0. ABSTRACT

In this paper, we present the methodology and infrastructure test bed developed as well as the activities conducted by CNES for the validation of the perception hardware and stereo-vision algorithms proposed for Exomars in the framework of our support to ESA.

1. INTRODUCTION

The reduced communication opportunities and the long signal propagation times between Earth and planet Mars make the rover tele-operation incompatible with the long daily traverses (approximately 100m per sol) required by the exploration of sites of scientific interest. The Exomars rover has therefore to be highly autonomous in the reaching of the target destinations designated by the scientists.

The autonomous navigation integrates a path planning function that computes a target trajectory from its knowledge of the rover's surroundings that are internally represented as a digital elevation map. To create the 3D model of the Martian ground, a stereo-bench, mounted on the rover mast, takes image pairs that are fed to a dense stereo-vision algorithm.

The necessity for the path planning function to choose trajectories that guarantee the permanent security of the rover leads to a security-efficiency dilemma in the definition of the security margins that have to be observed around potential obstacles. The search for a good compromise between rover security and optimal path to the target requires to have good confidence in the computed 3D model of the ground. An in-depth validation of the stereo-vision performances has therefore to be done prior to define the security margins.

In the Exomars project, CNES (the French Space Agency), acts as a technical support to ESA and has made available to the Project the results of its past developments on the Mars96 and MSR missions as well as of its R&D on rovers: stereo-bench design, stereo-vision and autonomous navigation algorithms, test facilities...

2. MARTIAN GROUND 3D MODELING

Stereo-vision is a well suited 3D modelling solution for robotic planetary exploration missions thanks to its low mass, low volume, low energy consumption and the relative ease of qualification for the space environment (wide temperature ranges, vibrations during launch and exploitation, chocks at landing, space radiations...). The acquisition process being fast, dense 3D models (millions of points) can be reconstructed and for some applications (eg visual motion estimation) it could be done while the rover is moving.

2.1 Vision Sensor: The Stereo-bench

Fig. 1 presents the prototype of a flight model stereo-bench studied by CNES. It consists of a pair of space-qualified 1024x1024 pixels cameras separated by a titanium plate (stereo base of 100mm), specifically designed for thermoelastic stability.

2.2 Vision Algorithms

The 3D model reconstruction is computed by texture correlation [1] to produce a high density depths map.

An initial optical calibration procedure [2] corrects the effects of geometric distortion, places the images of a given point of the landscape on the same row of the left

Figure 1: Stereo-bench
an right CCD and insures that the optical axes of the pinhole models of the two cameras with their optics are parallel.

The viewpoints difference between the two cameras (set apart by the stereo base) results in getting the image of an object point on different columns on the left and right images.

The interval between the two columns, called disparity $d_i$, is related to the distance $D_i$ to the object, the stereo base $b$ and the focal length $f$ as follows:

$$d_i = \frac{f \cdot b}{D_i}$$

For every pixel in the left image, the homologous pixel is searched on the corresponding row of the right image by comparison of the neighbourhood of the pixels (correlation windows). The column of the matching pixel defines the disparity value. A sub-pixel interpolation is then computed to break the pixel barrier and obtain decimal values instead of integer ones.

Computing the disparity for every pixel of the left image that have an homologous pixel in the right image results in an image of disparities.

A faster multi-resolution algorithm has also been developed by CNES and is currently under validation.

The stereo-correlation algorithm can either be implemented in software by the on-board computer or it can be effectively hardwired in an FPGA.

3. HIGH-DENSITY 3D MODEL VALIDATION

A manual validation procedure, based on the comparison of a few distances in the 3D model to their counterparts measured with tape measure in the physical scene is not suitable for Exomars as the 3D model produced by the perception chain may contain up to one million points. In addition, the visual inspection of the disparity maps might not be reliable enough to detect subtle reconstruction artefacts.

A global validation method was therefore developed. It is based on the comparison of the “real” 3D model (generated by stereo-vision) to a “virtual” 3D reference model (computed by a “virtual stereo-bench” from the high-density cloud of points, of the same scene, measured by a Laser Scanner). Hardware and software means were developed for the acquisition and the exploitation of the validation data.

4. VALIDATION MGSE

To lower the impact of external factors on the validation process, a specific MGSE (Mechanical Ground Support Equipment) was designed and built. It is mainly composed of a stereo-bench, a Laser Scanner (LS) and translation stages (used to reduce the parallax between the two acquisitions).

The LS measures a high density cloud of points of the validation scene with an accuracy of circa ±2mm. For the field of view covered by the stereo-bench, at full resolution, up to 30 million 3D points can be measured.
One motorized stage permits to swap between the LS and the stereo-bench acquisitions. Three manual translation stages are used during the calibration process of the MGSE to accurately place the viewpoints of the two instruments at the same 3D spatial position.

4.1 MGSE Configuration

The MGSE can be equipped with different mechanical wedges designed to set the pitch of the stereo-bench according to the region of interest in the scene: flush with the rover's wheels (ie at short distance to the subject) or a few meters in front of them. In case of large field of view optics, a single wedge could suffice.

While the Laser Scanner has a single viewpoint (the centre of the mirror), the two viewpoints of the stereo-bench (corresponding to the left and right cameras) lead to specific occlusions issues (Fig. 5).

As the real disparity is computed for the stereo-bench left viewpoint, placing the Laser Scanner on the left camera allows to maximize the quantity of available data in the left re-projected virtual disparity image. However, for the specific study of occlusions, it may be interesting to place the Laser Scanner on the right stereo-bench viewpoint.

4.2 MGSE Calibration

To minimize the effects of measurement uncertainties on the estimated performances of the stereo-vision system, a Laser Tracker that allows the measurement of an optical target's 3D position with sub-millimetric accuracy is used during the calibration process. Once the MGSE is moved to the target scene, the first step consists to measure the elements of the transfer matrix (rotations and translations) between the Laser Scanner and the Laser Tracker frames. For that, landmark balls are placed in the scene and the positions of their centres are measured with both instruments. The transfer matrix between the two frames is then computed by least squares fitting.

The position of the Laser Scanner in the Laser Tracker frame being known, it is possible to reduce the parallax between the acquisitions of the stereo-bench and the Laser Scanner. The parallax reduction is an iterative process that involves the measurement of the stereo-bench position and attitude (the stereo-bench is equipped with measurement mechanical reference elements) and the adjustment of the translation stages until the selected viewpoint is placed in the same position as was the centre of the Laser Scanner frame. The 3D position of the viewpoint is estimated by stacking the mechanical dimensions of the stereo-bench and the optical characteristics of the lenses (to minimize the impact of the mechanical tolerances stack-up, the stereo-bench mechanical parts were produced by high-precision machining). To lessen the number of iterations, the target position of the stereo-bench is estimated and displayed to the operator.

5. 3D ACQUISITIONS CAMPAIGNS

CNES has two Mars yards in Toulouse that can be arranged for 3D acquisitions.

The indoor site contains pozzolana-textured geometric objects that can be arranged on the ground to create various configurations that might be challenging for a 3D vision system.

For the validation work, a dozen indoor scenes were acquired at half the maximum resolution of the Laser Scanner (ie circa 7 LS points per stereo-bench pixel). On some scenes pozzolana rocks were added to the geometric objects.
The SEROM outdoor site represents a “realistic” Martian landscape composed of pozzolana grains and rocks of different sizes. While the indoor site has steady and repeatable illumination conditions, at the SEROM it was possible to test different lighting configurations (and the impact of shadows) depending on the position of the Sun.

A mechanical noise inside the Laser Scanner required to halve the resolution for the outdoors campaign resulting in only circa 1.75 LS points per stereo-bench pixel. Five outdoors scenes were still acquired.

The calibration procedure being time-consuming, changing the content of the scene (eg adding or moving rocks) was preferred to moving the MGSE to different places and recalibrating. That solution also permitted to monitor the evolution of the global performances wrt the scene content (starting with bare ground and step-by-step adding stones of increasing sizes).

To study the robustness of the stereo-vision algorithms with regard to image exposition, under-exposed and over-exposed images of the same scene were acquired.

6. 3D DATA FILTERING

To be exploited, the clouds of points acquired by the Laser Scanner have to be filtered: the 3D points located out of the interest zone (corresponding to the stereo-bench field of view) and the outliers that often appear close to the edges of the objects are discarded so that they don’t impact the estimation of the vision system performance.

The 3D-Filter software was developed to give the user the possibility to use a 3D tool (sphere or cube) to select in 3D the points that have to be either invalidated or filtered with a specified criterion (density of point or minimum distance).

Several displacement modes were implemented for both the virtual camera and the 3D tool to allow the user to visualize and accurately select the volumes to be processed. The critical functions are multi-threaded to take advantage of the multi-core processors and dramatically speed-up the processing of the tens of millions of 3D points measured by the Laser Scanner.
7. PERCEPTION WORKSHOP (PW)

The exploitation of the validation triplet, constituted by the stereo pair of images and the filtered cloud of points corresponding to a scene, is done with the help of the specifically developed Perception Workshop software. PW includes the various tools required for the validation of the stereo-vision algorithms.

7.1 Stereo-vision Computation

The stereo-vision tool allows to apply the geometric distortion correction to the raw stereo-bench images, to select the stereo-correlation algorithm (grey level, mono-resolution, fast multi-resolution) and to set all the associated parameters (image sub-sampling ratio, pre-processing type, correlation window size, search range, disparity filtering...) in order to study their impact on the performances and to define an optimal set of parameters for a given robotic mission.

7.2 Virtual Stereo-Bench

The tool simulates a virtual stereo-bench imaging the 3D model of the scene measured by the Laser Scanner. The filtered point cloud is meshed, textured and the depth map corresponding to the scene is computed. The parameters of the virtual stereo-bench are used to compute a virtual disparity map that can be compared pixel by pixel to the real disparity map. The attitude of the virtual stereo-bench defined during the MGSE calibration and parallax-reducing phase can be modified by the user in order to better fit the virtual model to the real model.

7.3 3D Viewer

The 3D viewer allows the simultaneous display of the real and virtual 3D scenes to visually estimate the accuracy of the fit. The viewer implements various in-scene displacement modes, texture display and lighting effects and includes position and distance measuring tools. It is very effective to highlight scale, position and attitude differences in the models.

7.4 Disparity Profile

This visual tool displays the horizontal and vertical sections of the disparity map at locations selected by the user. It is mainly intended to study the behaviour of the algorithms at occlusions and object edges.

7.5 Stereo Benchmarking

While the visual comparison tools enable the user to do a rough estimation of the performances, mathematical evaluation functions were developed to help the user to compare the similarity of a real scene and its corresponding virtual one. The classical window-based scores (SAD, SSD, ZNCC) were effective to detect and display the mismatching regions, but a score based on 3D distances differences computed from disparities turned out to be the best for the parametric studies.

7.5 MGSE Geometry Optimisation Loop

The MGSE calibration and parallax reduction steps allow to obtain an acceptable fit between the real and virtual models. An optimisation loop was nevertheless developed to fine tune the position and attitude of the virtual stereo-bench in order to improve the fit. The optimisation is performed using user-chosen parameters (attitude, position...) and a mathematical similarity criterion.

7.6 Stereo Base Optimisation Loop

The size of the stereo base can also be optimised leading to the possibility to use the MGSE as an indirect stereo base measurement tool (the accuracy of the Laser Scanner was verified with the Laser Tracker on the landmarks balls).

8. PRELIMINARY RESULTS

Due to the Laser Scanner technical problem during the outdoor campaign, the resolution of the acquired scenes was not sufficient for the Exomars validation purpose.

Waiting for the 2011 full-resolution 3D campaigns, a preliminary study was still and all conducted on the indoor and outdoor data to reveal trends in the influence of the different parameters on the performances of the stereo-vision system.

These results were obtained using a 100mm base stereo-bench, equipped with 4.65μm pixel size cameras, and 4.8mm focal length lenses.
The default configuration used for the study was: 3DAcq MGSE geometry optimisation enabled to obtain the best fit between real and virtual point clouds, full resolution images, 7x7 correlation window size and mono-resolution algorithm.

Nota bene: In Tab. 2 et Tab 3 the values corresponding to that configuration are displayed in bold.

### 8.1 Global Performances

Tab.1 shows the indoor and outdoor mean accuracies.

<table>
<thead>
<tr>
<th>Scene Type</th>
<th>Mean Error (mm)</th>
<th>Std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indoor scenes</td>
<td>5.05</td>
<td>0.64</td>
</tr>
<tr>
<td>Outdoor scenes</td>
<td>13.27</td>
<td>1.55</td>
</tr>
</tbody>
</table>

*Table 1: Accuracy depending on the scene type*

The main hypotheses for the lower performances observed on the outdoor scenes are: the halved Laser Scanner resolution and the use of the wedge allowing imaging of more distant objects (errors increase with distance).

### 8.2 Impact of the Image Resolution

From a 1024x1024 pixels image, 512x512 or 256x256 pixels images can be obtained by hardware or software binning. The reduced number of points of the sub-sampled images enables faster processing of the vision algorithms.

Tab. 2 shows that the computing time gain obtained by image resolution degradation is much more significant than the induced accuracy loss.

<table>
<thead>
<tr>
<th>Image sub-sampling</th>
<th>Mean Error (mm)</th>
<th>Mean accuracy degradation</th>
<th>Number of pixels to process</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/1</td>
<td>13.27</td>
<td>0%</td>
<td>100%</td>
</tr>
<tr>
<td>1/2</td>
<td>16.60</td>
<td>25.09%</td>
<td>25%</td>
</tr>
<tr>
<td>1/4</td>
<td>20.45</td>
<td>54.11%</td>
<td>6.25%</td>
</tr>
</tbody>
</table>

*Table 2: Sub-sampling ratio impacts*

The relatively moderate impact on the accuracy may be explained by the fact that the sub-pixel interpolation allows the calculation of decimal disparities, thus reducing the impact of the pixel size increase on the performances.

Before selecting a baseline resolution, the impact of the resolution on the accuracy has however to be studied at Laser Scanner full resolution on “tough” textures.

### 8.3 Impact of the Correlation Window Size

The size of the correlation window impacts the disparity image computation time. Its effect on the accuracy is shown in Tab. 3.

<table>
<thead>
<tr>
<th>Correlation window size</th>
<th>Mean Error (mm)</th>
<th>Mean accuracy gain wrt 7x7</th>
<th>Estimated complexity wrt 7x7</th>
</tr>
</thead>
<tbody>
<tr>
<td>9x9</td>
<td>12.45</td>
<td>6.18%</td>
<td>+60%</td>
</tr>
<tr>
<td>7x7</td>
<td>13.27</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>5x5</td>
<td>15.00</td>
<td>-13.04%</td>
<td>-50%</td>
</tr>
</tbody>
</table>

*Table 3: Correlation window size influence*

On the tested sizes, the bigger the correlation window, the better the accuracy. The 5x5 correlation window size appears however to be a good trade-off between accuracy and computing complexity.

### 8.4 Robustness to Exposure Time Difference

The purpose of that work was to study the robustness of the stereo-vision algorithms to exposure time in general and to a discrepancy in the exposure of the left and right images in particular (case of a problem in the auto-pose algorithm or uneven drifts in time of the cameras electronics exposed to the space environment).

Tab. 4 and Tab. 5 present for an outdoor scene, the correlation rates and the reconstruction errors encountered when using different exposure times. The main diagonal corresponds to exposure times leading to similar histograms for the left and right images.

<table>
<thead>
<tr>
<th>Left/Right Exposures</th>
<th>5 ms</th>
<th>48 ms</th>
<th>81 ms</th>
<th>87 ms</th>
<th>135 ms</th>
<th>170 ms</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 ms</td>
<td>27.24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>55 ms</td>
<td></td>
<td>88.50</td>
<td>66.87</td>
<td>57.10</td>
<td>15.33</td>
<td>07.27</td>
</tr>
<tr>
<td>90 ms</td>
<td></td>
<td>75.03</td>
<td>89.68</td>
<td>81.58</td>
<td>47.92</td>
<td>28.70</td>
</tr>
<tr>
<td>100 ms</td>
<td></td>
<td>62.31</td>
<td>82.68</td>
<td>90.09</td>
<td>77.61</td>
<td>55.69</td>
</tr>
<tr>
<td>150 ms</td>
<td></td>
<td>19.64</td>
<td>57.60</td>
<td>82.59</td>
<td>90.18</td>
<td>86.70</td>
</tr>
<tr>
<td>200 ms</td>
<td></td>
<td>106.10</td>
<td>32.99</td>
<td>55.30</td>
<td>87.42</td>
<td>90.19</td>
</tr>
</tbody>
</table>

*Table 4: Correlation rates (%) for cross-exposures*

Nota bene: Due to parts of the areas present in the left image, but not perceived by the right camera (left part of the image, occlusions in the scene), the maximum possible correlation rate for our system is 94%.

<table>
<thead>
<tr>
<th>Left/Right Exposures</th>
<th>5 ms</th>
<th>48 ms</th>
<th>81 ms</th>
<th>87 ms</th>
<th>135 ms</th>
<th>170 ms</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 ms</td>
<td>15.74</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>55 ms</td>
<td></td>
<td>14.75</td>
<td>16.93</td>
<td>16.96</td>
<td>17.55</td>
<td>17.97</td>
</tr>
<tr>
<td>90 ms</td>
<td></td>
<td>16.12</td>
<td>14.15</td>
<td>15.72</td>
<td>18.64</td>
<td>20.41</td>
</tr>
<tr>
<td>100 ms</td>
<td></td>
<td>16.50</td>
<td>16.11</td>
<td>14.10</td>
<td>16.45</td>
<td>18.38</td>
</tr>
<tr>
<td>150 ms</td>
<td></td>
<td>17.78</td>
<td>18.05</td>
<td>15.57</td>
<td>13.82</td>
<td>15.25</td>
</tr>
<tr>
<td>200 ms</td>
<td></td>
<td>19.70</td>
<td>20.10</td>
<td>18.11</td>
<td>14.98</td>
<td>13.94</td>
</tr>
</tbody>
</table>

*Table 5: Reconstruction error (mm) for cross-exposures*
Longer times give better performances and are more robust to exposure differences between the left and right cameras.

A preliminary study of the fast multi-resolution algorithm was also conducted. Tab. 6 and Tab. 7 present the observed performances when difference appears between left and right camera exposure times.

<table>
<thead>
<tr>
<th>Left/Right Exposures</th>
<th>5 ms</th>
<th>48 ms</th>
<th>81 ms</th>
<th>87 ms</th>
<th>135 ms</th>
<th>170 ms</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 ms 55 ms</td>
<td>81.90</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>90 ms</td>
<td>92.74</td>
<td>93.08</td>
<td>92.48</td>
<td>91.14</td>
<td>89.96</td>
<td></td>
</tr>
<tr>
<td>100 ms</td>
<td>93.11</td>
<td>92.22</td>
<td>93.11</td>
<td>92.79</td>
<td>91.89</td>
<td></td>
</tr>
<tr>
<td>150 ms</td>
<td>91.97</td>
<td>92.11</td>
<td>92.94</td>
<td>93.19</td>
<td>92.98</td>
<td></td>
</tr>
<tr>
<td>200 ms</td>
<td>90.33</td>
<td>91.26</td>
<td>92.23</td>
<td>92.80</td>
<td>93.15</td>
<td></td>
</tr>
</tbody>
</table>

*Table 6: Correlation rates (%) for cross-exposures*

<table>
<thead>
<tr>
<th>Left/Right Exposures</th>
<th>5 ms</th>
<th>48 ms</th>
<th>81 ms</th>
<th>87 ms</th>
<th>135 ms</th>
<th>170 ms</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 ms 55 ms</td>
<td>19.59</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>90 ms</td>
<td>13.63</td>
<td>15.68</td>
<td>15.25</td>
<td>17.41</td>
<td>18.35</td>
<td></td>
</tr>
<tr>
<td>100 ms</td>
<td>14.62</td>
<td>12.92</td>
<td>14.59</td>
<td>16.27</td>
<td>17.36</td>
<td></td>
</tr>
<tr>
<td>150 ms</td>
<td>14.07</td>
<td>14.87</td>
<td>12.78</td>
<td>14.53</td>
<td>15.88</td>
<td></td>
</tr>
<tr>
<td>200 ms</td>
<td>16.22</td>
<td>16.48</td>
<td>14.92</td>
<td>13.61</td>
<td>12.66</td>
<td></td>
</tr>
</tbody>
</table>

*Table 7: Reconstruction error (mm) for cross-exposures*

Besides being faster, the multi-resolution correlation algorithm seems to be much more robust to exposure times difference, gives good results with low exposure time acquisitions and shows better accuracy than the mono-resolution algorithm. It is therefore a potential excellent candidate for the ExoMars mission.

9. FURTHER WORK

9.1 Full Resolution Outdoors Campaigns

The Laser Scanner was serviced to solve the mechanical noise problem and is again fully operational for the extended full-resolution outdoors campaigns planned for summer 2011.

These campaigns will include the full validation of the performances of the fast multi-resolution stereo-correlation algorithm.

9.2 “Tough” Textures Campaigns

The stereo-vision algorithm being based on texture correlation between the left and right images, campaigns on various outdoors sites outside of CNES will also be conducted to create a scene database with a large variety of textures for the validation and the tuning of the stereo-vision algorithms.

9.3 Margins for Autonomous Navigation

In addition to the stereo-vision system intrinsic errors (measured using the geometry optimisation loop), the definition of the security margins for the Autonomous Navigation has to take into account the geometric errors specific to the mount of the stereo-bench on the rover.

The study of this second error item is also planned for 2011.

10. CONCLUSION

We presented the validation method in use at CNES for the validation of the performances of the Exomars stereo-vision system.

The preliminary results obtained on the 2010 campaign data are very promising but have to be confirmed on a full-resolution dataset including “tough” textures.

On the validation dataset, the multi-resolution stereo-correlation algorithm showed up to give the best results besides being faster and more robust wrt to images exposure times.

11. REFERENCES

[1] CNES Space Technology Course – *Spacecraft Techniques and Technology* – Volume 3 Platforms