Machine Vision and Applications (2003) 14: 5–14
Digital Object Identifier (DOI) 10.1007/s00138-002-0097-7

Machine Vision and Applications

A stereo-vision system for support of planetary surface exploration

Maarten Vergauwen, Marc Pollefeys*, Luc Van Gool

K.U.Leuven ESAT-PSI, Kasteelpark Arenberg 10, 3001 Leuven, Belgium;
e-mail: {Maarten.Vergauwen,Marc.Pollefeys,Luc.VanGool}@esat.kuleuven.ac.be

Abstract. In this paper, we present a system that was developed for the European Space Agency (ESA) for the support of planetary exploration. The system that is sent to the planetary surface consists of a rover and a lander. The lander contains a stereo head equipped with a pan-tilt mechanism. This vision system is used both for modeling the terrain and for localization of the rover. Both tasks are necessary for the navigation of the rover. Due to the stress that occurs during the flight, a recalibration of the stereo-vision system is required once it is deployed on the planet. Practical limitations make it unfeasible to use a known calibration pattern for this purpose; therefore, a new calibration procedure had to be developed that could work on images of the planetary environment. This automatic procedure recovers the relative orientation of the cameras and the pan and tilt axes, as well as the exterior orientation for all the images. The same images are subsequently used to reconstruct the 3-D structure of the terrain. For this purpose, a dense stereo-matching algorithm is used that (after rectification) computes a disparity map. Finally, all the disparity maps are merged into a single digital terrain model. In this paper, a simple and elegant procedure is proposed that achieves that goal. The fact that the same images can be used for both calibration and 3-D reconstruction is important, since, in general, the communication bandwidth is very limited. In addition to navigation and path planning, the 3-D model of the terrain is also used for virtual-reality simulations of the mission, wherein the model is texture mapped with the original images. The system has been implemented, and the first tests on the ESA planetary terrain testbed were successful.

Key words: Prototype systems – Planetary exploration – 3D reconstruction – Stereo calibration

1 Introduction

The work described in this paper was performed within the scope of the ROBUST[1] project of the ESA. In this project, an end-to-end system is developed for a planetary exploration mission.

1.1 System overview

The ROBUST system consists of three important parts.

- Planetary rover: the Nanokhod (Fig. 1 top), a small and simple rover, designed to carry instruments in the immediate surroundings of a lander. It is equipped with a tether cable that provides the rover with power and a data connection to the lander, allowing a very high ratio of instrument mass to rover mass [13]. The rover consists of a payload cab and two connected tracks that are able to traverse a wide variety of soil-structures. The payload cab is subdivided into four chambers, each of which carries an instrument that can perform in situ experiments. In a process called *docking*, all of the chambers can be put in contact with the soil to extract samples.
- Planetary lander: contains the imaging head, an on-board computer, and the control system for both the Nanokhod and the imaging head. Figure 1 (bot) shows one of the cameras.
- On-ground control system.

The imaging head is used for both recording images, from which a reconstruction of the planetary terrain is computed, and for controlling the motion of the rover, using light emitting diodes on the payload cab. The imaging head consists of a stereo head mounted on a unit (approximately 1.5 m high) that allows for pan and tilt motions. The two cameras of the stereo head are space-approved 1024×1024 CCD cameras. The stereo head has a baseline of 0.5 m.

Correspondence to: M. Vergauwen
* Postdoctoral Fellow of the Fund for Scientific Research – Flanders, Belgium (F.W.O. – Vlaanderen).

[1] The ROBUST consortium consists of the Belgian companies SAS and OptiDrive, the K.U.Leuven departments PMA and ESAT-PSI, and the German companies DLR and vH&S.

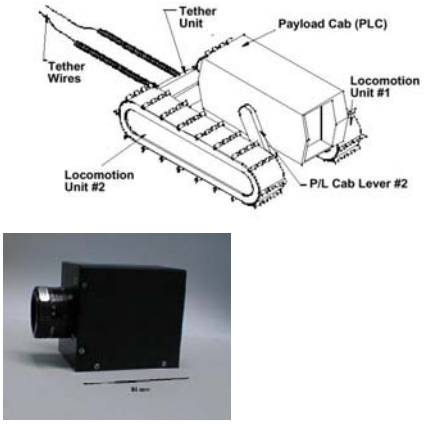

Fig. 1. Top: The Nanokhod; bot: one of the cameras of the stereo rig

1.2 Utilization

A typical utilization scenario will deploy the imaging head as soon as possible after the landing of the planetary system. Because of the strain on the parts during launch and landing, the imaging head needs to be recalibrated. To accomplish this, it takes images of the terrain that are sent to earth, where the calibration is performed using these images. From the same images a 3-D reconstruction of the terrain is then computed. Since the cameras have a limited field of view ($23^\circ \times 23^\circ$) the entire environment is not recorded at once but it is segmented into rings according to the tilt angle and each ring is divided into segments according to the pan angle of the imaging head (Fig. 2). The outermost boundary of the recorded terrain lies 20 m from the camera. For each of the segments, a stereo-image pair is recorded and sent down. The values of the actual pan and tilt angles can be read out from the encoders of the pan-tilt motors and sent down together with the corresponding images.

The 3-D reconstruction of the planetary terrain, which is obtained on earth, is then used by a team of scientists to indicate possible interesting sites for the Nanokhod to visit and extract and analyze samples from. A path-planning procedure is applied to compute a route for the Nanokhod to follow, which minimizes a cost by combining different factors, such as the risk of tipping over or the risk of losing visual contact. After a thorough verification through simulation, the route is up-linked to the lander. The control system on the lander observes the Nanokhod via stereo images of the light emitting diodes (LEDs) on the Nanokhod, which are taken by the imaging head. Differencing the images with the LEDs turned on with background images yields the coordinates of the LEDs in the stereo images and, hence, the pose of the Nanokhod on the terrain. This allows for the system to control the rover, such that it executes the uploaded tasks [14]. The results of the analysis of the samples, together with the telemetry of the rover, is sent back to the ground station, where the latter is visualized in a 3-D simulation environment.

In the following sections, the imaging-head calibration process, the reconstruction of the planetary terrain in 3-D, and the path-planning are explained in more detail. Results of a system test in a planetary testbed at the European Space Research and Technology Center (ESTEC) are presented, and some possible further developments are discussed.

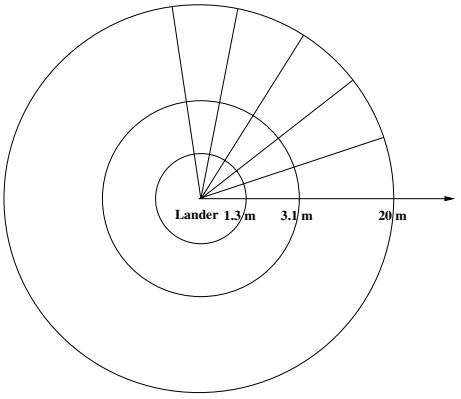

Fig. 2. Segmentation of the terrain

2 Calibration

Every planetary mission is a high-risk operation. During launch and landing, the lander and its contents are subject to extreme forces. The mechanical properties of the imaging head are likely to have been affected by mechanical and thermal forces. For high-accuracy equipment, such as the imaging head, a small change in these mechanical properties results in a large degradation of the results, unless the new properties can be estimated. The cameras themselves are built so that the intrinsic parameters during the mission can be assumed to be identical to the parameters obtained through calibration on ground. If the camera housing were not so rigidly built, the camera intrinsics would be likely to change during launch or landing. Algorithms exist that can retrieve these intrinsic parameters from images too [10, 16].

Traditional calibration algorithms rely on known calibration objects with well-defined optical characteristics in the scene. By taking pictures of these artificial objects, the pose of the cameras can be computed, yielding the extrinsic (mechanical) calibration of the stereo rig. There are two reasons why this scheme is not suited in our case, in which the imaging head is deployed on a distant planet. First, there is the problem of where to place the calibration objects. We need to be absolutely sure of the pose of these objects for the calibration to have any meaningful result. It is, of course, impossible to add objects to the terrain, so we have to consider placing calibration "markers" on the lander itself. A typical lander consists of a "cocoon" that opens after landing, comparable to an opening flower. The markers could be applied to the opening "petals." However, we are never sure of the exact position of these petals, which makes the markers much harder to use. Even if we were to dispose of accurate markers on the lander, a second problem would arise: During the design of the imaging head, robustness was a very important issue; for that reason, the number of moving items was minimized. Therefore, a fixed-focus lens was chosen. Since the accuracy of the stereo matching decreases with the square of the distance, the cameras are focused on the far range to gain as much accuracy in the far regions as possible. As a consequence, the images of near regions are blurred. Since the markers would be on the

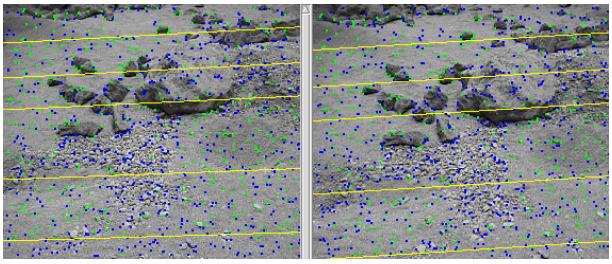

Fig. 3. Epipolar geometry of an image pair

lander, images of the markers would always be blurred, reducing the accuracy of the calibration up to the point that the markers would be useless. It is clear that standard calibration algorithms could not be used in our system. A new strategy had to be developed that only uses images of the (unknown) terrain to calibrate the imaging head.

The calibration procedure that was implemented for the ROBUST project is able to calibrate the imaging head using images of the terrain only. This means that the images sent down from the planet to earth to reconstruct the terrain can also be used for calibrating the imaging head. Therefore, the terrain-based calibration causes no important overhead on transmission. The calibration of the extrinsic (mechanical) properties of the imaging head is split into two parts, which are executed consecutively. First, the relative transformation between the two cameras is computed. Once this relative calibration is performed, a procedure can be executed which computes the transformations between the cameras and the lander. This boils down to computing the pan and tilt axes of the pan-tilt unit.

2.1 Relative calibration

The relative transformation between the two cameras of the imaging head can be computed from images of the terrain only. The algorithm to achieve this uses the concept of the essential matrix. This matrix represents the epipolar geometry between two views, including the internal parameters of the cameras as extra information. We make use of the fact that the relative transformation between the cameras does not change when the different segments of the terrain are recorded. This allows for different measurements of the epipolar geometry to be combined to yield one accurate solution.

If the essential matrix between the two views is computed, the relative transformation (position and orientation) between the two cameras can be calculated up to the baseline (i.e., the distance between the two cameras). Both steps are explained in more detail in the following.

2.1.1 Computing epipolar geometry

The first step in obtaining the relative calibration is the computation of the epipolar geometry of the stereo head. The epipolar geometry constraint limits the search for the correspondence of a point in one image to points on a line in the second image (Fig. 3).

The epipolar geometry between two images can be expressed in the form of the fundamental matrix $\mathbf{F}$ [3]. For all

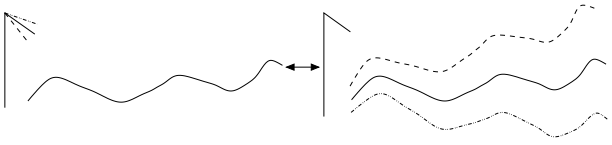

Fig. 4. Combining different segments

corresponding points m_1 in image 1 and m_2 in image 2, the following equation holds.

$$m_2^T \mathbf{F} m_1 = 0, \tag{1}$$

where $\mathbf{F}$ is a 3×3 matrix of rank 2. This matrix can be computed linearly from 8 or nonlinearly from 7 correspondences. To find back the epipolar geometry between two images automatically, a feature detector, called the Harris corner detector [4] is first applied to the images. This yields a set of salient features in the images, which are, typically, very well suited for matching.

Next, the extracted corners are matched automatically between pairs of images using cross correlation. This process yields a set of possible matches, which is, typically, contaminated with an important number of wrong matches or outliers, making it impossible to use standard methods, such as weighted least squares, to compute the fundamental matrix. A robust matching scheme, called random sampling consensus (RANSAC) [2], is used to compute and update epipolar geometry and matches iteratively. The results of the RANSAC procedure are the epipolar geometry in the form of $\mathbf{F}$ and the inlying matches (correspondences that satisfy Eq. 1).

In the case of the ROBUST, imaging-head data of the different segments of the terrain can be combined to compute the epipolar geometry much more robustly because the relative transformation between the cameras does not change (Fig. 4). Stereo images of different rings are obtained by tilting the imaging head. However, we can imagine the camera being kept steady and the terrain being tilted. This would result in the same stereo images. That is why the possible correspondences of the different rings and segments can be combined (as if they were recorded by one stereo view) to compute the epipolar geometry more accurately.

It is even the case that a specific degenerate case for the computation of the epipolar geometry is solved by the combination scheme we describe. Computing the epipolar geometry for a pair of images of a *planar* scene is impossible from correspondences only. If the planetary terrain is planar or close to it, computing the epipolar geometry for one pair of images becomes an ill-posed problem. By combining correspondences from different segments, this problem is solved.

2.1.2 Computing relative transformation

Once the epipolar geometry is computed in the form of the fundamental matrix $\mathbf{F}$, the relative transformation between the two cameras of the imaging head can be calculated. First, the essential matrix is obtained as $\mathbf{E} = \mathbf{K}^\top \mathbf{F} \mathbf{K}$ with $\mathbf{K}$, the 3×3 matrix with the intrinsic calibration of the cameras. The relative transformation $(\mathbf{R}, \mathbf{t})$ (up to scale) can easily be obtained from it [8], since $\mathbf{E} = [\mathbf{t}]_\times \mathbf{R}$. The relative rotation $\mathbf{R}$ can thus computed as $\mathbf{R} = \mathbf{U}\mathbf{B}\mathbf{V}^\mathbf{T}$ with $\mathbf{U}$ and $\mathbf{V}$ the

matrices containing the left and right singular vectors of $\mathbf{E}$ (i.e., $\mathbf{E} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathbf{T}}$), and $\mathbf{B}$ a matrix representing a rotation around the optical camera axis over $\frac{\pi}{2}$ or $-\frac{\pi}{2}$. The direction of the relative translation is easily computed, since the epipole (the projection of the second camera center in the first image) is known from the fundamental matrix.

The absolute value of the distance between the two cameras, i.e., the baseline, can not be computed from the images. It is unlikely, however, that this value will deviate much from the original value. Furthermore, since all measurements for 3-D reconstructions and localization will consistently be carried out using the same stereo head, this has very little influence. The only case where absolute measurements are really needed is for estimating tether consumption and the absolute size of obstacles to overcome. In these cases even a deviation of a few percent would be smaller than the uncertainty induced by other factors.

The computed values for $\mathbf{R}$ and $\mathbf{t}$ are used as an initialization for a nonlinear Levenberg–Marquardt minimization, which finds back the values of $\mathbf{R}$ and $\mathbf{t}$ that minimize sum of all distances between points and their corresponding epipolar lines. The result is a very accurate estimation of the relative transformation between the two images.

2.2 Pan-tilt calibration

To be able to bring all the measurements in a single frame, it is not sufficient to have the relative calibration between the cameras; the locations of the pan and the tilt axes are also needed. Since, for the same reasons as for the relative calibration, these values can change due to the stress that occurs during launch and landing, we need to calibrate these axes.

The evaluation of the pan and the tilt axes is more complicated than the relative calibration, but can also be achieved from the image data, if, at least, some overlap exists between neighboring image pairs. This is guaranteed in the pan direction due to the fixed vergence setup that does not allow 100% overlap between two views of the same stereo pair (hence, yields overlap with neighboring pairs). For the tilt axis, care should be taken to foresee a sufficient overlap between the rings of Fig. 2 for, at least, some image pairs.

2.2.1 Relative transformations between views

To calibrate the pan and tilt axes, stereo images of the same ring and the same segment are used, respectively. The overlap between consecutive stereo images is especially important in the strategy.

For the calibration of the tilt axis, a stereo image of the outer ring of a certain segment is recorded. The imaging head is commanded to execute a tilt motion and to record a stereo image of the second ring. We have to make sure that there is sufficient overlap between the two image pairs (Fig. 5). The area shaded from the bottom left to the top right is visible in the first view. The area shaded from the bottom right to the top left is visible in the second. The overlapping area is, of course, visible in both views.

Corresponding features in the images of the first image pair can be found, as explained in Sect. 2.1.1. Because we know the relative transformation between the two cameras, we can reconstruct the features in 3D. The same is done in the second image pair. Because of the overlap, some of the features will be visible in both image pairs. We can find correspondences of these features by running the matching algorithm from Sect. 2.1.1 on the two images of the left or the right camera. The corresponding features allow us to align the reconstruction of the second pair with the reconstruction of the first pair. This yields the relative transformation between the first and second imaging-head frames (Fig. 6 explains this procedure in more detail). Since only a motion of the tilt axis took place between the recording of the first and second stereo pairs, the tilt axis can be very easily computed from the relative transformation between the first and second imaging-head frames.

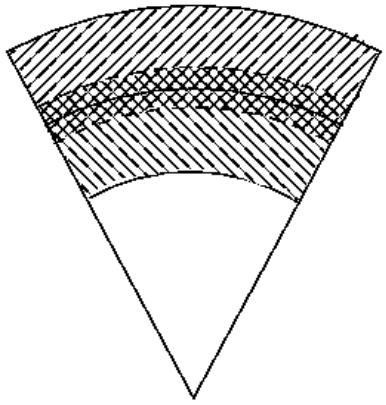

Fig. 5. Symbolic representation of the setup for the computation of a relative transformation for a tilt motion

For the pan axis, the computation of the relative transformation between two views is slightly different, as shown in Fig. 7.

It is clear that, in this case, there are almost no features that are present in all four images of the two views. Due to the verging of the two cameras, however, there are some features that can be seen in one stereo view and in one image of the other image pair. One of these features is represented by a yellow dot in Fig. 7. Again we find back corresponding features between the left image of the first pair and the right image of the second pair with the algorithm from Sect. 2.1.1, shown in Fig. 7 as the yellow arrow. Because the features are visible in both images of the first stereo view, they can be reconstructed in 3D. They are also visible in one image of the second view, and correspondences are found (the blue arrow in Fig. 7). One can apply a pose estimation of the camera of the second pair, in which the features are visible, yielding the pose of this camera in the frame of the first view. Since only a motion of the pan axis took place between the recording of the first and second stereo view, the pan axis can be computed from this result.

2.2.2 Actual calibration of pan and tilt axes

During the acquisition of the data, we try not to change the pan angle if a pure tilt rotation is executed and vice versa. In any real system, however, there will be deviations from the desired angles, which means that the computation of the tilt axis will not be correct. This is because the algorithm computes the

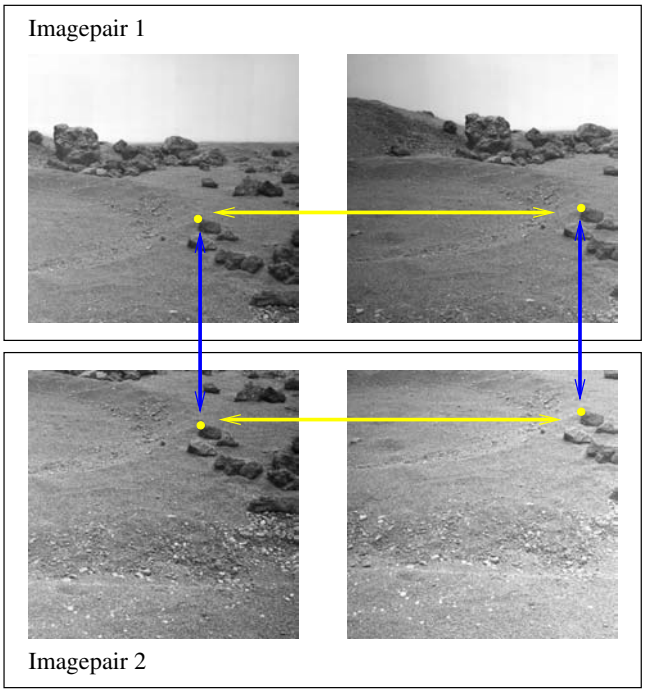

Fig. 6. Different correspondences for computing the tilt axis: The upper two images are the left and right images of a stereo pair taken for a tilt angle of 24°. The lower images are the left and right images of a stereo pair, taken for a tilt angle of 14°. The *light arrows* denote correspondences between images of the same stereo pair. The *dark arrows* denote correspondences between the *left* and *right* images, respectively

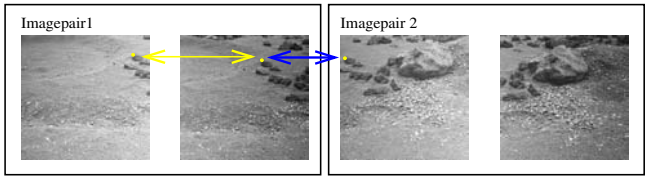

Fig. 7. Matching between different views for a pan motion

real rotation axis, which is not the tilt axis if there is even a small pan component. Nonetheless, there is a solution to this problem: In the second step, a good approximation of the pan axis was found. Therefore, if we account for the small deviations of pan, which are read out from the encoders, with the current computed value of the pan axis, we can recompute the tilt axis more accurately. This, in turn, allows us to update the pan axis, and so on. We can repeat this iterative procedure until the solution for the axes has converged. In reality, three iterations have proven to be sufficient.

2.3 Synthetic calibration experiment

The calibration algorithm has been tested on artificial data. A virtual planar scene with texture from a real image from Mars was constructed, and pairs of images were generated with a visualization toolkit. We chose a planar scene to show that our algorithms could deal with this degenerate case. First, the relative calibration between the two cameras was computed. During calibration, data from nine image pairs was combined, and 2591 corners were matched to calculate the relative transformation. We could compare the result with the ground-truth value of the relative transformation. For comparison, the rotational part of the relative transformation was represented as a rotation around an axis (l, m, n) with a certain angle θ. The ground truth was

$$(l_0, m_0, n_0) = (0, 1, 0) \text{ and } \theta_0 = 10^\circ.$$

The computed values were

$$(l, m, n) = (0.000397, 0.99951, 0.00018) \text{ and } \theta = 10.093^\circ.$$

The angle between (l, m, n) and (l_0, m_0, n_0) was 0.0252°. The difference between θ and θ_0 was 0.093°. Both values were small, meaning that the rotation was estimated accurately because of the combination of data.

The pan and tilt axes were calibrated from the same data. The angle between the computed tilt axis and the ground truth was 0.138°. The angle between the computed pan axis and the ground truth was 0.618°. The larger error for the pan axis can be explained from the fact that only correspondences between three images were used to estimate it, while correspondences from four images could be exploited to compute the tilt axis.

3 3-D terrain modeling

After the calibration of the imaging head is performed, the process of generating a 3-D model or models of the planetary terrain can start. This modeling is vital to accomplishing the goal of planetary exploration. Its input are all images of the terrain and the calibration of the imaging head. The output of the terrain modeling can have different forms, but the most important one is the digital elevation map (DEM). In the following sections, we will describe the different steps that are performed to obtain such a DEM. (We should mention that, in the unfortunate event of one failing camera on the imaging head, techniques exist to construct stereo panoramas from monocular images taken by a camera that rotates around an axis that does not go through the camera's center [9].)

3.1 Generation of disparity maps

On each pair of images recorded by the imaging head, a stereo algorithm is applied to compute the disparity maps from the left image to the right and from the right image to the left. Disparity maps are an elegant way to describe correspondences of two images if the images are *rectified* first. The process of rectification remaps the image pair to standard geometry with the epipolar lines coinciding with the image scan lines [11, 15]. The correspondence search is then reduced to a matching of the image points along each image scan line. The result (the disparity map) is an image in which the value of each pixel corresponds to the number of pixels that have to be moved to the left or right to find the corresponding pixel in the other image.

In addition to the epipolar geometry, other constraints, such as preserving the order of neighboring pixels, bidirectional uniqueness of the match, and detection of occlusions

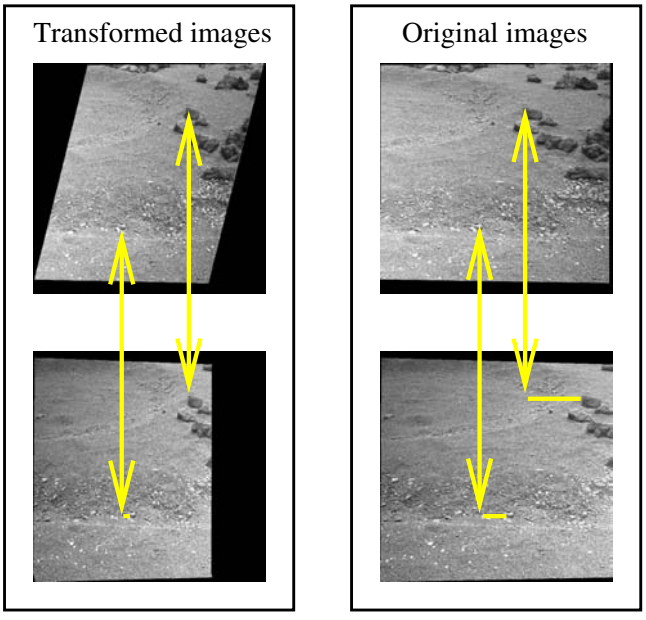

Fig. 8. Transforming rectified images decreases disparity ranges by a large factor. On the left, the disparity for two rocks in the transformed images is clearly much smaller than on the right in the original rectified images

can be exploited. The dense correspondence scheme we employ to construct the disparity maps is the one described in [6]. This scheme is based on the dynamic-programming scheme of Cox [1]. It operates on rectified image pairs and incorporates the above-mentioned constraints. The matcher searches at each pixel in the left image for the maximum normalized cross correlation in the right image by shifting a small measurement window along the corresponding scan line. Matching ambiguities are resolved by exploiting the ordering constraint in the dynamic-programming approach.

Due to the fixed vergence of the cameras, very large disparity ranges (from -100 up to 500 pixels) can appear between two images of an image pair, especially in the very near and far regions of the terrain. This forces the stereo algorithm to search in a very large search range and causes problems with memory management and speed. To deal with this problem we transform the images in such a way that the disparity range is reduced. This can be done assuming that the terrain can be approximated by a horizontal plane. We transform the rectified images such that a point on the horizontal plane would have disparity zero. Because the terrains, typically, don't differ too much from the horizontal plane, the disparities are greatly reduced. Figure 8 gives an example.

3.2 Digital elevation maps

A DEM can be seen as a collection of points in a "top view" of the 3-D terrain, where each point has its own height or "elevation." The algorithm proposed for generating regular DEMs in the ROBUST project fills in a "top view" image of the terrain completely, i.e., a height value can be computed for every pixel in the top view image, except for pixels that are not visible in the images because of occlusions. These occlusions are found in a very simple way.

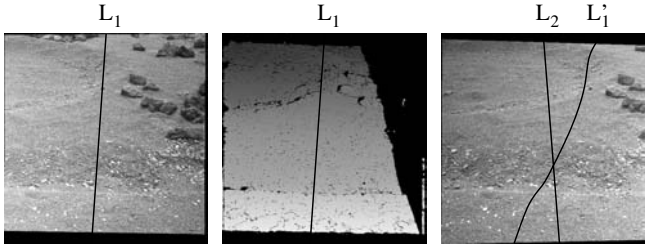

Fig. 9. Digital elevation map generation in detail. The figure contains the left rectified image (*left*), the corresponding disparity map (*middle*) and the right rectified image (*right*). The 3-D line, L, corresponding to a point of the DEM projects to L_1 and L_2, respectively. Through the disparity map, the *shadow* of L_1 on the right image can also be computed. The intersection point of L_2 and L_1 corresponds to the point where L intersects the surface

The terrain is divided into cells: the pixels of the DEM. For each cell, the stereo-pair image is selected in which the cell would be visible if it had a height of zero. A vertical line is drawn and the projection of this line in the left and right images of the stereo pair is computed. (Figure 9 illustrates the algorithm that is used to determine the height of the terrain on that line.)

The two images on the left and right are the disparity images of the left and right stereo images, respectively. We display the untransformed images here because they are smaller. The solid lines $\mathbf{L_1}$ and $\mathbf{L_2}$ are the projections of the vertical line in both disparity images. Image placing a light where the left camera is. This light shines on the vertical line, which throws a shadow on the terrain. In the left image, this shadow, of course, has the same projection as the line itself. In the right image, however, this is not the case. The projection of the shadow in this image is the smooth curve, $\mathbf{L_1'}$. The part of this curve above the intersection of $\mathbf{L_1'}$ and $\mathbf{L_2}$ is the "real" part of the shadow (i.e., it would be visible on the terrain). The other part can be seen as the "virtual" part of the shadow, coming from the part of the vertical line below the surface of the terrain. This shadow curve can be computed using the disparity in the left disparity image of every pixel of the projected line, $\mathbf{L_1}$. The intersection point of the vertical line and the terrain can then be found as the point where the shadow $\mathbf{L_1'}$ intersects the line $\mathbf{L_2}$. Some remarks can be made about this procedure.

- Occluded regions are detected easily. This is the case for cells that are not visible in both stereo images. The height value of these cells can not be computed, and these cells get a certain predefined value in the DEM, which marks them as unseen.
- This particular scheme makes it possible to generate regular digital elevation maps at any desired resolution, interpolating automatically if needed.
- For the parts of the terrain close to the boundary of a ring, different parts of the vertical line will be projected in different stereo views. Therefore, it is possible that information from two different stereo views has to be combined to find back the intersection. This is not a problem because the transformation between the views can easily be computed, since the calibration has been calculated.

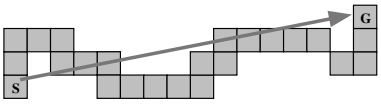

Fig. 10. Corridor

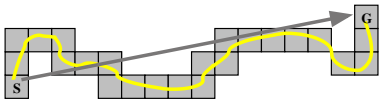

Fig. 11. Refined path

Fig. 12. Mosaic of images of the testbed taken by the stereo head

4 Path planning

Once the planetary terrain has been reconstructed, scientists can indicate sites of interest (points of reference or PORs) that the rover should visit. Paths have to be computed between successive PORs. For this purpose, a path-planning module has been developed. Given a terrain map, an initial rover position and heading, and a desired rover position and heading, the path planner will find a path, using an A* framework [7], which takes the rover from the initial state to the goal state and which is optimal in terms of a series of parameters, such as energy consumption, risk of tip-over, use of tether, etc.

4.1 Travel cost map

The travel cost map (TCM) provides a measure for the cost of traverse, based on metrics inherent in the terrain. In the current implementation, a simple metric based on the gradient of the DEM is used. Another metric characterizes the uncertainty of the terrain data; the farther from the lander camera the higher the uncertainty. Areas occluded by rocks also have high uncertainty.

4.2 The hierarchical approach

The rover can move according to a set of available operators (also called rover movements), which take the rover from one position and heading (this pair is also known as a state) to another position and heading. Each operator has an associated cost. The main term of this cost is computed from the above-mentioned TCM. Given that A* is computationally very complex, finding a path in a reasonably large terrain, using complex operators for the rover movements, can take a long time. This has led to the choice of a hierarchical approach to the path-planning problem.

4.2.1 Finding the corridor

At the first stage, a traverse is planned between the start and goal states using A* to cover the whole terrain. However, since the cells are somewhat larger than the size of the rover, the resolution is reduced so that the rover can move comfortably within the corridor. A low-resolution TCM is used for this. The transition operators are simple, i.e., forward, backward, left, and right, which allows us to apply a highly optimized and fast version of A*. The result is a corridor (Fig. 10) in which the rover may safely move.

4.2.2 Refinement of the Path

At the second stage, the path is refined using the high-resolution TCM. By restricting the search to cells marked in the restrain grid that was constructed in the previous stage, more complex operators and full available resolution can be used within reasonable time constraints.

The representation of the operators to take the rover from one state to another is kept very general, i.e., a rotation followed by a translation. The cost of applying an operator is determined by using a number of cost evaluation points. The cost is calculated as a weighted sum of the costs at these points evaluated at the resulting rover pose. The evaluation points represent parts of the "virtual" rover during and after the completion of the corresponding move. The positions of the evaluation points are calculated, based on the rover dimensions, the parameters of the rover movement, the desired safety margin, and the resolution of the TCM.

The result of the hierarchical A* algorithm is a high-resolution path (Fig. 11), represented by an ordered list of rover poses, bringing the rover from its start pose to the desired destination.

5 System test

A first test of the complete system was performed at the ESA–ESTEC test facilities in Noordwijk, The Netherlands, where access to a planetary testbed of about 7×7 m was available. The imaging head was set up next to the testbed. Its first task was the to record the terrain according to the setup shown in Fig. 2. A mosaic of the pictures taken by this process is shown in Fig. 12.

The autonomous calibration procedure was launched, and it computed the extrinsic calibration of the cameras based on the images. Once the calibration had been computed, the system rectified the images and computed dense disparity maps. Based on these, a digital elevation map was constructed. The result can be seen in Fig. 13. Because of the relatively low height of the imaging head (approximately 1.5 m above the testbed) and the big rocks in the testbed, a large portion of the digital elevation map could not be filled in because of occlusions. The digital elevation map of Fig. 13 was then used to construct a textured triangulated mesh model. Some views of this model can be seen in Fig. 14. A striping effect is visible in the texture of the model. This was due to a hardware problem with the cameras, which caused the images to suffer from an intensity gradient. Since the 3-D model combines data from all images, the boundaries are visible.

Once the reconstruction was complete, the model was used to plan a trajectory for the Nanokhod rover. At that time, an-

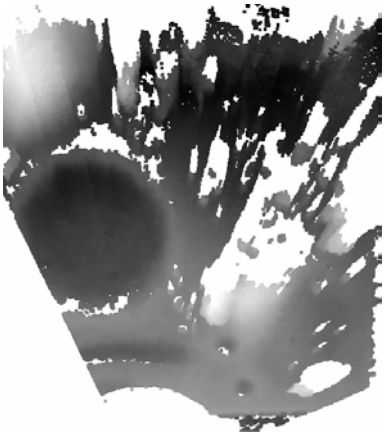

Fig. 13. Digital elevation map of the ESTEC planetary testbed. A significant number of cells are not filled in because they are located in occluded areas. Darker pixels are lower, lighter pixels have a greater height. The crater in the middle left of the DEM and the hill behind it can be clearly detected

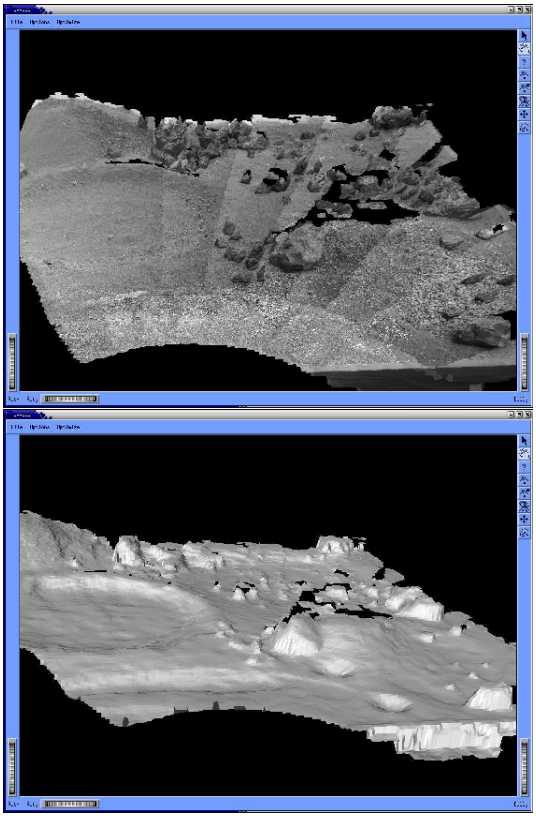

Fig. 14. A textured and a nontextured view of the reconstructed model of the ESTEC planetary testbed

other component of the system came into play: the *simulator*. This component allowed the operator to simulate low-level commands on the Nanokhod and imaging head in a virtual-reality environment. A simple gravitation model was used to predict the pose of the Nanokhod after a motion. An overview of the GUI of this component is shown in Fig. 15. The GUI is divided into four windows. The upper two show what the left and right camera of the stereo head were seeing. The bottom-right window is a 3-D interaction window, and the bottom-left

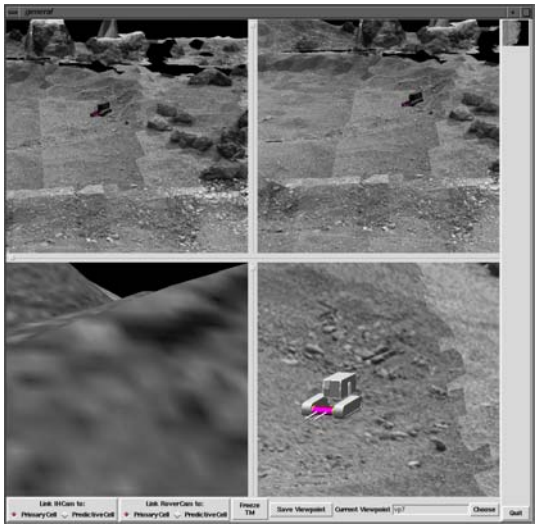

Fig. 15. Overview of the simulator. The upper two windows simulate the view of the two cameras. The bottom right window is a 3-D interaction window and the bottom left window gives a view from the cab of the rover

window shows a picture of what a camera mounted in the cab of the Nanokhod would have seen.

When the trajectory was simulated, the commands were uploaded to the lander system which executed them autonomously. The pose of the rover was observed using the imaging head. The lander needed the calibration parameters, computed on the ground station for this and made use of the LEDs that are present on the Nanokhod. The telemetry, computed by the lander could then (after downlink to the ground station) be played back by the simulator and compared with the simulated pose of the Nanokhod.

6 Future development

For the ROBUST project, calibration of the imaging head is a critical issue. Because of known specifics of the imaging head, the calibration algorithms can be targeted to this particular setup and take all known information on mechanics and intrinsics of the system into account. One can imagine situations where such information is not available. While the algorithms described in this paper can no longer be used, it is still possible to retrieve some calibration and 3-D reconstruction.

6.1 Uncalibrated 3-D reconstruction

How structure and motion can be retrieved from an uncalibrated image sequence is described in [12]. The 3-D modeling task is decomposed into a number of steps. First, successive images of the sequence are matched, and epipolar geometry is computed using the same approach as described in Sect. 2.1.1. The projection matrices of all views are computed by triangulation and pose-estimation algorithms. The ambiguity on the reconstruction is then reduced from pure projective to metric (Euclidean up to scale) by self-calibration algorithms (see also [10]). Finally, textured triangular mesh models of the scene are constructed.

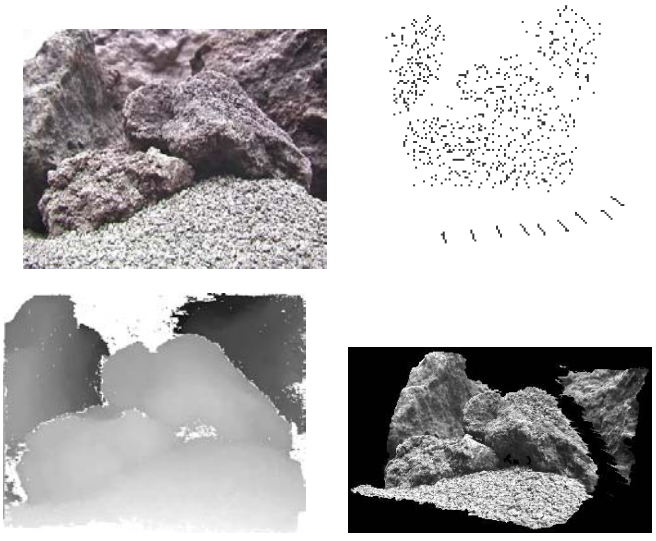

Fig. 16. Results of the reconstruction process of a rock on the planetary testbed. The upper left view shows one of the original images of the sequence. The upper right view displays reconstructed points and cameras. In the lower left view, a dense depth map is shown. The lower right view shows an arbitrary view of the 3-D reconstruction

6.2 Experiment

The technique described could be of use during a planetary exploration mission. The most important instrument of the payload cab of the rover is probably a camera. This camera will take images of samples and rocks on the terrain, but it could also be used to perform close-range reconstruction of the terrain, helping the rover to "dock" precisely onto the desired position. The camera would take images of the area of interest during the approach. These images could then be used as input for the reconstruction algorithm described above to generate a 3-D reconstruction. The resolution of this reconstruction would be far superior to that of the DEM obtained from the imaging head.

During testing of the ROBUST system on the planetary testbed at ESTEC, a preliminary test was performed. Images of some rocks on the testbed were taken by hand with a semi-professional digital camera. The images were processed by the 3-D reconstruction system. The resulting 3-D reconstruction of the rock is shown in Fig. 16. The results are very good and show that this strategy could be an interesting extension of the ROBUST system.

7 Conclusion

In this paper, we have proposed an approach for calibration and 3-D measurement from planetary-terrain images that allows for an important simplification of the design of the imaging system of the lander. The components described in this paper are part of an end-to-end system that can reconstruct an unknown planetary terrain and guide a rover autonomously on the planetary surface. The system has succeeded in a first test in a planetary testbed at ESTEC.

Acknowledgements. We acknowledge support from the Belgian IUAP4/24 IMechS project. We also wish to thank all partners of the ROBUST project for the collaboration. Special thanks go to Ronny Moreas of K.U.Leuven-PMA for his contribution to the path planning algorithm.

References

1. Cox I, Hingorani S, Rao S (1996) A maximum likelihood stereo algorithm, In: Comput Vision Image Underst 63:3
2. Fischler M, Bolles R (1981) RANdom SAmpling Consensus: a paradigm for model fitting with application to image analysis and automated cartography. Commun Assoc Comput Mach 24:381–395
3. Faugeras O (1992) What can be seen in three dimensions with an uncalibrated stereo rig. In: Computer Vision (ECCV'92), Lecture notes in computer science, vol 588. Springer, Berlin Heidelberg New York, pp 563–578
4. Harris C, Stephens M (1988) A combined corner and edge detector. In: 4th Alvey Vision Conference, pp 147–151
5. Knight J, Reid I (2000) Self-calibration of a stereo-rig in a planar scene by data combination. In: Proceedings of the International Conference on Pattern Recognition (ICPR 2000), Barcelona, IEEE Press, Piscataway, NJ, pp 411–414
6. Koch R (1997) Automatische Oberflachenmodellierung starrer dreidimensionaler Objekte aus stereoskopischen Rundum-Ansichten. PhD thesis, University of Hannover, Germany, 1996 also published as Fortschritte-Berichte VDI, Reihe 10, Nr499, VDI, Düsseldorf
7. Latombe JC (1991) Robot motion planning. In: International series in engineering and computer science, vol 124. Kluwer, Dordrecht
8. Maybank S (1992) Theory of reconstruction from image motion. Springer, Berlin Heidelberg New York
9. Peleg S, Ben-Ezra M (1999) Stereo panorama with a single camera. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE Press, NY, pp 395–401
10. Pollefeys M, Koch R, Van Gool L (1999) Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters. Int J Comput Vision 32:7–25
11. Pollefeys M, Koch R, Van Gool L (1999) A simple and efficient rectification method for general motion. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV'99), Corfu, Greece, September, IEEE Computer Society Press, Kerkyra, Greece, pp 496–501
12. Pollefeys M, Koch R, Vergauwen M, Van Gool L (1998) Metric 3-D Surface reconstruction from uncalibrated image sequences. In: Proceedings of the SMILE Workshop (post-ECCV'98). Lecture notes on computer science, vol 1506. pp 138–153, Springer, Berlin Heidelberg New York
13. Rieder R, Wanke H, Hoerner Hv (1995) Nanokhod, a miniature deployment device with instrumentation for chemical, mineralogical and geological analysis of planetary surfaces, for use in connection with fixed planetary surface stations. In: Lunar Planet Sci XXVI:1261–1262
14. Steinmetz B, Arbter K, Brunner B, Landzettel K (2001) Autonomous vision-based navigation of the Nankhod rover, In: Proceedings of the International Symposium on Artificial Intelligence and Robotics and Automation in Space (iSAIRAS 2001), Montreal, June. CSA, Hubert, Canada
15. Loop C, Zhang Z (1999) Computing rectifying homographies for stereo vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'99), Fort Collins, Colo, June. IEEE Press, NY

16. Zisserman A, Beardsley P, Reid I (1995) Metric calibration of a stereo rig. In: Proceedings of the IEEE Workshop on Representation of Visual Scenes, Cambridge, Mass. pp 93–100

Maarten Vergauwen was born in '74 and got a degree in electronical engineering at the Katholieke Universiteit Leuven in Belgium in 1997. Since then he has been a phd student at the Katholieke Universiteit Leuven. He has been involved in several projects among which the ESA projects Viable and Robust.

Marc Pollefeys is an assistant professor of Computer Vision in the Department of Computer Science of the University of North Carolina-Chapel Hill. Previously he was a post-doctoral researcher at the K.U.Leuven where he also received his M.S. and Ph.D. in '94 and '99 respectively. He has received several prizes, amongst which the Marr prize at ICCV'98, and was the author or co-author of over 70 technical papers. Marc Pollefeys has organized workshops and courses at major vision and graphics conferences, has been on the program committee of multiple conferences and is a regular reviewer for most of the vision, graphics and photogrammetry journals.

Luc Van Gool is professor for computer vision at the K.U.Leuven in Belgium and the ETH in Zurich, Switzerland. He leads a research team at both universities. He has authored over 100 papers in computer vision. He has been a PC member of several, major conferences and is in the editorial boards of several international journals. His main interests include 3D modeling, animation, tracking, gesture analysis, recognition, and texture. He received a David Marr Prize, a Tech-Art Prize and an EITC Prize from the EC. He was the coordinator of several European projects. He is a co-founder of the company Eyetronics.