Available online at www.sciencedirect.com
ScienceDirect
Computer Vision and Image Understanding 105 (2007) 73–85
Computer Vision and Image Understanding
www.elsevier.com/locate/cviu

Visual terrain mapping for Mars exploration

Clark F. Olson [a,*], Larry H. Matthies [b], John R. Wright [b], Rongxing Li [c], Kaichang Di [c]

[a] Computing and Software Systems, University of Washington, Bothell, 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
[c] Department of Civil and Environmental Engineering and Geodetic Science, The Ohio State University, 2070 Neil Avenue, 470 Hitchcock Hall, Columbus, OH 43210-1275, USA

Received 6 September 2005; accepted 18 August 2006
Available online 2 October 2006

Abstract

One goal for future Mars missions is for a rover to be able to navigate autonomously to science targets not visible to the rover, but seen in orbital or descent images. This can be accomplished if accurate maps of the terrain are available for the rover to use in planning and localization. We describe techniques to generate such terrain maps using images with a variety of resolutions and scales, including surface images from the lander and rover, descent images captured by the lander as it approaches the planetary surface, and orbital images from current and future Mars orbiters. At the highest resolution, we process surface images captured by rovers and landers using bundle adjustment. At the next lower resolution (and larger scale), we use wide-baseline stereo vision to map terrain distant from a rover with surface images. Mapping the lander descent images using a structure-from-motion algorithm generates data at a hierarchy of resolutions. These provide a link between the high-resolution surface images and the low-resolution orbital images. Orbital images are mapped using similar techniques, although with the added complication that the images may be captured with a variety of sensors. Robust multi-modal matching techniques are applied to these images. The terrain maps are combined using a system for unifying multi-resolution models and integrating three-dimensional terrains. The result is a multi-resolution map that can be used to generate fixed-resolution maps at any desired scale.

Keywords: Terrain mapping; Structure-from-motion; Stereo vision; Mars; Robot navigation; Localization

1. Introduction

For a Mars rover capable of long-range mobility, it is desirable to travel to science targets observed in orbital or descent images, but that are not visible to the rover at its starting position. However, current rovers, including the Mars Exploration Rovers (Spirit and Opportunity), are not able to navigate autonomously to distant targets with a single command. Furthermore, navigation errors can result in the loss of an entire day of scientific activity, since communication with rovers on Mars usually occurs only once per day. Navigation and localization accuracy can be improved using maps of the terrain that encompass the rover's location and the desired destination. Such maps allow improved planning in the route taken by the rover to reach its goal. They are also critical for localization, so that the rover knows when the goal has been reached. We have developed techniques to generate 3D terrain maps for Mars rovers that use all available images, including surface images from landers and rovers, orbital images from current and future Mars orbiters, and descent images from landers. (Descent images are the nested images taken by the lander as it descends to the surface of the planet.) These images provide mapping data at a variety of resolutions from the very high resolution in the surface images to the lower resolution in the orbital images.

For mapping the high-resolution rover and lander images on the surface, we use bundle adjustment techniques to optimize the estimated camera external parameters

* Corresponding author.
E-mail addresses: cfolson@u.washington.edu (C.F. Olson), larry.matthies@jpl.nasa.gov (L.H. Matthies), john.wright@jpl.nasa.gov (J.R. Wright), li.282@osu.edu (R. Li), di.2@osu.edu (K. Di).

1077-3142/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.cviu.2006.08.005

iteratively. These techniques are applied, in particular, to overlapping stereo images from panoramic image sets in order to create accurate maps of the terrain nearby the rover, since slight inaccuracies in the camera positions can lead to seams or ridges in the panoramic map that appear to be obstacles for planning purposes. This method automatically determines corresponding features between multiple pairs of stereo images, even in cases where the overlap is small. These correspondences are used to update the camera positions precisely and produce seamless maps from the stereo data. This technique has been used during ground operations for the Mars Exploration Rover (MER) mission.

Terrain that is distant from the rover cannot be mapped accurately using such techniques. We map this terrain using a combination of rover, lander, and orbital images. Onboard the rover, maps of distant terrain can be created using wide-baseline stereo vision. While conventional stereo vision performed on the rover has limited accuracy for distant terrain owing to the small distance between the stereo cameras (known as the baseline distance), we can achieve accurate mapping for distant terrain using wide-baseline stereo vision. With this technique, images from the same camera, but at different rover positions, are used to generate a virtual pair of stereo images with a large baseline distance. This introduces two problems. First, the relative positioning between the cameras is not well known, unlike conventional stereo vision, where the cameras can be carefully calibrated. In addition, the problem of determining the corresponding locations between the two images is more difficult owing to the different viewpoints at which the images are captured. We combine structure-from-motion and stereo vision techniques in our solution to these problems.

Images captured during the descent of a lander to the surface can be mapped using similar techniques. Each successive image is taken closer to the surface, so that the sequence represents a nested hierarchy of images with shrinking scale and growing resolution. This is useful for integrating the high-resolution surface data with the low-resolution orbital data. However, these images are more difficult to process, since the direction of movement is towards the terrain being imaged, which complicates image rectification. For this problem, we determine the terrain height at each image location by resampling the image multiple times, with each resampling representing a possible terrain height. For each location, the resampled image that yields the best match against the preceding image in the sequence is used to form an initial estimate of the terrain height. The estimates are refined using parabolic interpolation.

At the lowest resolution (and the largest scale), we use pairs of orbital images (or an orbital image and a high-altitude descent image) to compute three-dimensional terrain information. This builds upon the wide-baseline stereo methodology. However, for this case, the images may come from multiple sensors, such as different orbiters or an orbiter and a lander. For this reason, we must use techniques that can find correspondences between images even when the sensors have very different responses to various terrain features. Our approach is to transform the images into a new representation that measures the entropy in the image values around each pixel (treating the pixels in each neighborhood as samples from a random variable). This representation is robust and allows mapping using images from different types of camera.

The terrain maps computed from all of the sources of imagery and at all of the resolutions are compiled using SUMMITT (System for Unifying Multi-resolution Models and Integrating Three-dimensional Terrains), which is designed to merge disparate data sets from multiple missions into a single multi-resolution data set. This system registers the maps and combines them using an octree representation. The multi-resolution map can be used to generate terrain maps with any fixed resolution to use for planning and localization purposes.

A considerable amount of previous work has been done on robotic mapping [1]. Much of it concerns indoor robots, while we are concerned with mapping natural terrain with rovers and spacecraft. We concentrate on the use of cameras for mapping, but other sensors have also been used for mapping Mars, including laser altimeter [2] and delay-Doppler radar [3]. Aside from our own work, much of the early and recent work on terrain mapping for rovers has been performed at Carnegie Mellon University [4–7]. Work at CNRS (the National Center for Scientific Research in France) is also significant, where stereo mapping is performed using an autonomous blimp [8,9].

In Section 2, we discuss the use of wide-baseline stereo for mapping terrain that cannot be mapped effectively using conventional stereo vision. Section 3 describes our approach to mapping surface terrain close to the rover using stereo panoramas. Methods to map images captured during a spacecraft descent to the planetary surface are given in Section 4. Section 5 discusses techniques by which multi-model image pairs (such as from different orbiters or orbital/descent image pairs) can be used to create three-dimensional maps. The SUMMITT system for integrating multi-resolution data sets is described in Section 6. We give our conclusions in Section 7.

2. Wide-baseline stereo vision

Rovers currently map nearby terrain using stereo vision. However, the standard deviation of the error of stereo vision depth estimates increases with the square of the distance to the terrain position. For this reason, conventional stereo cannot accurately map terrain many meters away. One solution to this problem is to use a larger baseline distance (the distance between the cameras), since the error standard deviation is inversely proportional to this distance. This is problematic, however, since a rover with limited size cannot have two cameras with a large baseline distance. We achieve an arbitrarily large baseline distance

using two images captured by the rover at different positions. This technique has been called motion stereo [10,11] and wide-baseline stereo vision [12–16]. It allows terrain up to a few kilometers distant from a rover to be mapped. In general, when a stereo map of distant terrain is desired, the rover will capture an image from a suitable vantage point and then move perpendicular to the vector to the terrain to capture another image, since this provides the best geometry for recovering the terrain map.

While the use of wide-baseline stereo vision improves the accuracy of the range estimation, it introduces two new problems. First, conventional stereo systems use a pair of cameras that are carefully calibrated. In this case, the relative displacement and rotation between the camera positions are determined precisely prior to performing the stereo algorithm. For wide-baseline stereo, this is not possible, since the rover's knowledge of its position when each image is captured is prone to error much greater than a calibrated system would have. Second, the determination of correspondences between the images is more difficult for wide-baseline stereo, since the images have a larger difference in viewpoint than with conventional stereo. Our algorithm addresses these problems using a motion refinement step based on the structure-from-motion problem of computer vision [17] and robust matching between the images [18].

2.1. Motion refinement

Given a pair of wide-baseline images of the same terrain, the first step is to determine the relative positions of the camera at the locations where the images were taken. We assume that an initial estimate of this motion is available from the rover odometry (or other sensors). This estimate is used in two ways. It is used as a starting location for the iterative optimization of the motion. For this use, the estimate does not need to be accurate. The estimate is also used to determine the baseline distance between the camera locations. This parameter cannot be recovered during the optimization, since the same images would result if the problem was scaled to an arbitrary size.

We refine the initial estimate by determining corresponding points between the images and updating the motion to enforce geometrical constraints that must be satisfied for the points to be in correspondence [17]. These correspondences are determined using a simple and robust procedure. Distinctive features are first selected in one image using an operator that locates image pixels where there are gradients in multiple orientations. Candidate matches are detected in the other image at a reduced resolution using the sum-of-absolute-difference (SAD) measure to compare local neighborhoods. Our use of a simple measure here, rather than one that provides affine invariance [12–14,16] is based on a tradeoff between speed and matching errors. Experiments indicate that our system is able to detect a sufficient number of correct matches. Incorrect matches are usually detected and discarded using quality measures.

One or more candidate matches (depending on the candidate scores) determined using the SAD measure at the reduced resolution are carried forward to a candidate refinement step at the highest resolution, where they are again evaluated using the SAD measure. Some candidates are discarded at this stage based on the quality of match. The remaining candidates undergo an affine optimization step before the final match is selected. Even the final match is thrown out if the estimated standard deviation in the match position is too large or if the quality of the second best candidate is close to that of the best candidate.

After finding correspondences between the images, we optimize the translation T and rotation R that represent the motion of the camera between the two positions from which the images were taken:

$$p_2 = Rp_1 + T. \tag{1}$$

For this optimization, we use a state vector that includes the six parameters describing the relative camera positions (only five are recoverable, owing to the scale ambiguity) and the depth estimates of the features for which we have found correspondences. The objective function that we minimize combines the distances (one for each feature match) between the detected feature positions in the search image and the reprojected feature position in that image for the corresponding point according to the current motion estimate and feature depth. If there are n feature matches with coordinates (r_i, c_i) and the reprojected image locations of the corresponding features using the motion estimate are $(\tilde{r}_i, \tilde{c}_i)$, then a simple objective function would be

$$\sum_{i=1}^{n} (r_i - \tilde{r}_i)^2 + (c_i - \tilde{c}_i)^2. \tag{2}$$

We combine the distances in an M-estimator (following the discussion of Forsyth and Ponce [19]) and use the Levenberg–Marquardt optimization technique [20] to adjust the state vector in order to minimize this function.

After we have refined the motion estimate, we apply a rectification process that forces corresponding points between the images to lie on the same row in the images [21]. Fig. 1 shows an example of matching features that were detected in an image pair with a baseline distance of 20 m. Fig. 2 shows the images after motion refinement and rectification has been performed. In this example, the

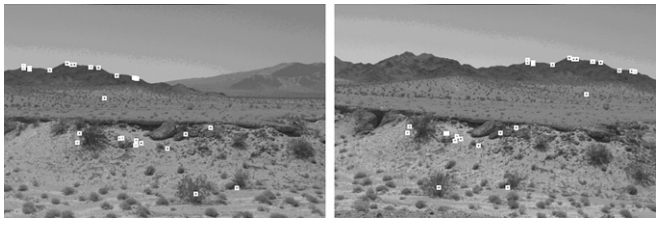

Fig. 1. Matching features detected in a wide-baseline stereo pair.

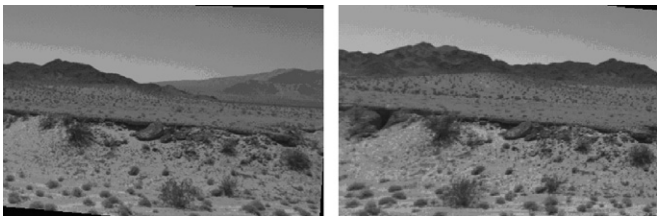

Fig. 2. Wide-baseline stereo pair after rectification. Corresponding features now lie on the same image row.

camera pointing angles converged by 20° from parallel so that the same rock was located at the center of both images.

It can be observed that there is relatively little movement between the features matched in the middle of the image. Features on the mountains move considerably to the right and foreground features move in the opposite direction. After rectification, all of the corresponding features lie on the same image row, facilitating dense matching.

2.2. Disparity estimation

Disparity is a measure of the difference in the image position for corresponding points between two images. After rectification, this difference should be exclusively in the x-coordinate. Given a pair of rectified images, we can compute the disparity for any point that is present in both images by searching along the corresponding row of the other image. Every position in one image is given a disparity estimate, unless no corresponding match can be found in the other image. We accomplish this by combining robust template matching [18] with an efficient stereo search algorithm [22,23]. This provides robustness to changes in appearance owing to different viewpoints while maintaining efficiency.

In order to determine which points are in correspondence between the two images, we could use a measure such as normalized correlation or the sum-of-squared differences (SSD) applied to a small image window (or neighborhood) around each point. However, owing to the difference in viewpoints, these measures do not produce good results for wide-baseline stereo vision [17]. Instead, we use a maximum-likelihood measurement that improves upon normalized correlation (and SSD) in two ways. First, normalized correlation compares only the pixels between the two images that are directly overlapping at some disparity of the image window with respect to the other image. If camera motion or perspective distortion causes pixels to move by different amounts between the two images, it will not be possible to find a window position where all of the pixels are correctly overlapped. Our distance measure allows pixels that are not directly overlapping to be matched by linearly combining the distance in the image with the difference in intensity. This is important, since we do not model perspective (or even affine) distortion in determining disparities. Computing the best distance for a pixel is no longer trivial with the formulation, since the best match may not be the overlapping pixel from the other image. However, the distances can be computed efficiently by precomputing a three-dimensional distance transform of the input data [18].

The second improvement over normalized correlation is that the possibility of an outlier is explicitly represented. In this application, any terrain feature that is visible in one image, but not in the other is an outlier for the matching process. Such outliers occur frequently for images taken from different viewpoints. In order to model such outliers, we use a probability density function (PDF) for each pixel distance that is a mixture of two terms, one for inliers and one for outliers, where each is weighted by an estimate of the probability of an outlier. Let $p_{\mathrm{i}}(D)$ be the PDF for inliers (typically modeled as a Gaussian function), $p_{\mathrm{o}}(D)$ be the PDF for outliers (a constant function works well in practice), and α be the expected fraction of inliers. The overall PDF is simply a linear combination:

$$p(D) = \alpha p_{\mathrm{i}}(D) + (1-\alpha)p_{\mathrm{o}}(D). \tag{3}$$

The disparities computed using this PDF are not sensitive to the expected fraction of inliers α, as long as it is not close to one. Our implementation uses $\alpha = 0.75$.

The overall likelihood for a neighborhood of pixels is the product of the probability density functions for each of the pixels in the neighborhood. In practice, it is efficient to sum the logarithms of the probability density functions, which can be precomputed.

$$\log L(D_1(d), \ldots, D_n(d)|d) = \sum_{i=1}^{n} \log p(D_i(d)), \tag{4}$$

where d is the disparity under consideration, n is the number of pixels in the image window, $D_i(d)$ is the distance for the ith pixel in the image window at this disparity, and $p(\cdot)$ is the probability density function for the distances. This yields an M-estimator for robust estimation of the disparity [24]. The use of a probability density function that models outliers prevents individual pixels that match very poorly from having an undue effect on the overall likelihood for a candidate match.

We use an efficient strategy that is common in stereo vision to perform dense matching between the rectified images using the measure described above. When two adjacent neighborhoods in the first image are compared to the second image at the same disparity, the computations performed overlap almost completely. Dynamic programming can be used to eliminate these redundant computations and perform dense matching efficiently [22,23]. Fig. 3 shows a result of performing dense matching using the images from Figs. 1 and 2. The disparities shown correspond to pixels at the same location in the left image of Fig. 2. Black locations indicate that no good match could be found. High quality results are achieved on the left side of the image, since the features in this area are also present in the right

Fig. 3. Disparity map computed from a wide-baseline stereo pair. The largest (positive) disparities are white. The smallest (negative) disparities are dark. Black values indicate that the disparity was discarded.

image. Towards the right side and bottom of the image, the results degrade, since there is no matching feature (or the match is difficult to find).

3. Mapping surface panoramas

We map surface panoramas using an automatic procedure that first selects candidate tie points in stereo images. Matches are detected both within stereo pairs and between adjacent stereo pairs to create an image network. Bundle adjustment is applied to the image network in order to improve the estimates of the camera and landmark positions. Finally, elevation maps and orthophotos are generated using dense matches detected between image pairs.

3.1. Automatic selection of tie points

We have developed a systematic method for automatic tie point selection [25–27]. For selecting tie points within one stereo pair (intra-stereo tie points), the first step in the procedure is interest point extraction using the Förstner operator [28]. The interest points are initially matched using normalized cross-correlation coefficients. However, within each stereo pair, mismatches may occur. The matches are verified based on the mutual consistency of the parallax between matches. The parallaxes are plotted versus the row of the interest point. Assuming that the terrain does not change sharply, the parallaxes should form a consistent curve in the plot. A median filter is used to identify and eliminate the mismatches. Next, an even distribution of the tie points is selected using a gridding method. Within each evenly spaced image patch, the single tie point with the largest variance in the image intensity is selected. Fig. 4 shows an example of automatically selected intra-stereo tie points from IMP (Imager for Mars Pathfinder) images.

Tie points between adjacent stereo images (inter-stereo tie points) are more difficult to find, since the images often have little overlap. In this case, a coarse elevation map is constructed using the intra-stereo data. The coarse map is used to predict both the overlapping areas in the inter-stereo images and approximate locations of the matches for selected interest points. We then search over a small search space in order to find the final match. This has resulted in approximately 90% success in test images. Verification is performed by examining the consistency of the parallaxes in this case also. Fig. 5 shows an example of automatically selected inter-stereo tie points from IMP images. Fig. 6 shows an example of tie points automatically selected from rover images. Ultimately, the selected intra- and inter-stereo tie points build an image network.

3.2. Bundle adjustment

A bundle adjustment [29] is applied to the image network to improve the accuracy of image orientation parameters as well as the 3D ground positions of the tie points. To achieve high accuracy, we model the correlation between the position and attitude of the stereo camera and use this correlation as constraints in the least-squares adjustment. A subset of the tie points is selected for the overall bundle adjustment with nine evenly distributed intra-stereo tie points per image and six evenly distributed inter-stereo tie points per image pair. These tie points are selected such that more weight is given to those points that appear in the most images.

In a Mars mission, it is unlikely that there will be a large number of control points that can be used to register surface images with orbital images. We, thus, use a free network with a rank deficient normal matrix for the bundle adjustment. No unique solution exists for such a network.

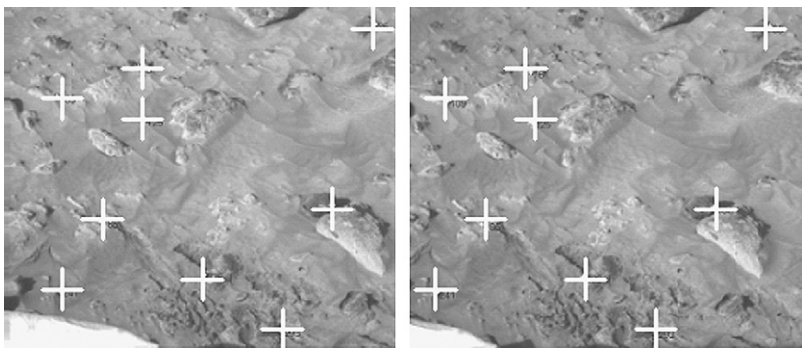

Fig. 4. Intra-stereo tie points in lander IMP (Imager for Mars Pathfinder) images.

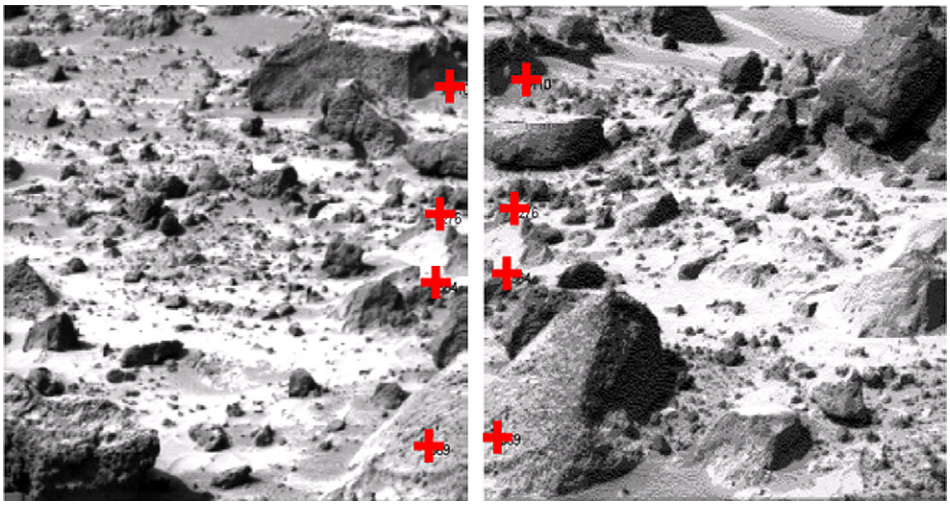

Fig. 5. Inter-stereo tie points in lander IMP (Imager for Mars Pathfinder) images.

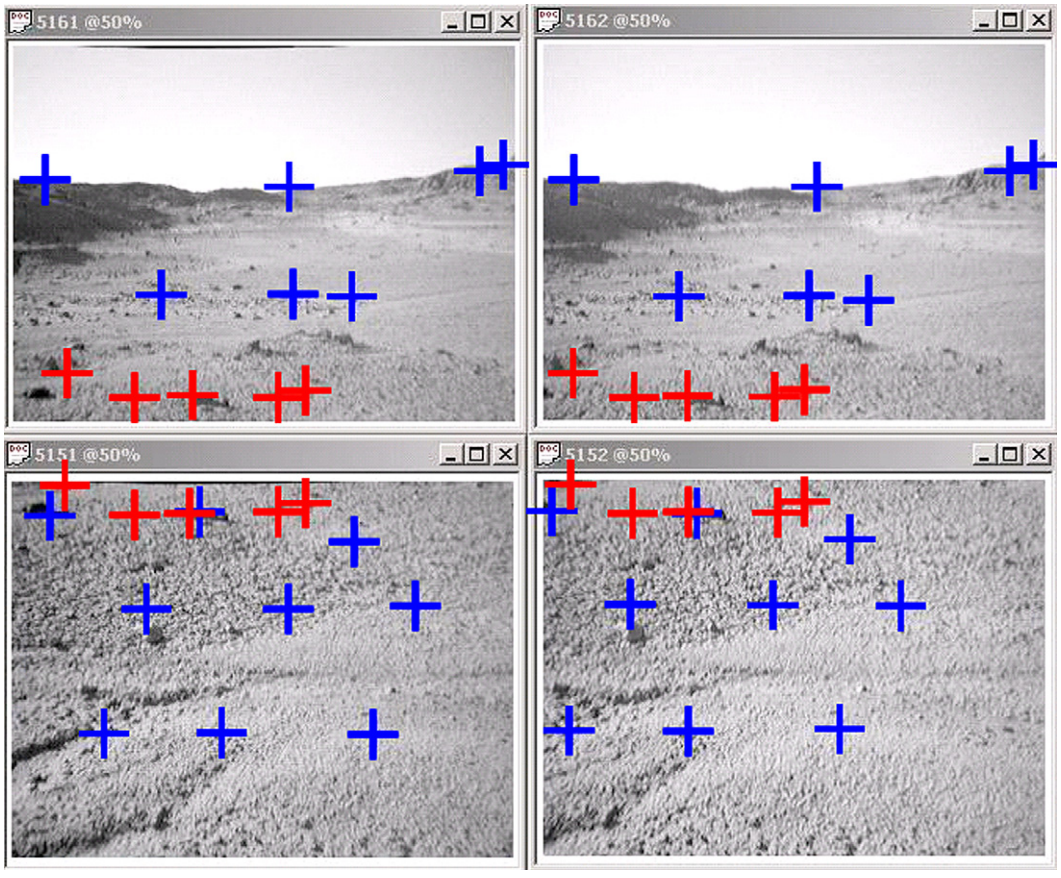

Fig. 6. Automatically selected tie points from rover images. (Red crosses are intra-stereo tie points and blue crosses are inter-stereo tie points.) (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this paper.)

We solve the normal equation using singular value decomposition to satisfy both the least-squares criterion and the minimum norm principle.

In an experiment on IMP data, the panorama consisted of 129 images that form either an upper panorama and a lower panorama with horizontal links, or a complete panorama with both horizontal and vertical links. In the image network, there were 655 tie points, 633 of which were automatically selected and 22 that were manually selected. The manually selected points were necessary to strengthen the image network, since some adjacent images have very little overlap. In practice, we expect to use an operational scenario that would produce a larger overlap between the adjacent images for use in mapping and prevent the need

for any manually selected points. Before adjustment, the precision was 4.61 pixels in the image space (distance between measured and reprojected image points) and 0.067 m in the object space (distance between 3D positions triangulated from adjacent stereo pairs). After bundle adjustment, the precision was 0.80 pixels in the image space and 0.05 m in the object space.

In an experiment processing terrestrial rover data, a panoramic image network was built by linking 36 mast images (18 pairs) with 220 automatically selected intra- and inter-stereo tie points. Before adjustment, the precision was 2.61 pixels in the image space and 1.62 m in the object space. After bundle adjustment, the precision was 0.55 pixels in the image space and 0.08 m in the object space. Clearly, the precision is considerably improved in both the image space and the object space by the bundle adjustment.

3.3. Elevation map and orthophoto generation

After bundle adjustment, image matching is performed to find dense correspondence points for elevation map generation. The epipolar geometry is used with a coarse-to-fine matching strategy to achieve both high speed and high reliability. The 3D ground coordinates of the matched points are then calculated by photogrammetric triangulation using the adjusted image orientation parameters. Finally, based on the 3D points, the elevation is generated using the Kriging interpolation method [30].

The orthophoto is generated by projecting the grid points of the refined elevation map onto the left (or right) image. A corresponding grayscale value is found and assigned to the grid point. In the area of overlap for adjacent images, the grid points will be projected to two or more images. The grayscale value is picked from the image

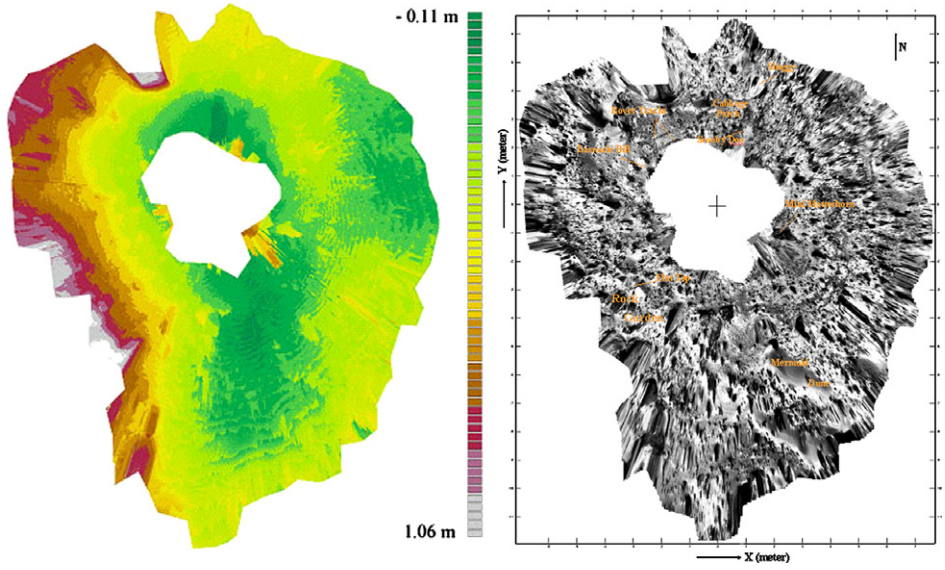

Fig. 7. Elevation map and orthophoto of the Mars Pathfinder landing site.

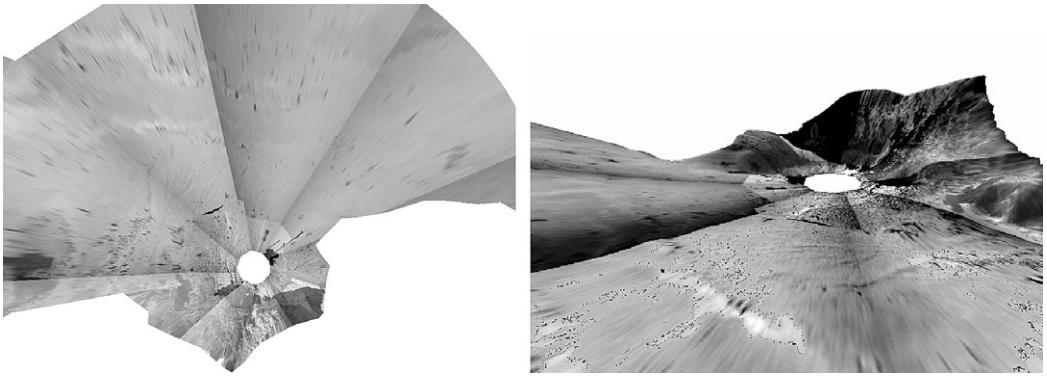

Fig. 8. 3D visualizations of the terrestrial rover test site.

in which the projected point is closest to its image center. The resultant elevation map and orthophoto of the Mars Pathfinder landing site are shown in Fig. 7. The 3D visualizations of the terrestrial rover data are shown in Fig. 8. Through visual checking of the orthophotos, we find no seams between image patches.

4. Mapping descent images

Images taken by a rover and lander on the surface give us high-resolution data for mapping, particularly near the landing site. We can also perform mapping using orbital images at a much lower resolution. One way of linking the data at very different resolutions is through the nested images that a lander captures as it descends to the planetary surface. These descent images yield data at a hierarchy of resolutions that can help pinpoint the landing site and integrate the high-resolution surface data into an encompassing multi-resolution map. Fig. 9 shows an example of descent imagery of the moon from the Ranger 9 mission. Fig. 10 shows a descent image sequence captured using a helicopter in the Mojave Desert for rover testing.

To perform matching between pairs of consecutive descent images, we can use techniques similar to those described for wide-baseline stereo, including motion refinement and dense disparity estimation [31]. For this case, the use of robust matching techniques for disparity estimation is less important, since the terrain is viewed from largely the same direction, but at different altitudes over the image sequence. A different problem arises for this case. Since the images are captured in the direction in which the spacecraft is traveling (down towards the surface), the epipoles are within the image boundaries and this has two consequences. First, we cannot rectify the images along epipolar lines, since they pass through the epipoles, and this would resample the image unevenly. In addition, the depth cannot be accurately recovered near the epipoles, since the computation is numerically unstable. Another issue that must be considered is that translations of the spacecraft relative to the surface tend to cause similar images to those caused by rotations of the spacecraft. We rely on the onboard inertial navigation system to provide accurate orientation data.

For the above reasons, our approach to recovering an elevation map from these images is somewhat different

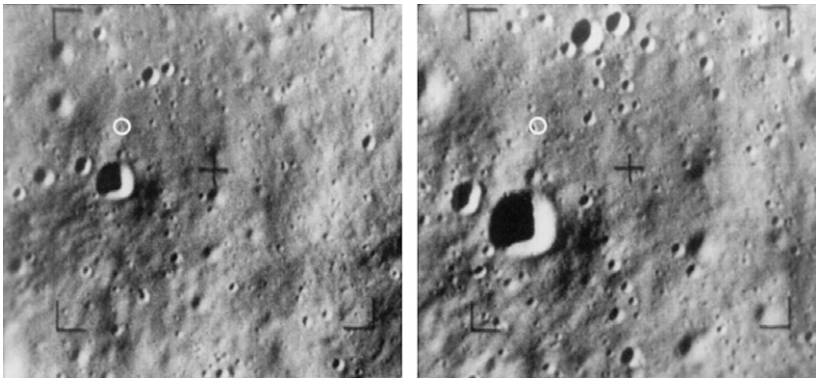

Fig. 9. Images of the moon captured by the Ranger 9 spacecraft during descent to the surface.

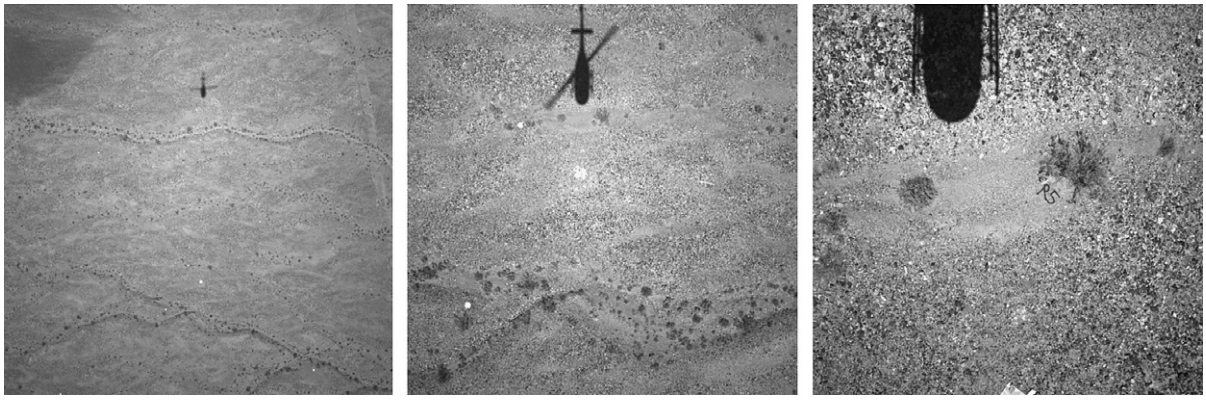

Fig. 10. Three images from a descent sequence of the Mojave Desert captured with a helicopter. The full sequence consists of eight images, including two at altitudes in between those shown above.

from the wide-baseline stereo vision methodology. As in the previous algorithm, we refine the motion by matching features between the images and performing an optimization that minimizes the difference between the matched feature positions and the predicted positions from the motion estimate. However, instead of performing rectification, we examine a discrete set of virtual planar surfaces perpendicular to the camera pointing axis that slice through the three-dimensional terrain. See Fig. 11.

For each of the elevations examined (each elevation corresponds to a virtual plane through the terrain) one image is warped to appear as it would from the second camera position if all of the terrain was at this elevation. The warped image is compared to the second image by computing the sum-of-squared-differences (SSD) in a neighborhood around each pixel. The scores in each neighborhood are weighted by a Gaussian modulation function, so that the pixels closest to the center of the neighborhood are given higher weight than the edges of the neighborhood. At each location, the elevation that produces the lowest of the modulated SSD scores is stored as the initial elevation estimate for that location:

$$S_k(x,y) = \sum_{i=-\frac{W}{2}}^{\frac{W}{2}} \sum_{j=-\frac{W}{2}}^{\frac{W}{2}} e^{\frac{-i^2-j^2}{2\sigma^2}} (I_1(x+i,y+j) - I_2^k(x+i,y+j))^2, \tag{5}$$

where W is the size of the image window examined, I_1 is the first image, and I_2^k is the second image warped according to the recovered motion parameters and the hypothesized depth z_k. Details of the warping can be found in [31].

We improve the accuracy of these estimates using parabolic fitting of the SSD scores achieved near the optimum. If $z_k(x,y)$ is the initial estimate for position (x,y) in the image, δ_z is the spacing between the planes slicing the terrain, $S_k(x,y)$ is the SSD score for this plane and $S_{k+1}(x,y)$ and $S_{k-1}(x,y)$ are the SSD scores for neighboring planes, then the fitted depth estimate is

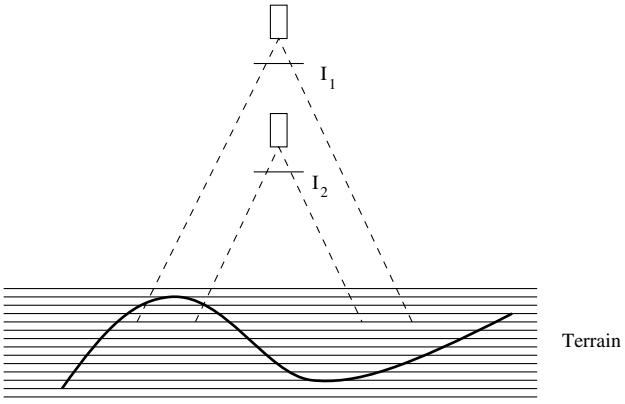

Fig. 11. The terrain is sliced with virtual planes in order to estimate the elevation at each location.

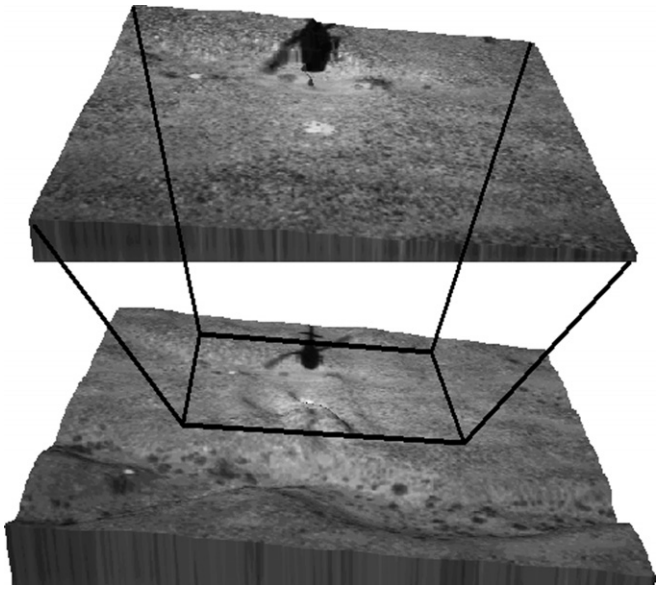

Fig. 12. Partial hierarchy of maps extracted from Mojave Desert images.

$$z(x,y) = z_k(x,y) + \frac{\delta_z(S_{k+1}(x,y) - S_{k-1}(x,y))}{2S_{k+1}(x,y) + 2S_{k-1}(x,y) - 4S_k(x,y)}. \tag{6}$$

These techniques have been tested on synthetic images, real images collected with a helicopter and Ranger images such as those in Fig. 9. Unfortunately, accurate camera models for the Ranger images are not available, so the maps constructed from those images are not quantitatively accurate. For the helicopter data, the initial motion estimates were computed using control points on the ground rather than using an inertial navigation system, as we would expect on a Mars lander. In addition, the camera positions do not follow a smooth path towards the surface, owing to the difficulty in maintaining the helicopter position as the altitude changes.

Fig. 12 shows two levels of the hierarchy of maps created from the descent images in Fig. 10. The maps were generated by rendering images according to the elevations computed by the algorithm. A channel can clearly be seen running through the lower image. Note that the location of the epipole can also be seen in the center of the lower image where the elevation estimates become somewhat unstable. Such areas can be refined using the map at the next higher resolution (lower altitude). In upper image, the epipole is not as noticeable, but it is present at the shadow of the helicopter. The epipoles are not precisely at the center of the images, since the motion between images was not directly downward and some rotation of the camera occurred between images.

5. Multi-modal data

After considering surface and descent images, one might wonder whether similar techniques can be used for orbital

images. In general, the answer is yes. We can even generate depth maps by performing matching between a descent image and an orbital image. If the same terrain location is imaged by the same sensor from a different location, then mapping can be performed using one of the previously described techniques. (If it is not imaged from a different location, we cannot extract any depth information.) On the other hand, if we want to perform matching between images captured from different sensors, then the problem becomes more complex, since the terrain will not necessarily have the same appearance in images from different sensors.

When we perform matching between images captured with different sensors, we must use a measure for matching image neighborhoods (essentially feature tracking) between images that is insensitive to the difference between the sensors, since different sensors have different responses to the various wavelengths of light. Our approach is to transform the images into a new representation that is less sensitive to sensor characteristics. The new representation replaces each pixel in the image with a local measure of the entropy in the image near the pixel [32]. This is based on the idea that each neighborhood in the image can be treated as a set of samples from a random variable. The entropy of the random variable is estimated from the sample values.

We estimate the probability density by, first, histogramming the values observed in the local neighborhood. The size of the neighborhood is selected empirically at present. A scheme to select the neighborhood size adaptively is a possible improvement for future work. This histogram is smoothed using a Gaussian function and the resulting distribution is used in the entropy calculation. Fig. 13 shows an example image and the same image after each pixel has been replaced with the local entropy computed using this method.

This entropy representation is insensitive to sensor characteristics. In fact, it has been used to perform matching between infrared images and conventional images that

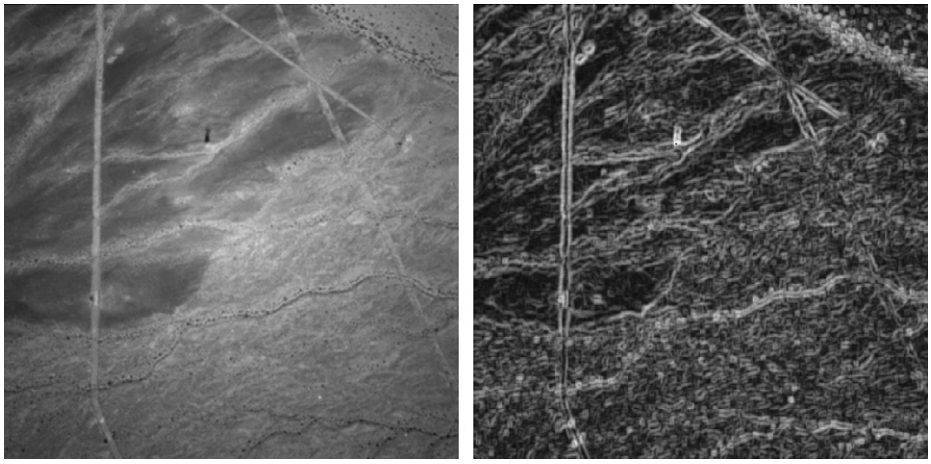

Fig. 13. An aerial image and its representation using a local entropy measure.

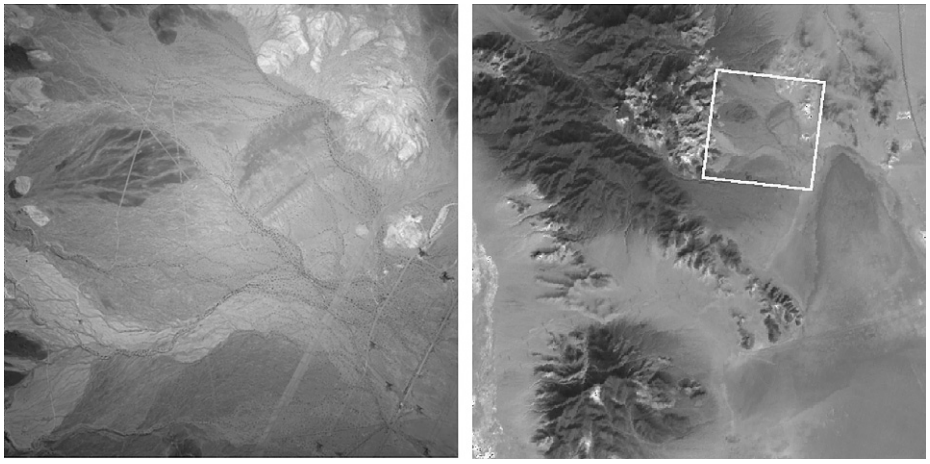

Fig. 14. Registration between an aerial and an orbital image. Left, aerial image of Mojave Desert. Right, registered location in an orbital image.

have a very different appearance [32]. Once the images have been transformed into the entropy representation, the images can be compared in a straightforward manner using normalized correlation.

This technique can also be used to determine the location of a descent image within an orbital image in order to pinpoint the landing site. For this problem, the search space is large, since the spacecraft position has six degrees-of-freedom. The terrain is usually roughly planar, so the transformation between the images can be modeled by an affine transformation.

Since we do not necessarily have a good enough initial estimate of the position to converge to the true position using iterative optimization, we search the pose space using different techniques. Translation in the image (such as would be caused by lateral motion of the spacecraft) can be searched efficiently using the Fast Fourier Transform (FFT), since cross-correlation corresponds to point-wise multiplication in the frequency domain. However, this technique cannot be used for the remaining parameters. Our approach is to sample these parameters coarsely. All translations are considered for each sample point using the FFT-based cross-correlation algorithm. When candidate matches are found, we refine them using iterative optimization. This combination of techniques has led to good speeds for multi-modal matching with a large search space [32]. Fig. 14 shows an example where an aerial helicopter image was registered with an orbital image encompassing the same location. Since the images were captured with different sensors, matching using straightforward techniques does not produce the correct result for this case.

6. Integrating data sets

We use SUMMITT for integrating maps from various sensors. SUMMITT is the System for Unifying Multi-resolution Models and Integrating Three-dimensional Terrains. It is a suite of software tools developed at the Jet Propulsion Laboratory for registering and merging terrain models from a variety of sources. The terrain models can then be output in a variety of formats to support activities requiring visualization of the terrain.

The fundamental data structure for the terrain data in SUMMITT is the octree. Octrees were chosen due to their inherent support of multi-resolution data, their ability to support rapid searches for nearest neighbors, and their small memory footprint. Multi-resolution support is necessary due to the disparate sources of terrain information that must be merged. Orbital cameras may be used to collect imagery and produce terrain models through stereo processing. In this case, the resolution of the terrain models could range from one meter up to multiple kilometers based on the quality and resolution of the optics, the orbital dynamics, and atmospheric constraints. Typically, the resolution of terrain models produced from orbital cameras will be relatively constant as the range to the terrain is generally fixed by roughly circular orbits. Descent cameras, on the other hand, are constantly moving during data collection and their range to the terrain is varied. Thus, a camera in the early stages of data collection, at a higher altitude, might be capable of producing models with a resolution near one meter. The same camera might produce models at one-centimeter resolution at a very low altitude. Finally, rover and lander cameras may produce models with sub-centimeter resolution. For example, on the Mars Exploration Rovers (MER) mission, the rovers carry three sets of stereo cameras ranging from nearly 180° FOV down to about 15° FOV. Given that the stereo baseline for each pair of cameras is comparable, the resolution of the terrain models produced will vary considerably among the cameras on a given platform. In addition, the models produced by a single camera will vary considerably in resolution due to the wide range of distances of the terrain from the camera. Thus, near-field objects may have up to 30 times finer resolution than far-field objects.

The terrain models are inserted into the octree by assuming that each sample is a volume with a cross-section equivalent to the model resolution. The coarser model data remain near the root of the tree while the finer data traverse nearer the leaves of the octree. Each individual model, except the first, must be registered to the overall terrain model contained within the octree, prior to being merged into the octree. The registration process uses the Iterative Closest Points (ICP) algorithm, as described by Zhang

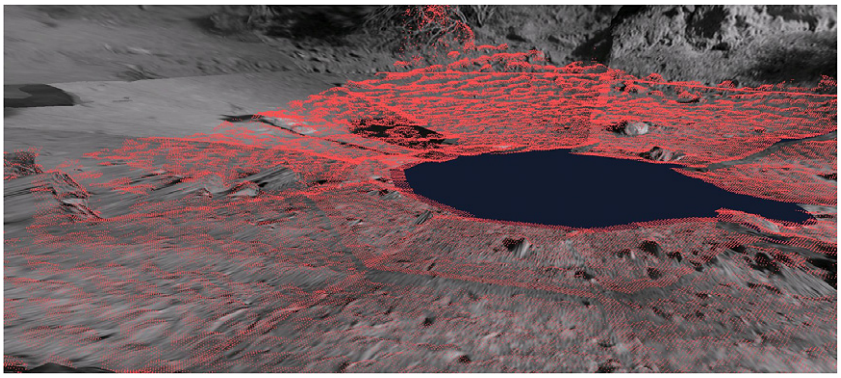

Fig. 15. Elevation map generated from registered rover panorama data.

[33] and applied by Nishino and Katsushi [34] and Miyazaki et al. [35], to compute an alignment transform. This transform is then applied to each sample in the model prior to inserting it into the octree. Since the ICP algorithm is an iterative process that requires finding nearest neighbors, the octree's support for rapid searches is very important to the success of the ICP algorithm.

Once models have been merged into the octree, terrain models can be generated. For many applications, multi-resolution triangle meshes are used. By treating the octree as a cloud of points, a variety of algorithms are available for this process, including those of Hoppe et al. [36] and Hilton et al. [37]. For rover missions, corresponding height maps are also required. The height maps are produced by binning the octree data in the (x, y) plane and then selecting the largest z value within each bin. The bins can be of any desired resolution to support the visualization requirements. Repeated extraction of height maps at different resolutions produces multi-resolution maps for applications requiring such support.

Fig. 15 shows an example of an elevation map generated using SUMMITT. In this case, panoramic stereo images were registered, integrated, and, finally, rendered with the original image data.

7. Discussion

In developing these techniques, some of the lessons learned are unsurprising in retrospect. The use of robust estimation procedures that do not break down in the presence of outliers has certainly been important. We also found that locating feature matches that broadly cover the area to be mapped is important. When the coverage was not as good, the estimated motion between the images tended to best fit areas in which features matches had been detected. The dense mapping was, therefore, better in these areas. In practice, this means that areas where the feature matching is difficult, such as nearby terrain that undergoes considerable perspective distortion in wide-baseline stereo, are not likely to be mapped well by this technique. However, it should be noted that we use wide-baseline stereo primarily to map distant terrain where this is not an issue, since conventional stereo can map the nearby terrain accurately. One possible improvement to our methodology would be to spend more computation in detecting more feature matches in difficult areas prior to estimating the camera motion.

The failure modes of these techniques are largely common ones for motion and depth estimation techniques. All of the techniques rely on images that capture the same terrain. If the terrain shown in the images does not overlap, then the techniques cannot succeed. Similarly, we require the images to have sufficiently distinctive features that they can be matched between images. Featureless images, or ones between which the appearance of the features changes drastically, prevent us from performing feature matching. This, in turn, disallows computation of the relative positions of the camera at the times when the images were captured and causes a failure. In both of these cases, where few (if any) matches are computed, the failure can be detected. A different type of failure can occur if some of the matches between the images are incorrect. While we use techniques designed to eliminate such errors in the matching process and to ameliorate the effects of them in the motion estimation process, quantitative error in the terrain mapping can occur when these errors are present. Our strategy has been to reduce the frequency of this type of error, since detecting when they have occurred is difficult.

8. Conclusions

Obtaining accurate maps of the terrain is critical for a rover to traverse long-distances on Mars. Without such maps, a rover may spend much time and energy venturing along what turns out to be a dead end. Maps are also critical for rover localization in a long-distance traverse, since estimates from odometry and other sensors will grow in error without bound unless corrected using a global map. We have described techniques for generating three-dimensional terrain maps spanning many resolutions, from the high-resolution maps generated using conventional stereo onboard a rover, through medium resolution maps generated using wide-baseline stereo or descent images, to the lower resolution (but wider area) maps generated from high-altitude descent images and orbital images. The data sets are combined using a system (SUMMITT) that efficiently registers and integrates the data sets in a multi-resolution context. This system also provides tools for map visualization that are useful for planning.

Acknowledgments

We gratefully acknowledge funding of this work by the NASA Mars Technology Program. This is an expanded version of a manuscript that appeared in the Proceedings of the IEEE Aerospace Conference [38].

References

[1] S. Thrun, Robotic mapping: a survey, in: G. Lakemeyer, B. Nebel (Eds.), Exploring Artificial Intelligence in the New Millennium, Morgan Kaufmann, Los Altos, CA, 2003, pp. 1–35.
[2] M.T. Zuber, Observations of the north polar region of Mars from the Mars orbiter laser altimeter, Science 283 (5396) (1998) 2053–2060.
[3] J.K. Harmon, R.E. Arvidson, E.A. Guinness, B.A. Campbell, M.A. Slade, Mars mapping with delay-Doppler radar, J. Geophys. Res. 104 (E6) (1999) 14065–14089.
[4] M. Hebert, C. Caillas, E. Krotkov, I.S. Kweon, T. Kanade, Terrain mapping for a roving planetary explorer, in: Proceedings of the IEEE Conference on Robotics and Automation, vol. 2, 1989, pp. 997–1002.
[5] I.S. Kweon, T. Kanade, High-resolution terrain map from multiple sensor data, IEEE Transactions on Pattern Analysis and Machine Intelligence 14 (2) (1992) 278–292.
[6] E. Krotkov, Terrain mapping for a walking planetary rover, IEEE Transactions on Robotics and Automation 10 (6) (1994) 728–739.

[7] D. Huber, M. Hebert, Fully automatic registration of multiple 3-d data sets, Image Vis. Comput. 21 (7) (2003) 637–650.
[8] S. Lacroix, I.-K. Jung, A. Mallet, Digital elevation map building from low altitude stereo imagery, Robot. Auton. Syst. 41 (2–3) (2002) 119–127.
[9] I.-K. Jung, S. Lacroix, High resolution terrain mapping using low altitude stereo imagery, in: Proceedings of the International Conference on Computer Vision, 2003, pp. 946–951.
[10] J. Huber, V. Graefe, Motion stereo for mobile robots, IEEE Transactions on Industrial Electronics 41 (4) (1994) 378–383.
[11] S. Negahdaripour, B.Y. Hayashi, Y. Aloimonos, Direct motion stereo for passive navigation, IEEE Transactions on Robotics and Automation 11 (6) (1995) 829–843.
[12] P. Pritchett, A. Zisserman, Wide baseline stereo matching, in: Proceedings of the International Conference on Computer Vision, 1998, pp. 765–760.
[13] A. Baumberg, Reliable feature matching across widely separated views, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, 2000, pp. 774–781.
[14] F. Schaffalitzky, A. Zisserman, Viewpoint invariant texture matching and wide baseline stereo, in: Proceedings of the International Conference on Computer Vision, vol. 2, 2001, pp. 636–643.
[15] J. Matas, O. Chum, M. Urban, T. Pajdla, Robust wide-baseline stereo from maximally stable extremal regions, Image Vis. Comput. 22 (2004) 761–767.
[16] T. Tuytelaars, L.V. Gool, Matching widely separated views based on affine invariant regions, Int. J. Comput. Vis. 59 (1) (2004) 61–85.
[17] C.F. Olson, H. Abi-Rached, M. Ye, J.P. Hendrich, Wide-baseline stereo vision for Mars rovers, in: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, 2003, pp. 1302–1307.
[18] C.F. Olson, Maximum-likelihood image matching, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (6) (2002) 853–857.
[19] D.A. Forsyth, J. Ponce, Computer Vision: A Modern Approach, Prentice-Hall, Englewood Cliffs, NJ, 2003.
[20] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Plannery, Numerical Recipes in C, Cambridge University Press, 1988.
[21] A. Fusiello, E. Trucco, A. Verri, A compact algorithm for rectification of stereo pairs, Mach. Vis. Appl. 12 (2000) 16–22.
[22] O. Faugeras, B. Hotz, H. Mathieu, T. Viéville, Z. Zhang, P. Fua, E. Théron, L. Moll, G. Berry, J. Vuillemin, P. Bertin, C. Proy, Real time correlation-based stereo: algorithm, implementations and applications, Tech. Rep. RR-2013, INRIA (1993). URL http://www.inria.fr/rrrt/rr-2013.html.
[23] K. Mühlmann, D. Maier, J. Hesser, R. Männer, Calculating dense disparity maps from color stereo images, an efficient implementation, Int. J. Comput. Vis. 47 (1/2/3) (2002) 79–88.
[24] C.V. Stewart, Robust parameter estimation in computer vision, SIAM Rev. 41 (3) (1999) 513–537.
[25] K. Di, R. Li, L.H. Matthies, C.F. Olson, High precision landing site mapping and rover localization by integrated bundle adjustment of MPF surface images, in: International Archives of Photogrammetry and Remote Sensing, vol. XXXIV, Part 4, Joint International Symposium on Geospatial Theory, Processing and Applications (ISPRS Commission IV Symposium), 2002, pp. 733–737.
[26] R. Li, K. Di, F. Xu, Automatic Mars landing site mapping using surface-based images, in: ISPRS WG IV/9: Extraterrestrial Mapping Workshop—Advances in Planetary Mapping, 2003.
[27] F. Xu, K. Di, R. Li, L.H. Matthies, C.F. Olson, Automatic feature registration and DEM generation for Martian surface mapping, in: International Archives of Photogrammetry and Remote Sensing, vol. XXXIV, Part 2, ISPRS Commission II Symposium on Integrated Systems for Spatial Data Production, Custodian and Decision Support, 2002, pp. 549–554.
[28] W. Förstner, A framework for low-level feature extraction, in: Proceedings of the European Conference on Computer Vision, 1994, pp. 383–394.
[29] B. Triggs, P. McLauchlan, R. Hartley, A. Fitzgibbon, Bundle adjustment—a modern synthesis, in: Vision Algorithms: Theory and Practice, vol. 1883 of Lecture Notes in Computer Science, 2000, pp. 298–372.
[30] M.A. Oliver, R. Webster, Kriging: a method of interpolation for geographical information systems, Int. J. Geogr. Inf. Syst. 4 (3) (1990) 313–332.
[31] Y. Xiong, C.F. Olson, L.H. Matthies, Computing depth maps from descent images, Mach. Vis. Appl. 16 (3) (2005) 139–147.
[32] C.F. Olson, Image registration by aligning entropies, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, 2001, pp. 331–336.
[33] Z. Zhang, Iterative point matching for registration of free-form curves and surfaces, Int. J. Comput. Vis. 13 (2) (1994) 119–152.
[34] K. Nishino, I. Katsushi, Robust simultaneous registration of multiple range images, in: Proceedings of the 5th Asian Conference on Computer Vision, 2002, pp. 454–461.
[35] D. Miyazaki, T. Ooishi, T. Nishikawa, R. Sagawa, K. Nishino, T. Tomomatus, Y. Takase, K. Ikeuchi, The great Buddha project: modelling cultural heritage through observation, in: Proceedings of the Sixth International Conference on Virtual Systems and Multimedia, 2000, pp. 138–145.
[36] H. Hoppe, T. DeRose, T. Duchamp, J. McDonald, W. Stuetzle, Mesh optimization, in: Proceedings of SIGGRAPH '93, 1993, pp. 19–26.
[37] A. Hilton, A.J. Stoddard, J. Illingworth, T. Windeatt, Marching triangles: range image fusion for complex object modelling, in: Proceedings of the International Conference on Image Processing, 1996, pp. 381–384.
[38] C.F. Olson, L.H. Matthies, J.R. Wright, R. Li, K. Di, Visual terrain mapping for Mars exploration, in: Proceedings of the IEEE Aerospace Conference, vol. 2, 2004, pp. 762–771.