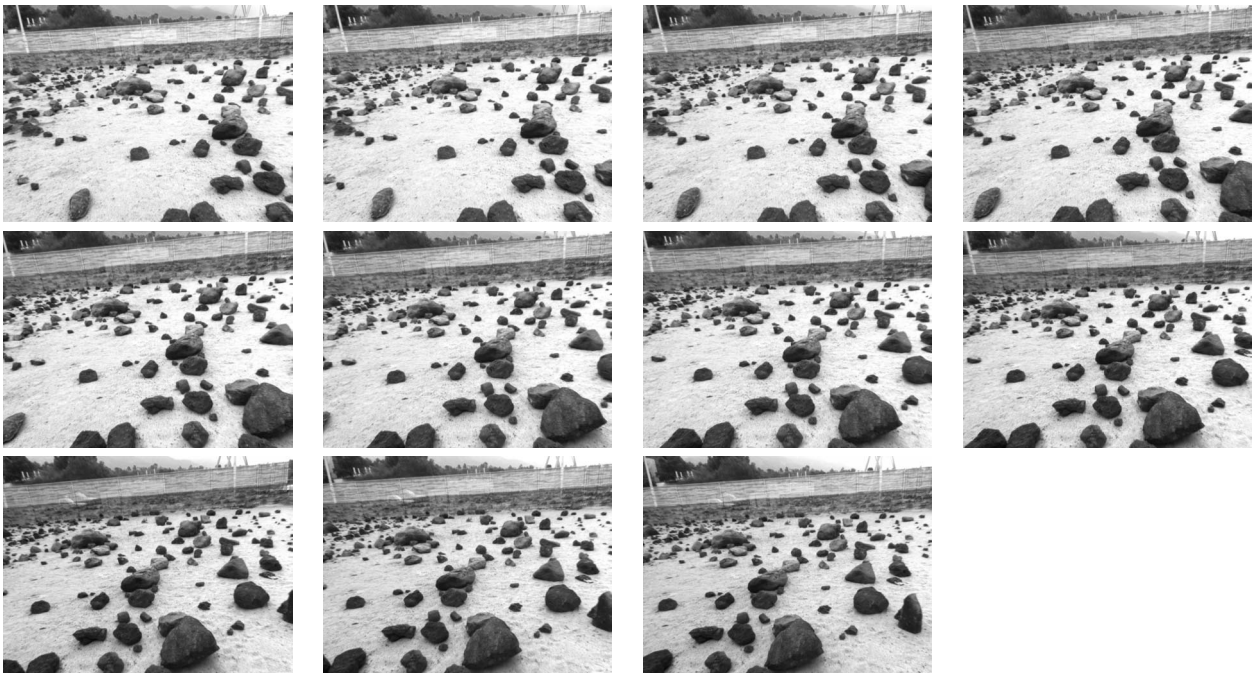

Figure 1: Image sequence from the JPL Mars Yard. The images were taken at intervals of approximately 20 cm.

search image at locations other than the implicit translation given by the relative position of the template. We use a measure that combines distance in the image with distance in intensity using a linear relationship.

The second improvement is that this measure explicitly accounts for the possibility of a pixel in the template with no corresponding match in the other image (for example, due to occlusion, which is common in wide-baseline stereo pairs.) We accomplish this using a maximum-likelihood measure that incorporates the probability distribution function of both pixels with matches and pixels without matches.

Matching is performed efficiently using this measure by adapting a common stereo search strategy. This strategy prevents redundant computations from being performed for adjacent templates at the same displacement in the search image by saving information. The cost is a small amount of memory.

We estimate the subpixel disparity and standard deviation by fitting the scores near the maximum in the most likely match [5]. Pixels are pruned from the output if the likelihood of the best match is low, the standard deviation is too large, or if the surrounding pixels do not form a large enough coherent block.

5 Experiments

We have tested this technique on many real images of natural terrain. Most of these tests used images that simulated terrain on Mars, that is, sandy and rocky terrain. Some tests used actual images of Mars from the Spirit or Opportunity rovers.

Our first experiment was performed using data collected in the Mars Yard at the Jet Propulsion Laboratory (JPL) using the Rocky 8 rover prototype. A sequence of 11 images captured using the rover mast navigation cameras was collected at roughly 20 cm intervals, so that we could consider wide-baseline stereo pairs with baseline distances ranging from 20 centimeters to 2 meters. The camera orientation remained roughly constant over this image sequence. See Figure 1.

The results of this sequence were interesting. Figure 2 shows the disparity maps that were created by using the first image in the sequence as one image in the wide-baseline pair and each other image as the second image in the pair. The disparities are shown relative to the pixel in the first image. Clearly the density of the stereo coverage is better with a smaller baseline. The reason for this is twofold. First, the region in the first image that is not visible in the second image grows leading to a triangular shaped area in the lower left that cannot be matched. In addition, matching between nearby terrain that was

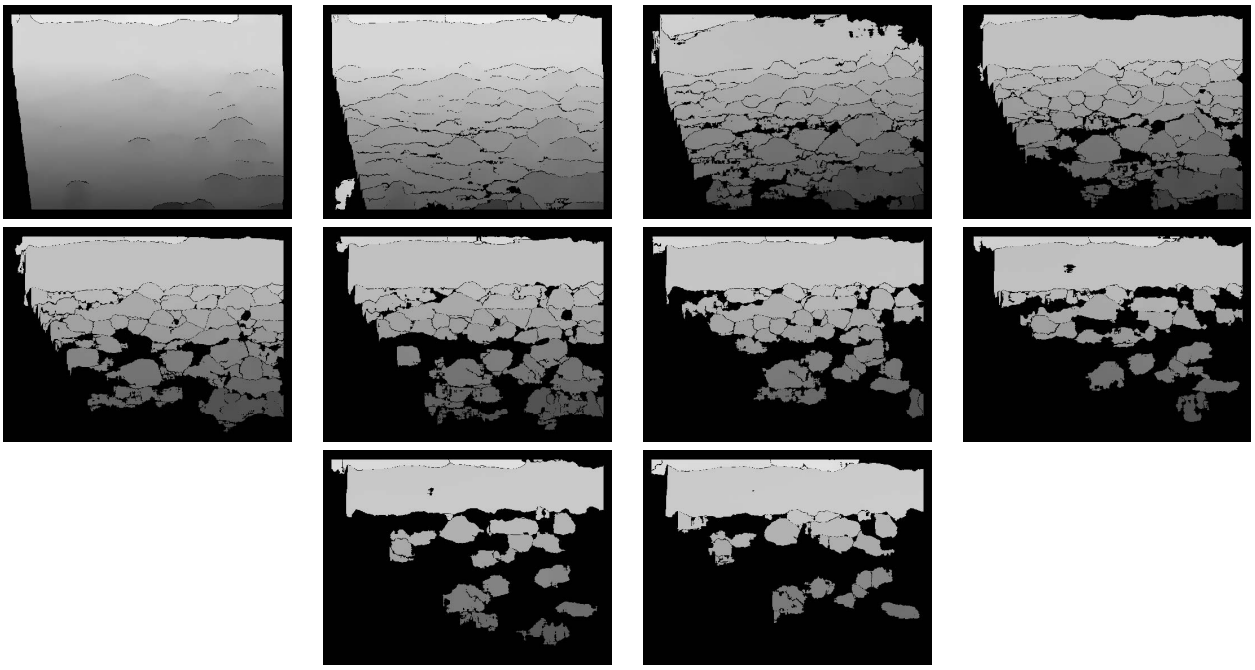

Figure 2: Disparity maps generated from Mars Yard sequence. Each disparity map was created using the wide-baseline stereo algorithm with image 1 as the first image and image n+1 as the second image. The baseline distance ranges from 20 cm for the first disparity map to 2 m for the last disparity map.

present in both images was of a lower quality owing to the difference in appearance between the images. This is expected. In fact, the goal of wide-baseline stereo is to map the more distant terrain. Note, however, that the current pruning techniques are not perfect. Some regions with good data are pruned and some outliers (on the left side) remain unpruned.

We can get an idea as to the quality of the range data for terrain at a greater distance by examining the shape and accuracy of the data for the wall present at the top of each image in the sequence. This wall was approximately 20 meters from the rover. Conventional stereo techniques with a 20 centimeter baseline did not yield highly accurate data of the wall owing to the distance. Similarly, the data using our techniques with a baseline of 20 centimeters produced range data, but without high accuracy. The estimated shape of wall (i.e. planarity) improved considerably as the baseline increased. However, quantitative error was present in the distance to the wall in some cases. The reason for the quantitative error appears to be that a few feature matching errors caused the motion estimation to converge incorrectly. Most of the mistracked features correspond to artificial objects (for example, the poles rising beyond the wall.) We expect this issue to be less of a problem in completely natural terrain.

Figure 3 shows rectified images for both a good case and a bad case for this data. When the motion estimation is correct, the wall is level in the rectified image. The reason for this is that when the camera axes are parallel to each other and perpendicular to the baseline, the correct rectification rotates the images so that the scanlines are parallel to the baseline. With an incorrect motion estimation, an error in the baseline direction causes a deviation from this geometry.

One drawback to this sequence is that none of the terrain is truly distant, so that the use of a wide baseline was not necessarily the best case. The following tests used more distant terrain. Figure 4 shows a wide-baseline stereo pair captured by Rocky 8 in the California desert during field testing. The foreground of these images has been cropped, since wide-baseline stereo was not successful on the nearby terrain. The second row of images shows the matches that were detected and used for motion estimation. The final row shows the left image after rectification and a disparity map corresponding to the rectified left image. In this case the baseline distance was 5 meters and the cameras had a narrow field-of-view producing high resolution in the im-

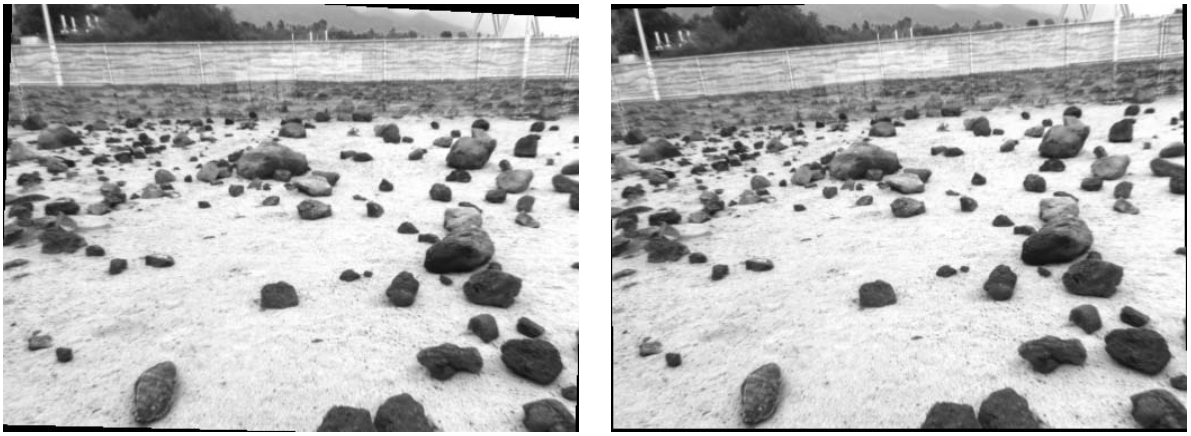

Figure 3: Rectified images from the JPL Mars Yard. The left image shows a good result and the right image shows a lower quality result. Note that the wall in the background is level in the good result.

ages. The range results to the ridge show very high qualitative accuracy, even though the ridge was over 1 kilometer from the cameras. Note that the motion of the rover was not perpendicular to the camera pointing angle and this has resulted in a small rotation of the terrain in the rectified image.

Figures 5 and 6 show two wide-baseline stereo pairs of the Endurance crater on Mars acquired by the Opportunity rover in April 2004. In these cases, good results were obtained on the distant side of the crater, but no results were obtained on the closer terrain that varied significantly between the images. While the primary goal of wide-baseline stereo is to compute range data for the distant points, advanced matching techniques that are not based on template matching may allow us to compute range data for this type of close terrain. It is interesting to note that there are some regions where the sparse matching was successful, but the dense matching either failed or was pruned. This is an area for future improvement.

6 Summary

We have developed a wide-baseline stereo vision technology for Mars rovers. The techniques are able to handle inaccurate motion estimates and images captured from different viewpoints. This allows distant terrain to be mapped for localization and navigation. We have tested these techniques using images of natural, desert terrain from Earth and Mars. The tests indicate that we can achieve high-quality qualitative results. The quantitative accuracy is still under investigation. We have seen that incorrect feature tracking does sometimes adversely affect the motion estimate. The incorrect tracking has occurred primarily for artificial objects, rather than natural terrain. We believe this can be addressed through the use of a random sampling technique, such as RANSAC, to eliminate incorrect matches prior to the iterative estimation step.

Acknowledgments

We gratefully acknowledge funding of this work by the NASA Mars Technology Program. We thank Max Bajracharya, Dan Helmick, and Rich Petras for collecting the Rocky 8 field test data, Dan Clouse for the data collection performed in the JPL Mars Yard, Mark Maimone for help with the MER data, Richard Madison for pointing out some issues during testing, and the MER team for an inspiring mission with publicly available images.

References

[1] W. Förstner. A framework for low-level feature extraction. In *Proceedings of the European Conference on Computer Vision*, pages 383–394, 1994.

[2] D. A. Forsyth and J. Ponce. *Computer Vision: A Modern Approach.* Prentice Hall, 2003.

[3] A. Fusiello, E. Trucco, and A. Verri. A compact algorithm for rectification of stereo pairs. *Machine Vision and Applications*, 12:16–22, 2000.

[4] R. Li, F. Ma, L. H. Matthies, C. F. Olson, and R. E. Arvidson. Localization of Mars rovers using descent and surface-based image data. *Journal of Geophysical Research - Planets*, 2002.

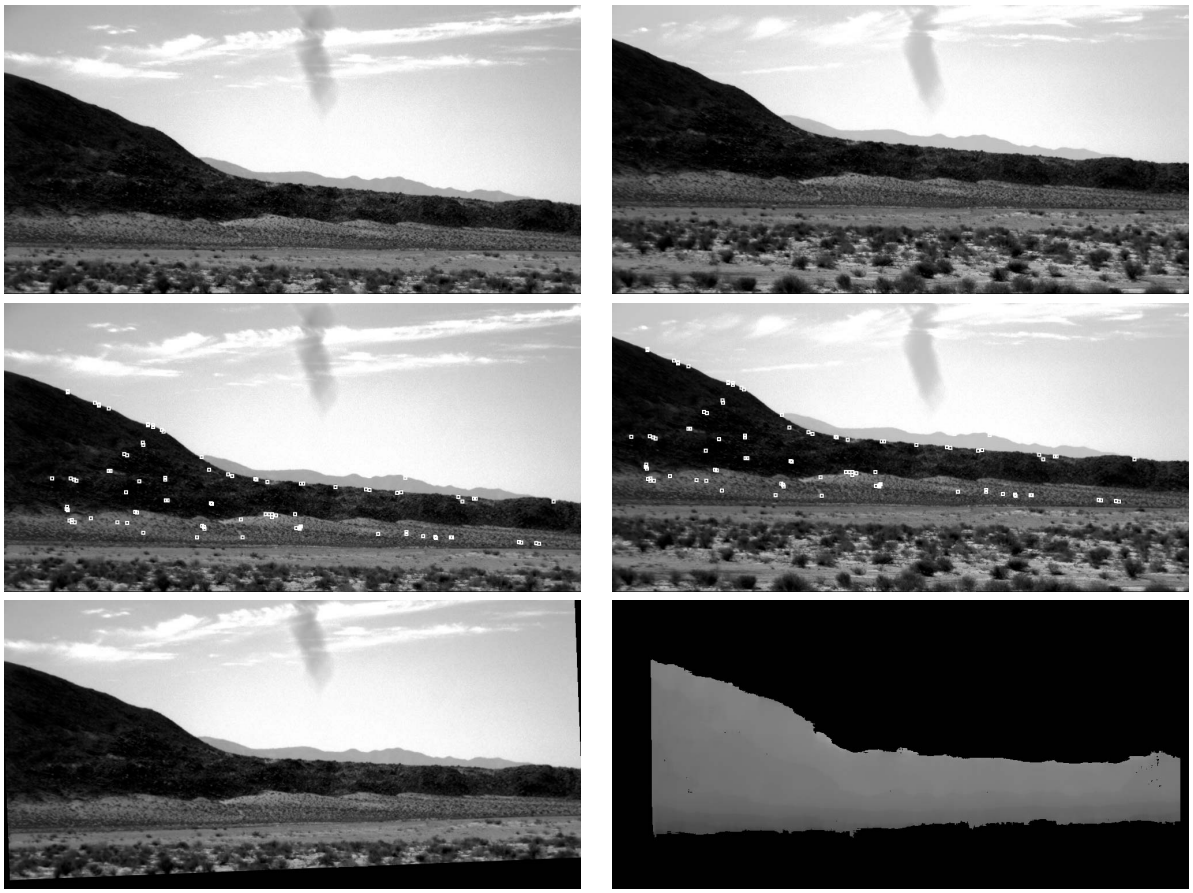

Figure 4: Wide-baseline stereo pair of the California desert captured by Rocky 8 prototype. The second row shows the feature matches that were detected between the images. The third row shows the left rectified image and the disparity map computed from the stereo pair.

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

[6] C. F. Olson. Maximum-likelihood image matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(6):853–857, June 2002.

[7] C. F. Olson, H. Abi-Rached, M. Ye, and J. P. Hendrich. Wide-baseline stereo vision for Mars rovers. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1302–1307, October 2003.

[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 Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 453–458, 2000.

[9] C. F. Olson, L. H. Matthies, Y. Xiong, R. Li, F. Ma, and F. Xu. Multi-resolution mapping using surface, descent, and orbital images. In *Proceedings of the 6th International Symposium on Artificial Intelligence, Robotics and Automation in Space*, 2001.

[10] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Plannery. *Numerical Recipes in C*. Cambridge University Press, 1988.

[11] P. Pritchett and A. Zisserman. Wide baseline stereo matching. In *Proceedings of the International Conference on Computer Vision*, pages 765–760, 1998.

[12] F. Schaffalitzky and A. Zisserman. Viewpoint invariant texture matching and wide baseline stereo. In *Proceedings of the International Conference on Computer Vision*, volume 2, pages 636–643, 2001.

[13] C. Strecha, T. Tuytelaars, and L. Van Gool. Dense matching of multiple wide-baseline views. In *Proceedings of the International Conference on Computer Vision*, volume 2, pages 1194–1201, 2003.

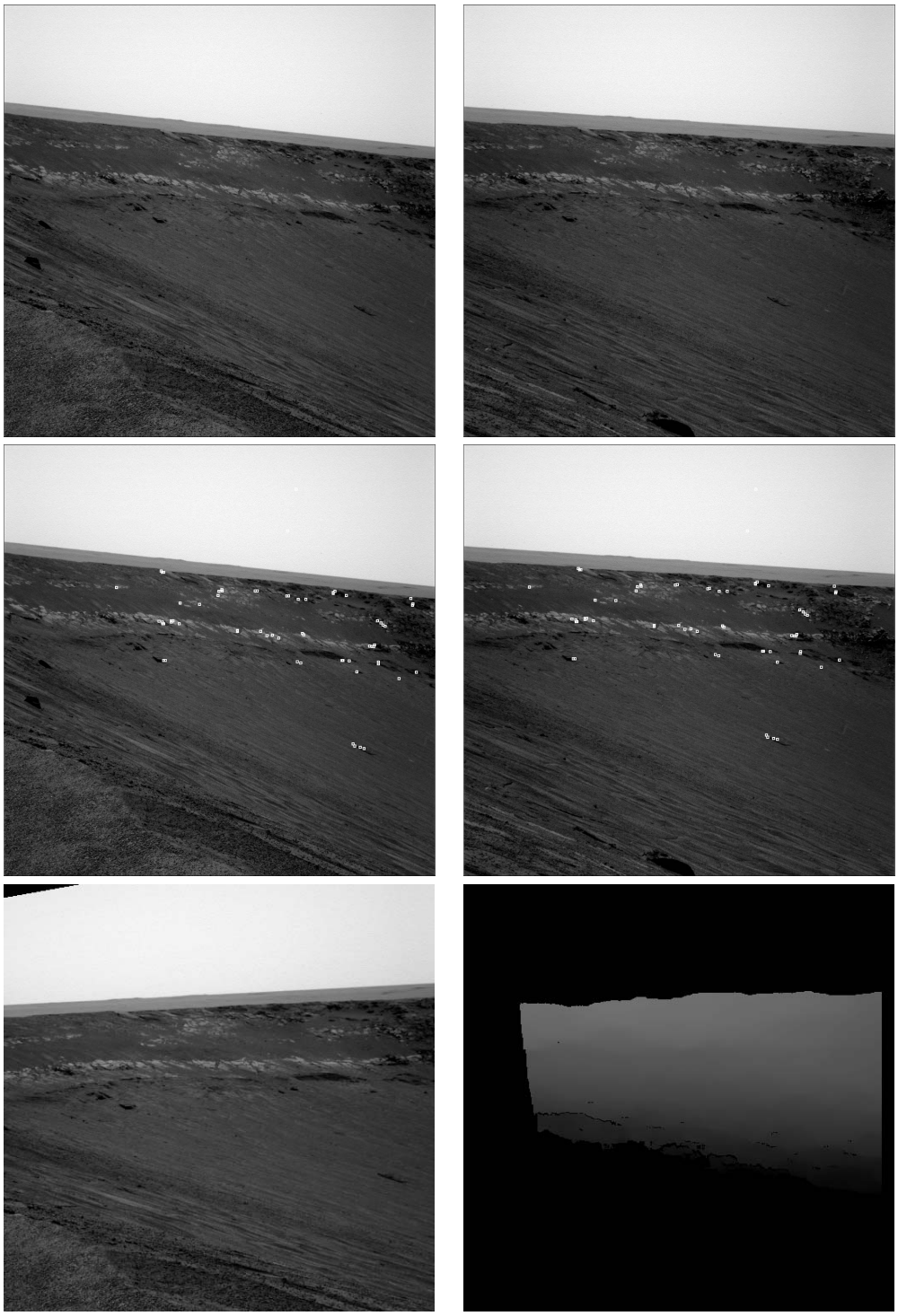

Figure 5: Wide-baseline stereo pair of the Endurance crater on Mars. The images were captured by the Opportunity rover. The middle row shows the features that were matched between the images. The bottom row shows the left rectified image and disparity map computed from the images.

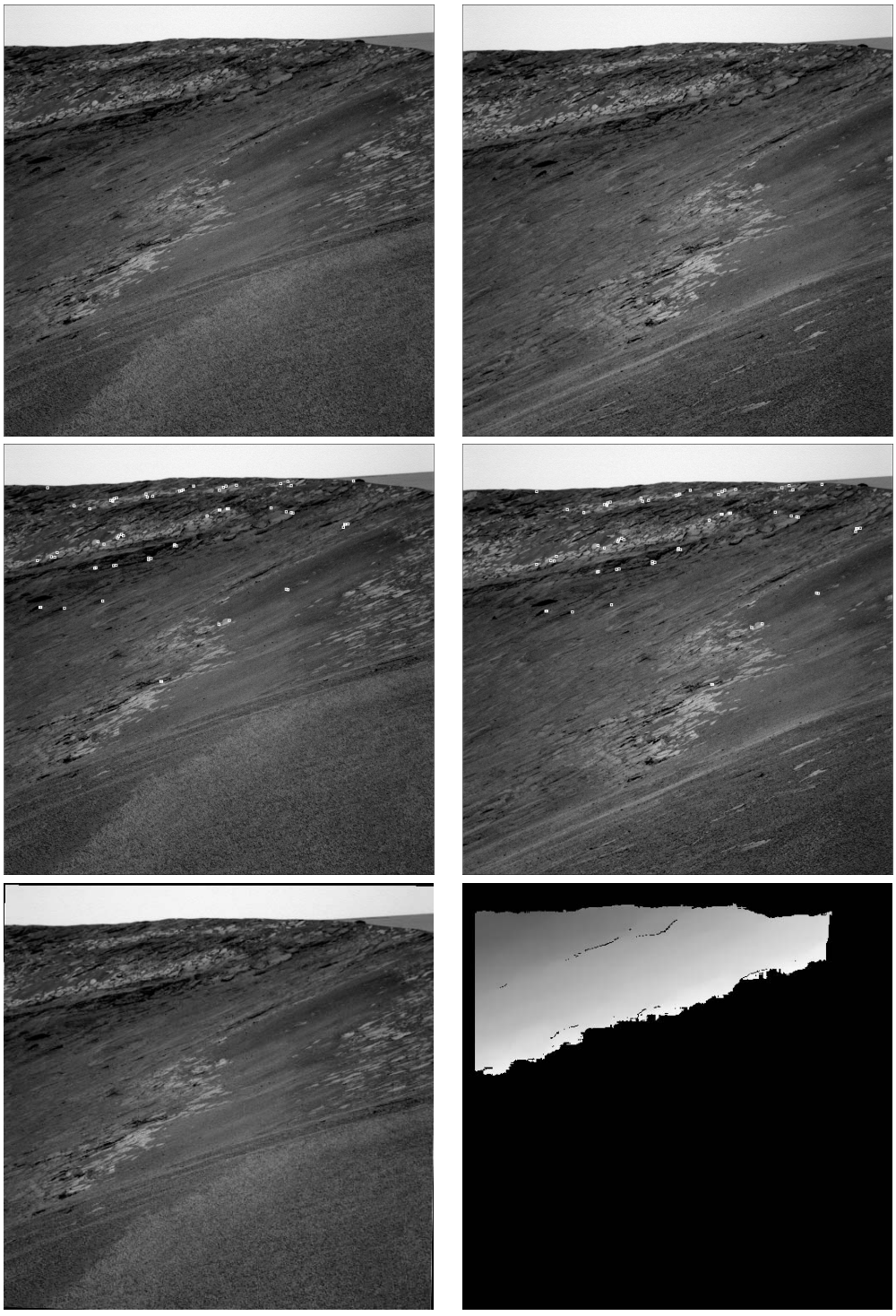

Figure 6: Wide-baseline stereo pair of the Endurance crater on Mars. The images were captured by the Opportunity rover. The middle row shows the features that were matched between the images. The bottom row shows the left rectified image and disparity map computed from the images.