A stereo vision system for support of planetary surface exploration

Maarten Vergauwen, Marc Pollefeys, Luc Van Gool
Centre for Processing of Speech and Images, K.U.Leuven,
Kasteelpark Arenberg 10, B-3001 Leuven-Heverlee, Belgium

Abstract. In this paper a system will be presented that was developed for ESA for 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 modelling of 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. Due to practical limitations it is infeasible to use a known calibration pattern for this purpose and therefore a new calibration procedure had to be developed that can work on images of the planetary environment. This automatic procedure recovers the relative orientation of the cameras and the pan- and tilt-axis. The same images are subsequently used to recover the 3D 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. The fact that the same images can be used for both calibration and 3D reconstruction is important since in general the communication bandwidth is very limited. In addition to the use for navigation and path planning, the 3D model of the terrain is also used for Virtual Reality simulation of the mission, in which case the model is texture mapped with the original images.

Keywords: Stereo vision, calibration, terrain reconstruction, planetary rover.

1 Introduction

The work described in this paper was performed in the scope of the Robust[1] project of the European Space Agency. In this project an end-to-end system is developed for a planetary exploration mission.
The *robust* system consists of three important parts.

- The planetary rover is a small and simple rover designed to carry instruments in the immediate surroundings of a lander. It is equipped with a tether cable, providing the rover with power and data connection to the lander. This allows a very high ratio instrument-mass/rover-mass (Rieder et al. 1995).
- The planetary lander contains the imaging head, an on-board computer and the control system for both rover and imaging head.
- The on-ground control system performs all operations on earth.

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, using light emitting diodes on the payload cab of the rover for the latter. It consists of a fixed stereo head,

[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.

mounted on a unit that allows for pan and tilt motions and which is approximately 1.5 meter high. The two cameras of the stereo head are fixed-focus space approved 1024x1024 CCD cameras. The stereo head has a baseline of 0.5 meter.

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 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 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. The outermost boundary of the recorded terrain lies at twenty meters 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 are sent down together with the corresponding images.

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 effects. For high accuracy equipment, such as the imaging head, a small change in these mechanical properties results in 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 parameters obtained through calibration on ground. If the camera housing were not so rigidly built and the camera intrinsics were likely to change during launch or landing, more advanced calibration algorithms would have to be used (Zisserman et al. 1995, Pollefeys et al. 1999a). Both for reasons of robustness and calibration, the cameras have a fixed-focus. To satisfy the accuracy specifications the full resolution is needed in the far range, resulting in blurred image for the short range. Traditional calibration algorithms rely on known calibration targets with well-defined spatial coordinates. In our case the only available location for these markers would be on the lander itself. However, for reasons explained above only blurred images could be obtained from markers located on the lander. Therefore, marker based calibration is not feasible. The calibration procedure that was implemented for the robust project calibrates the imaging head using images of the terrain. In fact, before using the images for reconstruction of the terrain, the calibration of the imaging head is first computed from them. Therefore, this calibration approach minimizes transmission overhead. The calibration of the extrinsic (mechanical) properties of the imaging head is split into two parts that are executed consecutively. First the relative transformation between the two cameras is computed. 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 effective 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 do this uses the concept of the essential matrix. This matrix represents the epipolar geometry between two views, taking into account the internal parameters of the cameras. 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, 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). 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. In figure 1 a pair of images of the ESTEC planetary testbed that were acquired with the imaging head is shown. A few corresponding epipolar lines are superimposed on the images. Note that the pattern of these lines is not perfectly symmetric although it should be according to the design. This illustrates the requirement for an image-based calibration procedure.

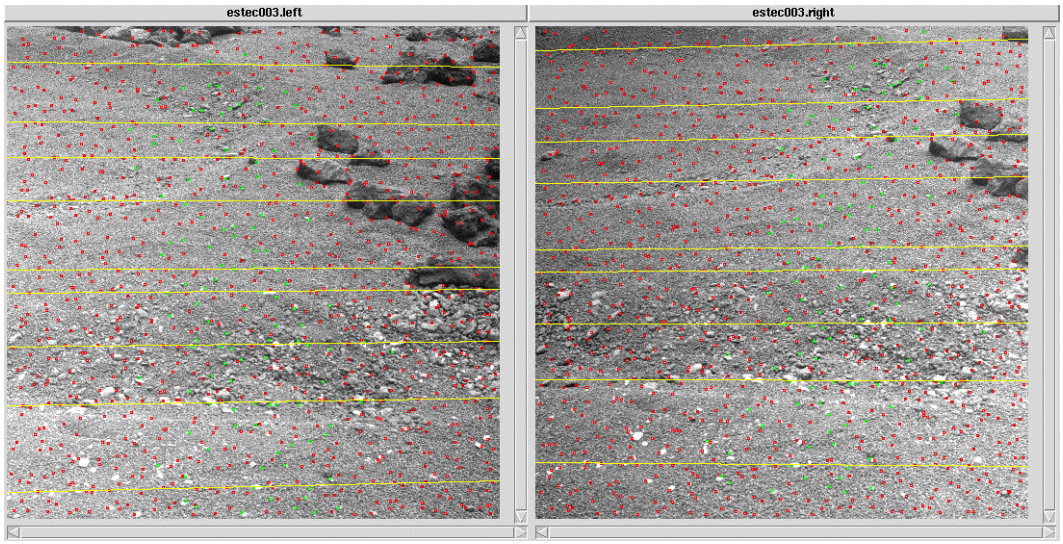

Figure 1 Epipolar geometry of an image pair

To find back the epipolar geometry between two images automatically a corner detector (see Harris and Stephens, 1987), is applied to the images. Next, a set of potential matches is obtained by comparing the neighbourhoods of features using normalized cross-correlation. These matches are typically contaminated with an important fraction of wrong matches. Therefore a robust matching scheme, called RANSAC (Fishler and Bolles, 1981), is used to compute and update epipolar geometry and matches iteratively. It is important to note that in the case of a fixed imaging head the data of the different image pairs can be combined to compute the epipolar geometry much more accurately because the relative transformation between the cameras does not change. This also solves some problems with the computation of the epipolar geometry that could occur with an (almost) planar terrain. Indeed, even for a planar terrain combining image pairs recorded at different tilt angles would correspond to observing a virtual scene containing multiple planes. Once the epipolar geometry has been computed and taking into account the known intrinsic camera parameters, the relative transformation can easily be computed (Maybank, 1992). This relative transformation is refined through a non-linear least squares algorithm that minimizes the sum of all distances between points and their corresponding epipolar lines. The absolute value of the distance between the two cameras, i.e. the baseline, cannot be computed from the images. It is unlikely, however, that this value will deviate a lot from the original value. Since all measurements for 3D reconstructions and localization will consistently be carried out using the same stereo head, this has only 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.

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, but also the pan- and the tilt-axis are needed. For the same reasons as for the relative calibration these values can change due to the stress that occurs during launch and landing. The evaluation of the pan- and the tilt-axis is more complicated, but can also be achieved from the image data, at least if some overlap exists between neighbouring image pairs. This is guaranteed in the pan direction due to the fixed vergence set-up that does not allow 100% overlap between two views of the same stereo pair (and thus yields overlap with neighbouring pairs). For the tilt axis care should be taken to foresee a sufficient overlap between at least some image pairs. To compute the different axes, the relative motion between stereo pairs is required. Matches can be obtained in a way similar to the one described in the previous section, however, in this case most features have already been reconstructed in 3D for one of the stereo pairs and a robust pose estimation algorithm (i.e. also based on RANSAC) can be used to determine the relative transformation between the stereo pairs. Once this has been done for all the overlapping pairs, the results can be combined to yield one consistent estimate of both pan and tilt axes. In this case available constraints such as the rotation angles (read from the encoders) are also enforced.

2.3 Synthetic Calibration experiment

The calibration algorithm has been tested on artificial data. A planar scene with texture from a real image from Mars was constructed and pairs of images were generated from it. First the relative calibration between the two cameras was computed. During calibration, data of 9 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 is represented as a rotation around an axis (*l, m, n*) with a certain angle θ. The ground truth was $(l_0, m_0, n_0) = (0, 1, 0)$ and $\theta_0 = 10^o$ and the computed values were $(l, m, n) = (0.000397, 0.99951, 0.00018)$ and $\theta = 10.093^o$. The angle between (l, m, n) and (l_0, m_0, n_0) was 0.0252 degrees. The difference between θ and θ_0 was 0.093 degrees. Both values are 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 degrees. The angle between the computed pan axis and the ground truth was 0.618 degrees. The larger error for the pan axis can be explained from the fact that only correspondences between three images are used to estimate it while correspondences from four images can be exploited to compute the tilt axis. During calibration of the real system, better results can be expected because much more image pairs of a non-planar scene will be used.

3 3D Terrain modelling

After the calibration of the imaging head is performed, the process of generating a 3D model or models of the planetary terrain can start. This modelling is vital to accomplish the goal of planetary exploration. Its input is all images of the terrain and the calibration of the imaging head. The output of the terrain modelling can have different forms but the most important is the Digital Elevation Map (DEM). In the following sections we will describe the different steps that are performed to obtain such a DEM.

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 between two images if the images are *rectified* first. The process of rectification re-maps the image pair to standard geometry with the epipolar lines coinciding with the image scan lines (Loop, 1999; Pollefeys et al. 1999b). The correspondence search is then reduced to a matching of the image points along each image scan-line. The result -a disparity map- 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. In addition to the epipolar geometry other constraints like preserving the order of neighbouring pixels, bi-directional uniqueness of the match and detection of occlusions can be exploited. The dense correspondence scheme we employ to construct the disparity maps is the one described in (Koch, 1996). This scheme is based on the dynamic programming scheme of Cox (Cox et al. 1996). 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.

3.2 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. The algorithm proposed for generating regular DEM´s 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 to the cameras because of occlusions. The terrain is divided into cells, i.e. 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 image of the stereo pair is computed. Figure 2 illustrates the algorithm that is used to determine the height of the terrain on that line.

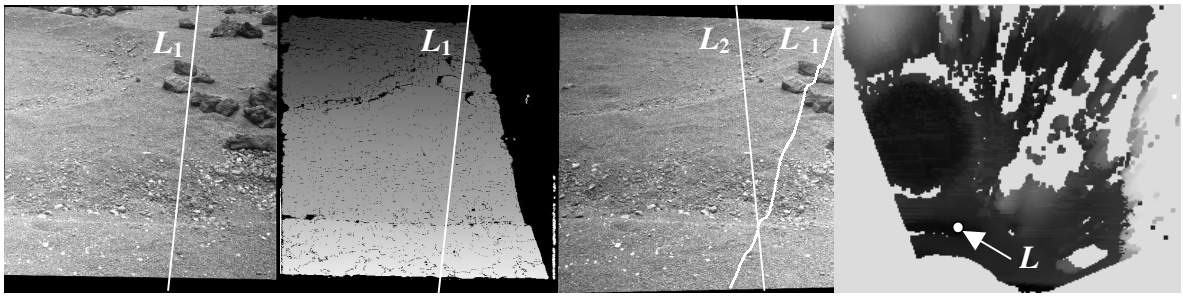

Figure 2 Digital Elevation Map generation. The figure from left to right: left rectified image, the corresponding disparity map, the right rectified image and the DEM for all pairs. The 3D line L corresponding to a point of the DEM projects to L_1 resp. L_2. Through the disparity map the shadow of L_1 on the right image can also be computed (L'_1). The intersection point of L_2 and L'_1 corresponds to the point where L intersects the surface.

4 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 by 7 meters was available. The imaging head was set up next to the testbed. Its first task was the recording of the terrain according. A mosaic of these pictures can be seen in figure 3.

Figure 3 Mosaic of images of the testbed taken by the stereo head.

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 DEM shown on the right of figure 2 was constructed. Because of the relatively low height of the imaging head (approximately 1.5 meters above the testbed) and the big rocks in the testbed, a large portion of the DEM could not be filled in because of occlusions. The DEM was then used to construct a textured triangulated mesh model. Once the reconstruction was complete, the model was used to plan a trajectory for the rover. At that time, another component of the system came into play, i.e. the simulator module. This component allows for the operator to simulate low-level commands on the rover and imaging head in a virtual reality environment. A simple gravitation model is used to predict the pose of the rover after a motion. An overview of the graphical user interface (GUI) of this component is shown in figure 5.

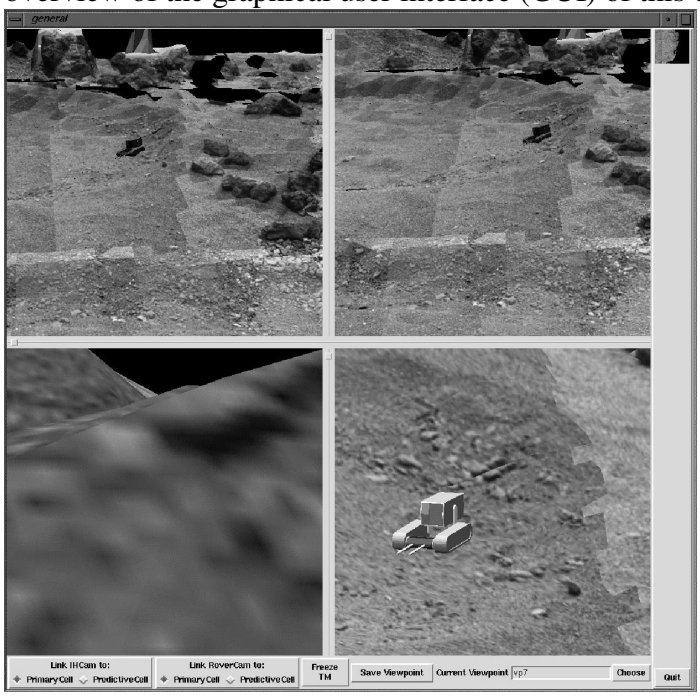

Figure 4 Snapshot of the simulator.

The GUI is divided into 4 windows. The upper two show what the left and right cameras of the stereo head are seeing. The bottom right window is a 3D interaction window and the bottom left window shows a picture of what a camera, mounted on the rover would see. When the trajectory was simulated, the commands were uploaded to the lander system that 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 LED's that are present on the rover. 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 rover.

5 Future development

For this 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 set-up 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 3D reconstruction. In (Pollefeys et al. 1998) it is described how structure and motion can be retrieved from an uncalibrated image sequence. The 3D modelling 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 section 2.1. 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 Pollefeys et al. 1999a). Finally textured triangular mesh models of the scene are constructed. During testing of the robust system on the planetary testbed at ESTEC, a test was performed. Images of some rocks on the testbed were taken by hand with a digital camera. The images were automatically processed by the 3D reconstruction system. The results can be seen in figure 5.

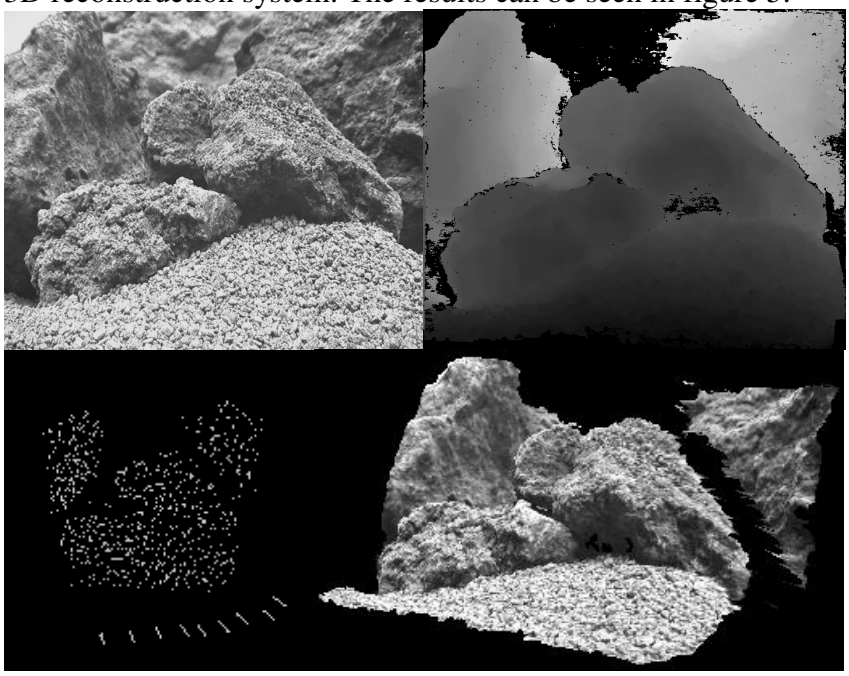

Figure 5 Results of the reconstruction process of a rock on the planetary testbed: one of the original images (upper-left), reconstructed points and cameras (lower-left), dense depth map (upper-right) and a view of the 3D reconstruction (lower-right).

The described technique 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 make images of samples and rocks on the terrain but 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 3D reconstruction. As can be seen from figure 5, the resolution of this reconstruction would be far superior to the one of the DEM, obtained from the imaging head.

6 Conclusion

In this paper an approach for calibration and 3D measurement from planetary terrain images was proposed which allowed 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 a first test in a planetary testbed at ESTEC.

Acknowledgments

Marc Pollefeys is a post-doctoral researcher of the fund for scientific research-Flanders. 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. I. Cox, S. Hingorani and S. Rao. A Maximum Likelihood Stereo Algorithm. Computer Vision and Image Understanding. Vol. 63, No. 3, May 1996.
2. M. Fischler and R. Bolles. RANdom SAmpling Consensus: a paradigm for model fitting with application to image analysis and automated cartography, Commun. Assoc. Comp. Mach. 24:381-95. 1981.
3. C. Harris and M. Stephens. A combined corner and edge detector. Proc. Fourth Alvey Vision Conference. pp.147-151, 1988.
4. J. Knight and I. Reid. Self-calibration of a stereo-rig in a planar scene by data combination, Proc. International Conference on Pattern Recognition (ICPR 2000), pp. 411-414. Barcelona. 2000.
5. R. Koch. Automatische Oberflachenmodellierung starrer dreidimensionaler Objekte aus stereoskopischen Rundum-Ansichten. PhD thesis. University of Hannover. Germany. 1996 also published as Fortschritte-Berichte VDI. Reihe 10. Nr.499. VDI Verlag. 1997.
6. S. Maybank, Theory of reconstruction from image motion, Springer-Verlag. 1992.
7. M. Pollefeys, R. Koch and L. Van Gool. Self-Calibration and Metric Reconstruction in spite of Varying and Unknown Internal Camera Parameters", International Journal of Computer Vision, 32(1). 7-25, 1999.
8. M. Pollefeys, R. Koch and L. Van Gool, A simple and efficient rectification method for general motion, Proc.ICCV'99 (international Conference on Computer Vision), pp.496-501, Corfu (Greece), 1999.
9. M. Pollefeys, R. Koch, M. Vergauwen and L. Van Gool, Metric 3D Surface Reconstruction from Uncalibrated Image Sequences, Proc. SMILE Workshop (post-ECCV'98), LNCS 1506, pp.138-153, Springer-Verlag, 1998.
10. R. Rieder, H. Wanke, H. v. Hoerner. e.a. 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, Lunar and Planetary Science. XXVI, pp.1261-1262, 1995.
11. C. Loop and Z. Zhang. Computing Rectifying Homographies for Stereo Vision. IEEE Conf. Computer Vision and Pattern Recognition (CVPR'99), Colorado, June 1999.
12. A. Zisserman, P. Beardsley and I. Reid, Metric calibration of a stereo rig, Proc. IEEE Workshop on Representation of Visual Scenes, Cambridge, pp.93-100, 1995.