Towards Cooperative Air/Ground Robotics: Issues Related to Environment Modeling

Simon Lacroix, Il-Kyun Jung, Anthony Mallet and Raja Chatila
LAAS / CNRS
7, Ave du Colonel Roche
F-31077 TOULOUSE Cedex
FRANCE

Abstract

In this paper, we consider the issues raised by the building of a consistent unique environment model on the basis of aerial and ground data. The ability to build and maintain such a model is a key feature to develop effective cooperation schemes between air and ground robots, and makes a lot of sense for planetary exploration robotics. A brief discussion of the problems and possible solutions is given: digital elevation maps (DEMs) appear to be a well suited representation to integrate both aerial and ground data. Some preliminary results related to the building of a fine resolution DEM on the sole basis of a set of aerial stereovision images are presented.

1 Introduction

In all the applications contexts where the development of exploration and intervention robotics is considered, air/ground robotic ensembles bring forth several opportunities from an operational point of view. Be it for planetary exploration, but also for environment monitoring and surveillance, demining or reconnaissance, a variety of scenarios can be envisaged. For instance, orbiters or aircrafts can operate in a preliminary phase, in order to gather informations that will be used in both mission planning and execution for the ground vehicles. But one can also foresee cooperative scenarios, where aircrafts would support ground robots with communication links and global informations, or where both kinds of machines would cooperate to fulfill a given mission.

Whatever the cooperation scenario, we believe that to foster the development of such ensembles, one of the most important issue to address is the building and management of environment representations using data provided by all possible sources. Indeed, not only each kind of machine (terrestrial, aerial) can benefit from the information gathered by the other, but also in order to plan and execute cooperative or coordinated actions, both kinds of machines must be able to build, manipulate and to reason on shared coherent environment representations.

In a planetary exploration context, studies and projects showed that free floating balloons (where the sole controlled parameter is altitude) could be considered in a Martian atmosphere [1]. Similar devices have also been envisaged for other planets [2]. For mapping and exploration missions, the use of orbiter or aircraft imaging prior or during ground rovers mission would definitely improve efficiency. Indeed, if global consistent informations on the environment are necessary for the operators to define the rover missions, they can also be used by the rover *during* its traverse, to define navigation strategies for instance [3], or to guaranty an error bounded position estimate during the traverses. Of course, such benefits are all the more interesting when considering *long range* missions, in terms of time and surface explored.

Aircrafts would in such cases rather play an assistance role, the core of the mission (fine mapping or sample collection for instance) being achieved by the rover. Addressing the problem of building common environment representations between aerial and grounds rovers is also very useful when considering only ground rovers, since it provides the rovers with the ability to consider informations gathered by an orbiter, during descent [4], or acquired by sensors supported by balloons tethered to the rovers.

In the next section, we discuss some issues related to environment modeling using both aerial and ground sensors, and sketch some possible solutions. Digital elevation maps appear to be a well suited representation to tackle most of these issues, especially in planetary environments, that are unstructured. After an introduction to these representations (section 3), section 4 and 5 present algorithms to build such maps, respectively with ground and low altitude stereovision data. A discussion concludes the paper.

2 Issues Related to Cooperative Environment Modeling

Basically, the problem is to integrate the information acquired from the rovers and aircrafts into a global coherent representation. For that purpose, one must define algorithms and representations that support the following characteristics of the data:

- Data resolution: the resolution of the gathered information can significantly change, depending on whether it has been acquired by a ground or an aerial sensor.

- Uncertainties: similarly, there are several orders of magnitude of variation on data uncertainties between ground an aerial data.
- Viewpoints changes: beside the resolution and uncertainties properties of the sensors that can influence the detection of specific environment features, the difference of viewpoints between ground and aerial sensors generates occlusions that considerably change the effectively perceived area, and therefore the detectable features.

Depending on the considered context, these difficulties may vary: for instance, for aerial sensors, occlusions caused by overheads are very unlikely to occur in planetary contexts, whereas the presence of vegetation or buildings on Earth make some area unperceivable from air. Also, the aircraft altitude of course strongly influences the properties of the perceived data in terms of precision and resolution.

But before tackling the problem of integrating ground and aerial data into a global consistent environment representation, robust to these variations, one must consider the *registration* issue: in order to merge these informations, a way to establish the spatial correspondence between the two kinds of data is required. This would be straightforwardly solved if sensor positions were always precisely known, which is far from being a realistic hypothesis.

Designing algorithms that solve the registration problem is therefore a key prerequisite to address cooperative environment modeling. Moreover, such algorithms can help to localize the rover with respect to initial maps of the environment, as provided by an orbiter for instance[1]: the consideration of this problem, illustrated in figure 1 is also very relevant when considering only ground rovers.

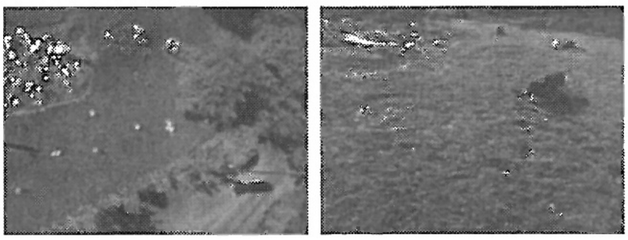

Figure 1: *An aerial image of our experimentation site (left), and a view taken from the ground. From which position in the left image was the right image taken ?*

Several kinds of solutions can help to solve this registration problem :

- Feature detection and recognition: the ability to detect, localize and associate features extracted from both sources of information is certainly a good way to proceed. The difficulty here lies in the conception of algorithms able to detect features in multi-scale data: geometric features can be directly matched between the two kinds of data only if the data resolution is comparable. The detection and identification of higher level features, *i.e.* not only geometric or topographic, but *objects* with a semantic interpretation, can be more useful, especially when there is a big resolution change in the data. Indeed, one can conceive independent semantically driven algorithms for the two sources of data that are able to detect and recognize the same kinds of objects.
- Region matching: most of the algorithms that structure and interpret aerial data produce a description in terms of regions with a semantic interpretation. Matching such regions with regions detected on the environment model built by ground rovers is a solution to consider, even though it will probably provide only qualitative positioning solutions.
- Skyline based approaches: several contributions describe algorithms that match the skyline perceived from the ground with skylines derived from an initial digital elevation map of the environment (*e.g.* [5]). This requires the initial geometric knowledge of a very large area around the rover in order to predict the skyline that will be perceived, the possibility to extract the skyline from images perceived by the rovers (which is not so easy in terrestrial environments), and the presence of distinguishable features on the horizon line (which is not the case in most of planetary environments - *cf* the landscape images taken during the Viking and PathFinder missions).

3 Digital Elevation Maps

Digital elevation maps (DEMs) that represent the ground surface as a function $z = f(x, y)$ defined on a regular Cartesian grid are a very popular way to to model unstructured terrains (*e.g.* see [6, 7, 8]). Indeed, such models are very convenient to manipulate, and are able to catch the gist of the perceived surface geometry: several feature detection algorithms can rely on such maps [9, 10]. The registration of this kind of models can also be performed using particular signatures that do not correspond to explicit geometric features (*e.g.* the *spin-images* in [11]), or thanks to point matching algorithms. Registration of such map in a multi-scale context has been considered in [12].

Range data. Dense 3D data are required to build a digital elevation map. For ground rovers in the context of planetary exploration, stereovision has various advantages over laser range finders, *e.g.* from the point of view of weight and energy consumption. Range data acquired from aerial devices can be produced by Lidar or Radar altimetry, but vision sensors are the ones that produce the finest resolution data. For very low altitude aircrafts, one can also consider stereovision, which directly produces dense 3D data: a stereo bench baseline of about 1 meter is large enough to

[1] A problem often referred to as the *drop-off problem*.

get 3D data up to one hundred meters depth. One of the advantage of stereovision is that to each 3D point produced by stereovision is associated a pixel in the video images: 3D and video data are therefore directly registered, which can be useful to detect features on the ground or to estimate the motion between stereovision frames.

Our stereovision algorithm is a classical pixel correlation algorithm: a dense disparity image is produced from a pair of images thanks to a correlation-based pixel matching algorithm (we use either the ZNCC criteria or the census matching criteria), false matches are avoided thanks to a reverse correlation. With good quality images, the reverse correlation phase is enough to remove the few false matches that occur. But when the images are not of very good quality, three thresholds defined on the correlation score curve are used to discard false matches (on the value of the highest score, on the difference between this score and the second highest peak in the curve, and on the "sharpness" of the peak).

Ground Data Versus Aerial Data. The problem of building digital elevation maps from 3D data essentially stems from uncertainties in the data (that grows quadratically with the distance), and to the fact that data resolution decreases drastically with the distance for ground sensors (figure 2). However, the problem is better conditioned when using aerial images looking downwards: not only data resolution on the ground is more regular, but also uncertainties on the 3D data "fit" better a representation of the uncertainties in the digital map.

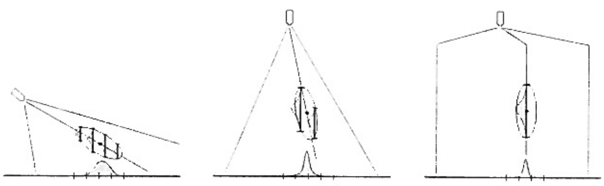

Figure 2: *Sensor errors with respect to an horizontal grid. In the leftmost case, which is the one encountered with range images acquired from a rover, the precision of the data, which is essentially in the range estimate, should be transformed into an occupancy probability in the DEM cells. In the rightmost case, which correspond to data acquired from a flying device, the problem is much better conditioned: the precision on the data can be directly represented by a precision on the elevation computed for the cells. The middle case corresponds to a very low altitude acquisition, which is the case of our aerial images: in our algorithm, we however consider that we are close to the rightmost case. One can also guess on these figures that the variable resolution of the data on the ground play an important role for ground rovers, and is much more regular with aerial images.*

Building digital elevation maps. Although there has been several contributions to the problem of building digital elevation maps with data acquired from rovers, we think that it has still not been addressed in a satisfactory way. The main difficulty comes from the uncertainties on the 3D input data, that can hardly be propagated throughout the computations and represented in the grid structure with a simple mean/standard deviation couple. To take into account these data properties, the best representation would therefore be a 3D occupancy grid [13].

Currently, our algorithm to build a digital elevation map comes down to computing the elevation of a cell by averaging the elevations of the 3D points that are projected into it. The standard deviation on the elevation is also straightforwardly computed, and since to each 3D point is associated a luminance value, it is also possible to compute a mean luminance value for each map cell. We made some comparisons of this simple algorithm with the building of 3D occupation grids, using a Bayesian and a Dempster-Shafer approach to update the probabilities: given the density of the data, the simple approach gives very similar results.

One of the intrinsic drawback of representing terrains with DEMs is that they consider that the perceived surface is convex: there is no way to represent vertical features and overheads with a single elevation value for each cell. We are currently considering a much more sophisticated algorithm, akin to 3D occupancy grids, that maintains a discrete probability density function for the cell elevations.

4 Building DEMs with Sensors on Board a Rover

One of the problems when dealing with rough terrains is the large number of occlusions caused by terrain irregularities and the decrease of data resolution on the ground (figure 3). To palliate this, the DEM is continuously updated as the robot navigates, which requires knowing the robot 3D position with a precision of the order of the grid size. Odometry being not reliable on rough terrains, in our system the 3D position estimate is provided by a visual motion estimation technique [14], which is running in parallel with the navigation loop.

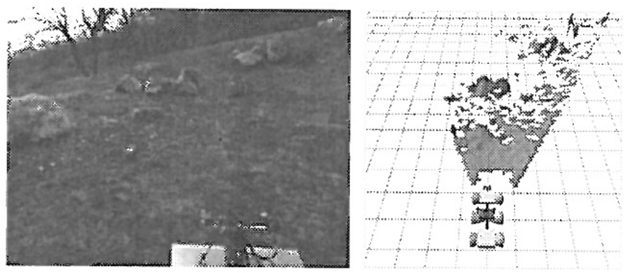

Figure 3: *A digital elevation map built with a single stereo image pair: several areas are unperceived because of terrain irregularities and the drastic decrease of data resolution with distance.*

The technique is able to estimate the 6 parameters of the robot displacements in any kind of environments, provided it is textured enough so that pixel-based stereovision works well: the presence of no particular landmark is required. The motion parameters between two stereo frames are computed on the basis of a set of 3D point to 3D point matches, established by tracking the corresponding pixels in the image sequence acquired while the robot moves (figure 4 - as in stereovision, pixels are tracked using either the

ZNCC or the census matching criteria). A lot of attention is paid to the selection of the pixel to track: on one hand, in order to avoid wrong correspondences, one must make sure that they can be faithfully tracked. This is done thanks to the application of an autocorrelation function on the image, that gives for every pixel an indicator related to the expected *precision* of the tracking algorithm. On the other hand, in order to have a precise estimation of the motion, one must choose pixels whose corresponding 3D point is known with a good accuracy. For that purpose, we use an error model of the stereovision algorithm, that consider the sharpness of the correlation curve as an indicator of the precision of the computed 3D coordinates [15].

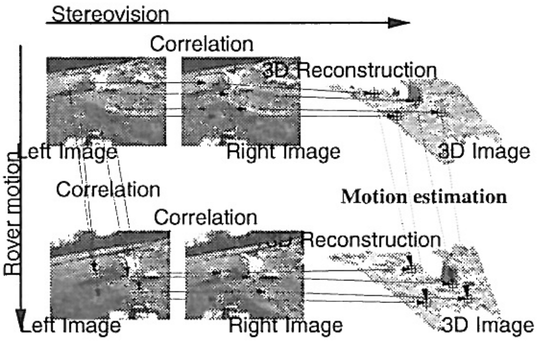

Figure 4: *Principle of the visual motion estimation technique. The steps of stereovision go from left to right, time goes from top to bottom. Given a stereovision image pair, a set of pixels are selected on the left image. They are tracked and matched in the new stereovision frame: this produces 3D points associations (right - the 3D images are represented as DEMs for readability purposes), from which the motion is estimated.*

Figure 5 presents a DEM resulting from the integration of 50 stereovision pairs. An important feature of DEMs built from stereovision data is the ability to associate a luminance value to every perceived DEM cell (figure 6.)

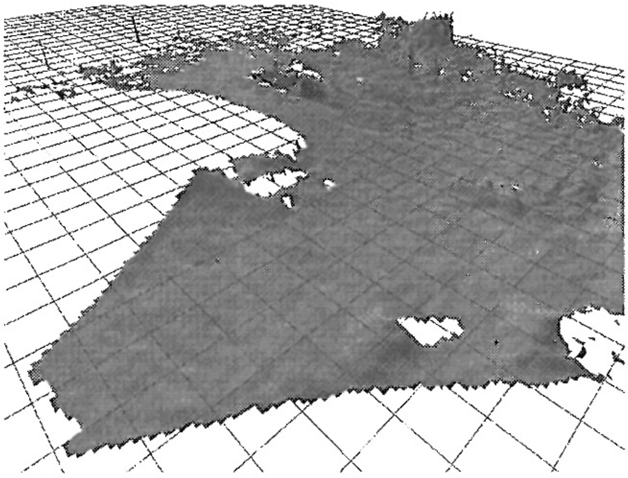

Figure 5: *A digital elevation map built during a 20 meter-long run, using 50 stereovision pairs. The displayed grid is* $1m \times 1m$, *the actual DEM grid is* $0.1m \times 0.1m$.

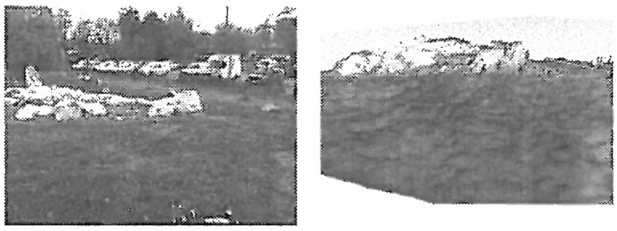

Figure 6: *The right image is a view of the digital elevation map built using several images of the scene showed in the left image.*

5 Building DEMs with Low Altitude Imagery

We are currently equipping a small blimp [16] with sensors (GPS, inclinometers and a stereo bench) and on board computing devices. We started to study the dynamics of airships in order to determine motion control laws, and began to acquire some stereovision data to tackle environment modeling issues. We present here how we could build a fine resolution spatially consistent map on the sole basis of stereo images (more details can be found in [17]).

Experimental Setup. To acquire low altitude data, we equipped a $25m^3$ tethered blimp with a stereo bench (figure 7). The stereo bench is composed of two black and white 752×582 pixels CCD cameras; the baseline is approximately 1.20 meters. Calibration of the bench has been made possible with a set of images pairs of a calibration frame laid on the ground. A few hundreds of stereo pairs have been acquired over an approximately $3000m^2$ surface, at altitudes varying from 10 to 40 meters, during three different acquisition runs. The surface over which the blimp flew contains elements of various natures: rocky areas, grass areas, fences, tall trees and bushes, a small building and a parking lot (figure 7).

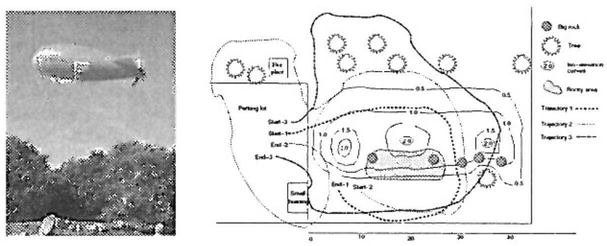

Figure 7: *The tethered balloon we used to get low altitude stereo images (left), and the three acquisition runs drawn on a sketch map of the area (right)*

Figure 8 presents a result of the stereovision algorithm on a pair of images acquired from the balloon.

Motion Estimation. Since the blimp is not equipped with any positioning device, one must determine the relative positions of the system between successive stereo

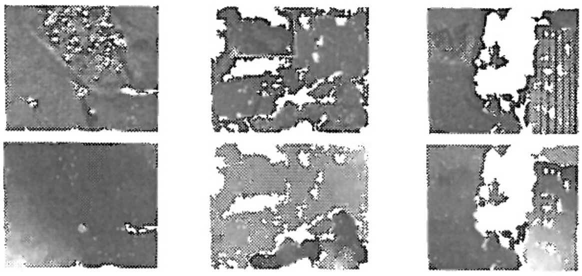

Figure 8: *Results of the stereovision algorithm on three example images. Original images where non-matched pixels are white are on the top line, disparity images are on the bottom line. The disparity is here coded with grey levels: the darker the pixels, the closer they are to the cameras. The left result is an ideal case, where most of the pixels are matched. The middle and right examples show that the stereovision algorithm hardly find matches in the shadowed areas.*

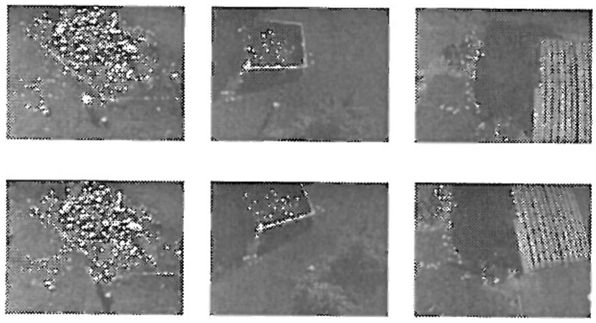

Figure 9: *Three results of the interest point matching algorithm, applied between the example images of figure 8 and the following ones in the acquisition sequence. The white squares indicate the interest points that were matched. Note that the viewpoint changes in the middle and right images are quite important: nevertheless, no erroneous matches were established.*

frames in order to build a global digital elevation map. For that purpose, we adapted the motion estimation algorithm initially developed for ground rovers [14]. But when applied with a ground rover, the tracking phase is initiated using a first estimate of the motion provided by odometry. This is not possible with the blimp images, as no initial motion estimate is provided: it is not possible to focus the search area in the images for the pixel tracking.

This phase is therefore replaced by an algorithm that matches interest points detected on a pair of grey level images taken from arbitrary points of view. First matching hypotheses are generated using a similarity measure of the interest points. Hypotheses are confirmed using local groups of interest points: group matches are based on a measure defined on an affine transformation estimate and a on correlation coefficient computed on the intensity of the interest points that are consistent with the estimated affine transformation. Once a reliable match has been determined for a given interest point and the corresponding local group, new group matches are found by propagating the estimated affine transformation (more details on this algorithm can be found in [18]).

The algorithm provides dense matches and is very robust to outliers, i.e. interest points generated by noise or present in only one image because of occultations or non overlap.

Figure 9 shows some results of the interest points matching algorithm on some images acquired with the blimp: not surprisingly, no interest points are matched in the shadow and low texture areas. This is essentially due to the poor image quality, that generate unstable corners in such areas. Fortunately, the algorithm is extremely robust[2], and is not affected by this.

First results. Thanks to these three algorithms (stereovision, motion estimation and map building), we could build various digital elevations maps integrating several tens of stereo images. In the absence of any attitude estimation device on-board the blimp, the plane of the DEM is defined orthogonally to the view axis of the first stereovision images considered.

Figure 10 presents a digital elevation map built with 120 stereovision pairs, covering about $1500m^2$, with a cell resolution of $5 \times 5cm^2$. The trajectory executed by the blimp, which could be recovered using the localization algorithm, is an about $100m$ long loop. The last images overlaps the first images, and no discrepancies can be observed in the final model on this area: the localization algorithm gave here an *extremely precise position estimate*, with a final error of the order a map cell size, *i.e.* about 0.1%. Figure 11 show two maps built with images corresponding to the second trajectory, and figure 12 show the map built with all the images of the third trajectory.: the position estimation in this latter case drifted of about 1%.

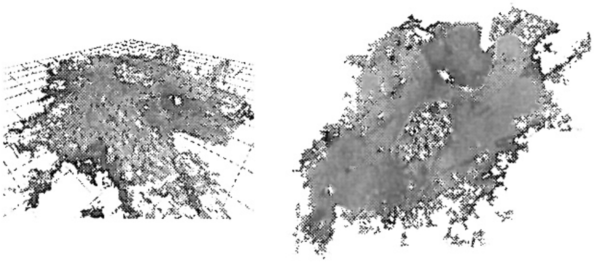

Figure 10: *Final digital elevation map produced with the 120 stereovision images corresponding to the first trajectory in the sketch map of figure 7. Left: 3D model, in which some peaks are artifacts due to the presence of the moving operator. Right: a top view of this model, which is actually an ortho-image of the terrain. The "vertical" projection of the blimp positions are shown as small black frames in this image.*

Note that one can distinguish in all this figures the various positions of the operator, that produces peaks in the elevation map and whitish areas in the luminance values (the operator wore a white shirt). This is because the algorithm that fills the map simply updates the cells without any consistency check. Such artifacts can be easily removed by checking if the elevation of the points of the 3D image to fuse is consistent with the current elevation and corre-

[2]It has been tested with hundreds of images of various 3D scenes, taken in various conditions and environments.

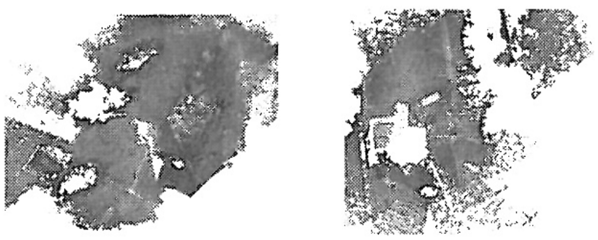

Figure 11: *The two ortho-images corresponding to DEMs built from images of the second trajectory in figure 7.*

Figure 12: *Final digital elevation map produced with the 80 stereovision images corresponding to the third trajectory of figure 7. The overall position estimation error can be seen on top of the right image, where the first and the last images overlap (see the calibration frame). The absolute translation error is about* $1m$ *at the end of the trajectory.*

sponding standard deviation of the map cells. Unconsistent 3D points correspond to moving parts in the environment, and can just be discarded.

All the images were processed off-line (no CPU was onboard the blimp). However, the computations times are compatible with an on-line implementation, most of the time being consumed by stereovision and interest points matching, which both take about one second on an Ultra-10 Sparc CPU. A very interesting point is that only two cameras are required do build such a map and to recover the blimp motions: there is no need to carry any positioning device. This is an important point when considering planetary exploration contexts, where the payload/volume ratio is much worse than within a dense atmosphere.

6 Discussion

These first results show the feasibility of building fine resolution elevation maps with a good resolution from aerial images solely from stereo data. Images from higher altitudes could also provide good resolution of the DEM if better resolution cameras were used, in addition to subsampling and super resolution processing.

The next step is to match the DEM built from the aerial imagery with the DEM built from the ground. Important to note is that both DEMs have similar resolutions. We are currently developing and testing several methods using features detected on both maps or iconic-based approaches with correlation of the elevations in both maps. Luminance being also provided in the two cases, the correlation algorithms can use this information together with the elevation.

References

[1] A. Vargas, J. Evrard, and P. Mauroy. Mars 96 aerostat: an overview of technology developments and testing. In *13th AIAA Lighter Than Air Systems Technology Conference, San Francisco, CA (USA)*, pages 29–36, June 1997.

[2] A.G DiCicco, K.T. Knock, and G.E. Powell. Balloon experiments at venus (bev). In *11th AIAA Lighter Than Air Systems Technology Conference, Clearwater Beach, Fl. (USA)*, pages 144–154, May 1995.

[3] R. Chatila, S. Lacroix, T. Siméon, and M. Herrb. Planetary exploration by a mobile robot : Mission teleprogramming and autonomous navigation. *Autonomous Robots Journal*, 2(4):333–344, 1995.

[4] L. Matthies, C. Olson, G. Tharp, and S. Laubagh. Visual localization methods for mars rovers using lander, rover and descent imagery. In *International Symposium on Artificial Intelligence, Robotics and Automation in Space, Tokyo (Japan)*, pages 413–418, July 1997.

[5] F. Cozman and E. Krotkov. Automatic mountain detection and pose estimation for teleoperation of lunar rovers. In *International Symposium on Experimental Robotics, Barcelona (Spain)*, pages 164–172, June 1997.

[6] M. Hebert, C.Caillas, E. Krotkov, I.S. Kweon, and T.Kanade. Terrain Mapping for a Roving Planetary Explorer. In *IEEE International Conference on Robotics and Automation, Scottsdale, Az. (USA)*, pages 997–1002, 1989.

[7] I.S. Kweon and T. Kanade. High-resolution terrain map from multiple sensor data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):278–292, Feb. 1992.

[8] C. Olson. Mobile robot self-localization by iconic matching of range maps. In *8th International Conference on Advanced Robotics, Monterey, Ca (USA)*, pages 447–452, July 1997.

[9] I.S. Kweon and T. Kanade. Extracting topographic features for outdoor mobile robots. In *IEEE International Conference on Robotics and Automation, Sacramento, Ca (USA)*, pages 1992–1997, April 1991.

[10] P. Fillatreau and M. Devy. Localization of an autonomous mobile robot from 3d depth images using heterogeneous features. In *International Workshop on Intelligent Robots and Systems, Yokohama (Japan)*, 1993.

[11] D. Huber and M. Hebert. A new approach to 3-d terrain mapping. In *IEEE/RSJ International Conference on Intelligent Robotics and Systems (IROS '99)*, pages 1121–1127, 1999.

[12] D. B. Gennery. Visual terrain matching for a mars rover. In *IEEE International Conference on Computer Vision and Pattern Recognition, San Diego, California (USA)*, pages 483–491, 1989.

[13] A.P. Tirumalai, B.G. Schunck, and R.C. Jain. Evidential reasoning for building environment maps. *IEEE Transactions on Systems, Man and Cybernetics*, 25(1):10–20, Jan. 1995.

[14] A. Mallet, S. Lacroix, and L. Gallo. Position estimation in outdoor environments using pixel tracking and stereovision. In *IEEE International Conference on Robotics and Automation, San Francisco, Ca (USA)*, pages 3519–3524, April 2000.

[15] S. Gautama, S. Lacroix, and M. Devy. On the performance of stereo matching algorithms. In *Workshop on Vision, Modelling and Visualization, Erlangen (Germany)*, Nov. 1999.

[16] S. Lacroix. Toward autonomous airships: research and developments at laas/cnrs. In *3rd International Airship Convention and Exhibition, Friedrichshafen (Germany)*, July 2000.

[17] S. Lacroix, I-K. Jung, and A. Mallet. Digital elevation map building with low altitude stereo imagery. In *9th International Symposium on Intelligent Robotic Systems, Toulouse (France)*, July 2001.

[18] I-K. Jung and S. Lacroix. A robust interest point matching algorithm. In *8th International Conference on Computer Vision, Vancouver (Canada)*, July 2001.