

# Autonomous Operations of a Micro-Rover for Geo-Science on Mars

Maarten Vergauwen, Marc Pollefeys\* and Luc Van Gool  
ESAT-PSI, K.U.Leuven, Belgium

## 1 Introduction

The work described in this paper was performed in the scope of the ROBUST<sup>1</sup> project of the European Space Agency (ESA). In this project an end-to-end system is developed for a planetary exploration mission. For a general overview of this project we refer to [2]. This paper will focus on two subsections of the system: the Calibration of the Imaging Head and the Terrain Reconstruction.

### 1.1 Imaging Head

The Imaging Head is both used for recording images from which a reconstruction of the planetary terrain is computed and for controlling the motion of the rover. It consists of a stereo head, mounted on a unit which allows for pan and tilt motions and which is approximately 1.5 meter high.

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 which are sent to earth where the calibration is performed using these images. From the same images a 3D reconstruction of the terrain is then computed. Since the cameras have a limited field of view (23x23 degrees) the entire environment is not recorded at once but it is segmented into rings and each ring is divided into segments. For each of the segments a stereo image pair is recorded and sent down.

## 2 Calibration

The mechanical properties of the Imaging Head are likely to be affected by mechanical and thermal effects. For such high accuracy equipment, 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 identical to the ones obtained through pre-launch calibration on ground.

\*Postdoctoral Fellow of the Fund for Scientific Research - Flanders (Belgium)(F.W.O. - Vlaanderen).

<sup>1</sup>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.

### 2.1 Using markers ?

Traditional calibration algorithms rely on known calibration objects in the scene. The pose of the cameras can be computed from images of these artificial objects, yielding the extrinsic calibration [6]. This scheme can't be used in this case because one can not be certain of the correct pose of calibration "markers" on a deployed lander. Even if one did dispose of accurate markers, a second problem arises. For sake of robustness, no zooming and focusing system was added to the cameras. To gain accuracy in the far regions, the cameras are focussed at infinity. As a consequence, the images of near regions are blurred. Since the markers would be on the lander, images of them would always be blurred, reducing the accuracy of the calibration. A new strategy had to be developed that only uses images of the terrain to calibrate the Imaging Head.

### 2.2 Employed strategy

The calibration procedure that was implemented for the ROBUST project is able to calibrate the Imaging Head using images of the terrain only. The calibration of the extrinsic (mechanical) properties is split into two parts which are executed consecutively. First the relative transformation between the two cameras is computed. This is explained in Section 3. Once this relative calibration is performed, a procedure can be performed which computes the relative transformations between the cameras and the lander. This boils down to computing the pan and tilt axes of the pan-tilt unit. Section 4 deals with this problem.

## 3 Relative calibration

The relative transformation between the two cameras of the Imaging Head can be computed from images only. The algorithm to do 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 the different segments of the terrain are recorded,

which 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).

### 3.1 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. If one wants to find back the epipolar geometry between two images automatically, a filter, called the “Harris Corner Detector” [3] is applied to the images first. The result consists of salient points or *corners* in the images. Next, the 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 mismatches or outliers. Therefore a robust matching scheme, called RANSAC, is used to compute and update epipolar geometry and matches iteratively.

In the case of the ROBUST Imaging Head the data of the different segments of the terrain can be combined to compute the epipolar geometry because the relative transformation between the cameras does not change. Stereo images of different rings are obtained by tilting the Imaging Head. However, one could imagine the camera to be kept steady and the terrain to be tilted. This would result in the same stereo images. That’s why the possible correspondences of the different rings and segments can be combined to compute the epipolar geometry more accurately.

### 3.2 Computing the relative transformation

Once the epipolar geometry is computed in the form of the fundamental matrix  $F$ , the relative transformation between the two cameras of the Imaging Head can be calculated. First the essential matrix is constructed. This is easily done since  $E = K^T F K$  with  $K$  the 3x3 matrix with the intrinsic calibration of the cameras. To derive the relative translation and rotation from the essential matrix, we refer to the work of Maybank et al. [4]. There is one parameter we can not calibrate: the actual value of the baseline. We can however assume that this value will not deviate much from the mechanical specs. If there were some change in the actual value of the baseline, the consequences of fixing it to the (wrong) value of the specs are not harsh because all measurements (terrain reconstruction and rover localization) are done within the same measurement system. The

computed values for  $R$  and  $t$  are used as an initialization for a non-linear Levenberg-Marquardt minimization which finds back the values of  $R$  and  $t$  that minimize sum of all distances between points and their corresponding epipolar lines.

## 4 Pan-tilt calibration

Computing the relative transformation between the two cameras is an important part of the calibration but it does not suffice. For rover localization and generation of terrain reconstructions, the transformations between the cameras and the Imaging Head and between the Imaging Head and the lander need to be known as well.

### 4.1 The Imaging Head frame

For sake of clarity a virtual “**Imaging Head frame**” is defined in “the middle” of the two cameras. This means that the relative translation and rotation between the left camera and the Imaging Head frame is equal to the translation and rotation between the Imaging Head frame and the right camera.

### 4.2 From Imaging Head to Lander

Calibrating the relative transformation between the Imaging Head frame and the lander is more complicated because it implies calibration of the pan and tilt axes. This transformation depends on the actual angle of rotation around both the pan and tilt axis. From the world’s point of view, the pan axis is never altered but the orientation of the tilt axis depends on the pan angle. If we look from the point of view of the IH however, it is the tilt axis that never changes and the orientation of the pan axis depends on the tilt angle. The latter view will be employed because it fits very well in the philosophy where one derives the entire chain of calibration transformations from the cameras, which are the only measurement device, to the lander.

### 4.3 Relative transformations between views

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

**Tilt** For the calibration of the tilt axis, a stereo image of the outer ring of a certain segment is recorded. The IH is commanded to execute a tilt motion and to record a stereo image of the second ring. There should be

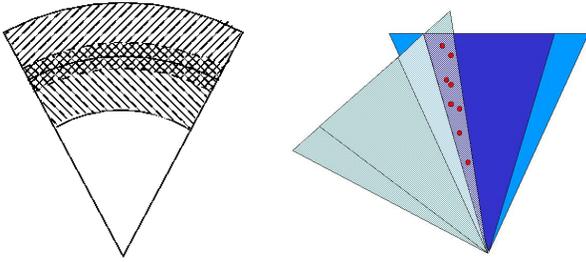


Figure 1: Symbolic representation of the setup for the computation of a relative transformation for a tilt motion(left) and for a pan motion (right)

sufficient overlap between the two image-pairs. This setup is shown on the left in figure 1. Corresponding features in the images of the first image pair can be found as explained in Section 3.1. 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. We can find correspondences between features that are visible in both pairs by running the matching algorithm of Section 3.1 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 IH frame.

**Pan** For the pan axis, the setup is shown on the right in figure 1. It is clear that in this case there are almost no features that are present in all 4 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. These features are represented by a dot in the right image of figure 1. 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 of Section 3.1. The features can be reconstructed in 3D in the first stereo view. They are also visible in one image of the second view, so 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. Using the relative transformation between camera and IH from Section 4.1, we have found back the relative transformation between the two stereo views.

#### 4.4 Actual calibration of pan and tilt axes

The previous section provides us with a set of relative transformations between IH frames. Part of these come from tilt motions, the other part from pan motions. Be-

cause that each of these transformations comes from a pure rotation, the rotation axes can easily be computed. Because -from the point of view of the cameras- the pan axis changes according to the tilt angle, one first has to “undo” the influence of the tilt rotation before one can use the relative transformation to compute the pan axis.

## 5 3D Terrain modeling

After the calibration of the IH is performed, the process of generating a 3D model or models of the planetary terrain can commence. This modeling is vital to accomplish 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 is the Digital Elevation Map (DEM). In this section we will describe the different steps that are performed to obtain such a DEM.

### 5.1 Generation of disparity maps

On each image pair recorded by the Imaging Head, a stereo algorithm is applied to compute the disparity maps from the left image to the right and vice versa. Disparity maps are an elegant way to describe correspondences between two images if the images are **rectified** first. This process re-maps the image pair to standard geometry with the epipolar lines coinciding with the image scan lines [5]. The correspondence search is then reduced to a matching of the image points along each image scan-line. The result (the disparity maps) is an image where the value of each pixel corresponds with the number of pixels one has to move to left or right to find the corresponding pixel in the other image. The dense correspondence scheme we employ to construct the disparity maps is the one described in [1]. It operates on rectified image pairs and incorporates some extra 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 constrain in the dynamic programming approach. The algorithm was adapted to yield sub-pixel accuracy by employing a quadratic fit of the disparities.

### 5.2 Digital Elevation Maps

A digital elevation map or DEM can be seen as a collection of points in a “top view” of the 3D terrain where each point has its own height or “elevation”. Classical approaches to generate DEMs from disparity maps or depth maps consist

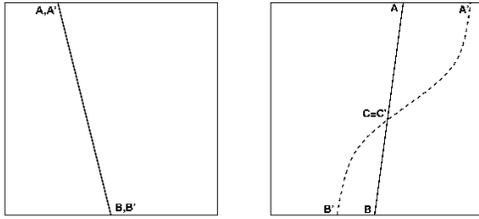


Figure 2: DEM generation in detail: left and right disparity image with the projection of the vertical line and the “shadow”

of two steps. First, for each stereo image pair the disparity images are used to construct depth images where the value of each pixel corresponds to its depth. Then a limited amount of points of each depth image is reconstructed in 3D. These points form the DEM. The problem of this scheme is that the resulting DEM is not regular in 3D.

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 because of occlusions. These occlusions are found in a very simple way.

The terrain is divided into rectangular cells. For each cell the stereo pair is selected in which the cell would be seen if it had a height of zero. A vertical line is drawn in this cell and the projection of this line in the left and right disparity image of the stereo pair is computed.

Figure 2 illustrates the algorithm that is used to find the height of the terrain on that line. The solid line  $A-B$  is the projection of the vertical line in both disparity images. Now imagine 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 from  $A'$  to  $B'$ . The part of this curve from  $A'$  to  $C'$  is the “real” part of the shadow (i.e. it would be visible on the terrain). The part from  $C'$  to  $B'$  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  $A-B$ . The intersection point  $C$  of the shadow  $A'-B'$  and the line  $A-B$  is the intersection of the vertical line and the terrain.

Occluded regions are easily detected since in this case no intersection point  $C$  exists. The height value of occluded cells can not be computed and these cells get a certain value

in the DEM which marks them as unseen. This particular scheme also 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 data of two different stereo views has to be combined.

## 6 Conclusion

The approach for calibration and 3D measurement from Martian terrain images proposed in this paper allowed for an important simplification of the design of the imaging system of the lander. Preliminary results indicate that the system will meet the requirements.

## Acknowledgments

We acknowledge support from the Belgian IUAP4/24 ‘IMechS’ project. We also wish to thank all partners of the ROBUST project for the collaboration.

## References

- [1] L. Falkenhagen, “Hierarchical Block-Based Disparity Estimation Considering Neighbourhood Constraints”. *Proc. International Workshop on SNHC and 3D Imaging*, pp. 115-122, Rhodes, Greece, 1997.
- [2] B. Fontaine, D. Termont e.a.: “Autonomous operations of a micro-rover for geo-science on Mars”, In *Proceedings of Workshop on Advanced Space Technologies for Robotics and Automation (ASTRA 2000)*, Noordwijk, December 2000.
- [3] C. Harris and M. Stephens: “A combined corner and edge detector”, In *Fourth Alvey Vision Conference*, pp. 147-151, 1988.
- [4] S. Maybank, “Theory of reconstruction from image motion”, *Springer Verlag*, 1992.
- [5] M. Pollefeys, R. Koch and L. Van Gool, “A simple and efficient rectification method for general motion”, *Proc international Conference on Computer Vision 99*, pp.496-501, Corfu (Greece), 1999.
- [6] R. Tsai, “A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf tv cameras and lenses.” *IEEE Journal of Robotics and Automation*, 3(4): pp. 324-344, 1987