ROBUST VISUAL ODOMETRY FOR SPACE EXPLORATION: 12TH SYMPOSIUM ON ADVANCED SPACE TECHNOLOGIES IN ROBOTICS AND AUTOMATION

Dr Andrew Shaw(1), Dr Mark Woods(1), Dr Winston Churchill(2), Prof. Paul Newman(2)

(1) *SCISYS,23Clothier Road, Bristol, BS4 5SS, UK Email: andy.shaw@scisys.co.uk*
(2) *Dept. Engineering Science, Parks Road, Oxford, OX1 3PJ,UK Email: winston@robots.ox.ac.uk*

ABSTRACT

As part of the ESA X-ROB study[1], a Visual Odometry (VO) based localisation component was selected for evaluation and bread-boarding as part of a GNC solution for the European Ground Prototype (EGP). Existing ExoMars Rover, EGP and Sample Fetch Rover (SFR) requirements served as a baseline for VO evaluation.

The solution is based on the Oxford Visual Odometry (OVO)[1] kernel, with provision for further performance improvements fusing additional sensor data. The extensive test results show that it satisfies the critical reference requirements with a high level of performance, and additionally meets performance goals on flight-representative hardware.

1. INTRODUCTION

In the domain of space exploration several challenges are present; low processing speeds, strict power and mass requirements, high precision instrument placement etc. The XROB project used COTS technology for fast development and demonstration of a modular robotic hardware and software architecture. Various problems exist when performing localisation in an unknown environment with limited resources, some of which include:

- Using Inertial Measurement Units (IMU) for translational information is often not possible due to the relatively slow movement of the platform. Integrated velocity and distance information is within the noise measurements of the accelerometers.

- When using images for location matching over a long period of time various conditions may change reducing the accuracy of the matching values. Varying lighting conditions can have large effects, especially if shadows are being cast. They remove information, and as a vehicle moves its perspective on the environment changes also.

- Most navigational instruments have a tendency to drift.

Visual localisation can use a variety of cues such as:

- Easily recognisable large landmarks whose positions are known a-priori to a level of accuracy required to perform the mission (e.g. landmarks on the horizon)

- Less easily recognisable landmarks whose positions are not known to the platform before they are encountered (e.g. rocky desert)

- Less easily recognisable objects whose positions may change over the time of the mission (e.g. a rocky desert with dust storms)

- Scenes that are significantly similar in appearance with no obvious landmarks that would enable straight forward distinctions between locations to be made (e.g. dunes)

- Scenes that are relatively sparse of features and present particularly difficult challenges in terms of tracking position, estimation motion and planning (e.g. dunes)

A major question for VO is that as a visual feature based technique there is uncertainty over how well it can maintain consistent estimates over long ranges, as the process is a relative pose change estimate which will be subject to accumulated drift.

SCISYS have brought state of the art non-space technologies into the space arena particularly for autonomous rover exploration. The Oxford Visual Odometry (OVO) [1] technique was evaluated and found to have significant potential for application in this field. Motion estimation from images can be achieved in a variety of ways. The presented OVO system takes a feature point approach. Feature points are locations in the image that are easily identifiable and repeatable across multiple images, for example corners. By understanding how these points move in the image as the camera moves, we can compute the motion of the camera. By tracking even relatively modest numbers of features (~100) we can achieve very good performance.

[1] ESA Work Statement Ref XROB-RQ-001-ESA

The extensive tests results demonstrated that the component satisfies the critical reference requirements for the EGP [2], MER [3] and SFR [4] and a range of mobile platform missions with a high level of performance.

Requirements

As part of the study, a review of past, present and future mission requirements in terms of localisation was performed, a subset of these is contained in Table 1.

Table 1: Performance requirements of the localisation system.

Provide a 6DOF relative pose estimate at 1Hz
Provide a pose with an accuracy equal or better than 1% of the distance travelled (@10cm/s and 10deg/s)
Provide a pose with an accuracy equal or better than 5% of the distance travelled (@200cm/s and 20deg/s)
The Rover shall know its relative location, w.r.t. the last stopping place where topographic information was acquired for path generation, to better than 10 cm at all times during it locomotion
Pose accuracy shall be verified in a quarry with Martian analogue and at the ESTEC site at ranges upto 100m
Cameras shall have auto exposure, shutter and gain adjustments without manual input
The system shall not be sensitive to projected shadows
Shall provide pose estimated in daylight conditions
Shall provide a strategy to continue operations when the sun is in the FOV

2. TECHNICAL APPROACH

In the estimation it is assumed that all observations of feature points within a single image happen at exactly the same time. This is important as it vastly simplifies the estimation process. As a result cameras are used with global shutter systems, instead of rolling shutters that can easily produce warped images under even modest motion. Choosing to use a stereo camera alleviates the depth perception problem encountered with a mono camera, meaning metric motion estimation can be achieved. However it does result in some additional engineering considerations. When dealing with two images, it is important they are both captured at *exactly* the same time, for the same reason that global shutters are important. Further, it is helpful to assume the rows across the pair of images are exactly aligned, as this aids feature matching. For these reasons it is helpful to buy specifically manufactured stereo cameras where both sensors are houses in a single case, and these issues are dealt with. The two images from a stereo camera are often referred to as the "left" and "right" image and the pair are referred to as a "frame".

Image Rectification

Images from cameras are not perfect; they typically suffer from distortion to some degree. To fix this, before further processing is done, both images are rectified. The rectification parameters can be computed once for each camera, and then stored for future use. The rectification process is then very fast.

Figure 1: An example of image rectification carried out during the Tubney Quarry trials: the top image is un-rectified, the bottom image has been rectified. The effect is most obvious in the corners of the image, e.g. the hill in the upper right has been stretched.

Feature Extraction

Once the image has been rectified we search for feature points within the image, these are also known as points of interest. There is a vast literature on point of interest computation. In our system we use the FAST corner extractor. FAST produces are large number of candidate corners at small computational cost. Obviously some of these corners will be better than others, so they are ranked based on a "good corner" score (here we use the Harris score). To achieve robustness to

motion blur we run FAST at different scales of the original image, known as “pyramid levels”.

Spreading Features Across the Image

As previously mentioned, features are ranked according to a “good corner” score. If we just took the top *N* features, we typically find them clustered around a few strong corners in the image. Allowing this to happen results in a poorly constrained estimate. Instead we force the features to spread out across the image using a quad tree. This constrains the number of features in any particular point in the image, as well as the maximum number of features tracked. OVO typically tracks between 100-150 features in total across the image, this is illustrated in Figure 2. Through experimentation is has been shown that this number of tracked features is sufficient to provide the <1% error in pose estimate.

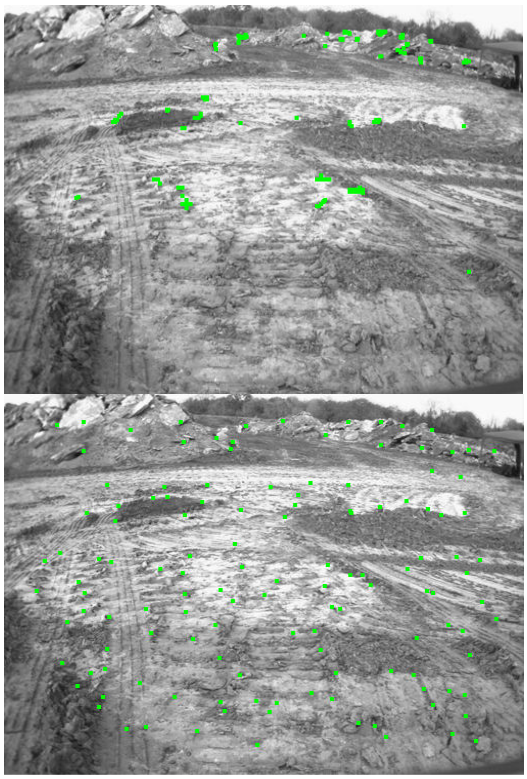

Figure 2: Extracted feature points, (top) identified points using a "good corner" score showing clustering, (bottom) feature extraction using quad trees producing across image dispertion

Left To Right Matching

When features are initialised for the first time, they need to be found in both the left and right image of the stereo pair. We use a stereo camera where we can assume the rows between the left and right image match up. Therefore if we have a feature on row *k* in the left image, then we have very cheap 1D search for the same feature on row *k* in the right image. We use mean SAD (sum of absolute differences) to compute the best matching score.

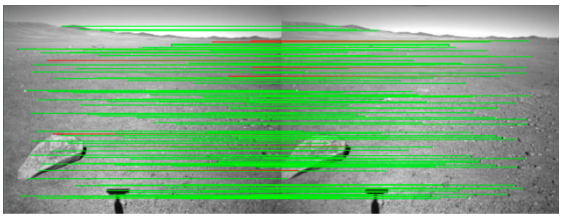

Figure 3: Image showing feature points matched across left and right image pair

Matching results between the left and right image of a stereo frame pair.

Temporal Feature Matching

Temporal matching is the process of matching two frames from time *t* to *t+1*. This is not as straight forward as left to right matching, as the camera could have undergone arbitrary motion between the two capture times. Therefore a more discriminative way is needed to match feature points. This approach uses a relatively new feature descriptor (a descriptor is a way of representing a point in an image), called Binary Robust Independent Elementary Features (BRIEF)[5]. BRIEF feature descriptors are very fast to compute and match meaning we can find good feature matches between two frames very quickly.

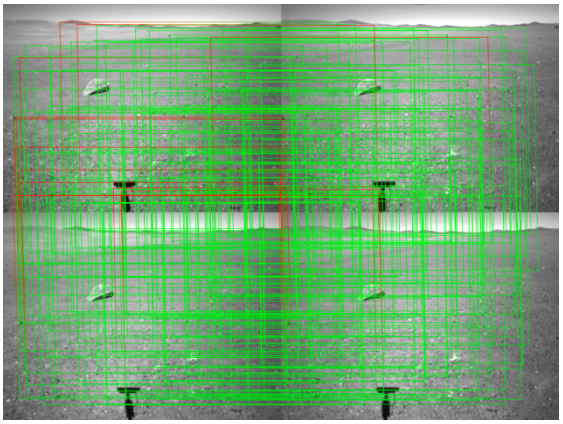

Figure 4: This image shows matching between two stereo frames. The upper image pair is the frame at time t+1, while the lower pair is the frame at time t.

Sub-Pixel Features

BRIEF provides an integer pixel location within the image, however better performance is achieved if we refine this estimate if with sub-pixel matching. This refinement is performed using Efficient Second Order Minimization (ESM)[6].

Motion Estimation

Once a set of refined feature matches between two stereo frames have been generated, the 6-DoF motion between the two frames can be computed starting with an initial RANSAC (RANdom SAmple Consensus)[7] step that highlights and removes any outliers. These will be incorrect matches that crept in during the temporal matching step. Least squares minimization is then performed using an *m*-estimator for robustness. After this step, any new outliers are also removed. Once the motion estimation has been completed and outliers removed, new features are added in the appropriate locations.

All pose estimates generated by the system are in the Nadir coordinate frame where x points along the vehicle forward direction, Y points to the right of the vehicle and Z points along the gravity vector.

3. PLATFORM

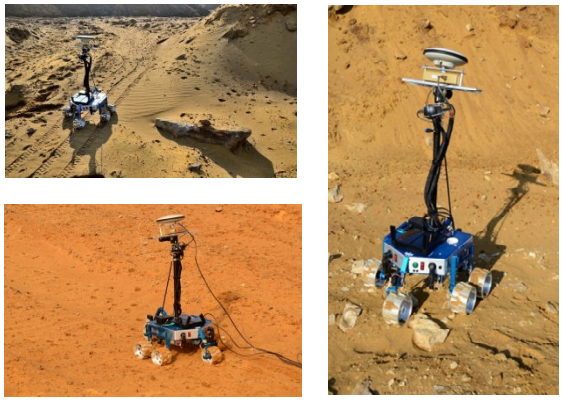

Figure 5: Various views of Tubney Quarry terrain with SCISYS Indie platform, Note dGPS receiver attached to the PTU of Indie to minimise offset errors

The SCISYS robotic platform INDIE, shown in Figure 5, which has the following configuration:

- Six wheeled passive platform design with high resolution motor feedback. This passive system provides ability to traverse extremely rough terrain without getting stuck.
- Front and rear wheel steering, coupled with differential drive system, allows on the spot rotation.
- Pan and tilt camera suite consisting of a high resolution (1280x960) stereo camera pair, a bumblebee stereo pair and a low resolution camera for real-time video streams. Provides motion through 300 * 180 degrees.
- Novatel DGPS module, position data accurate within 0.02 metres.
- 9 DOF inertial measurement unit with 3 axes of acceleration data, 3 axes of gyroscopic data, and 3 axes of magnetic data.
- On-board Processing; Intel core 2 duo mobile 2.33GHz processor, 2GB memory, 32GB solid state drive, Ubuntu 11.10 OS, direct C++ API, with language independent communication through CORBA

During testing of the localisation component the rover was placed in the environment and the system initialised, setting the local (relative) reference frame to zero and using the DGPS to provide the pose in the global (absolute) reference frame. Before testing the quarry environment was mapped and referenced prior to starting the test to obtain context, this data was down sampled to provide "orbital" data with a resolution of approximately 1m. The rover was then driven to a set of predefined way points, during the traverse the localisation component provided a continual pose estimate of the trajectory of the rover and an assessment of the accuracy was performed against the DGPS pose estimates. Both data sets were time stamped to allow easier matching.

The testing included both long "straight" traverses as well as "closed loop" traverses.

ESA EGP

The XROB stereo bench used during the Indie trials along with a relevant computer platform and SW was installed on the EGP (Figure 6) for the indoor trials at TAS-I. The only integration required for the test was the mechanical connection of the manual PTU mechanism and the electrical for power. There was no software interface to the EGP platform allowing no real-time access to wheel odometry and IMU readings for fusion. During real-time trials only Visual Odometry aspects were tested.

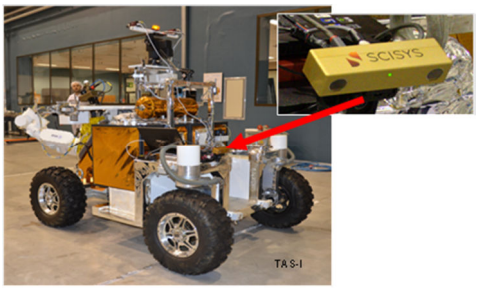

Figure 6: EGP Platform during the VO testing at the TAS-I clean room

4. GROUND TRUTH

To allow performance metrics to be obtained on the localisation system an external method of positioning the vehicle in the environment was required. The initial idea was to use a method of the triangulation of reflective markers and a laser range finder to calculate the actual 6DoF pose of the vehicle with respect to an origin. When using this method the actual pose estimates were in the order of ~5mm providing a good ground truth estimate, although this process was not continuous and it took approximately 4 minutes for each acquisition. This method was subsequently replaced in favour of a Novatel Differential GPS system that provided a continuous ~2cm accuracy at 1Hz pose estimate.

To set the system up a base station was deployed in the environment and allowed to settle for approximately 20-30mins after which the system can be used with high-accuracy (<2cm) feedback, using the system before this period could mean that the accuracy is reduced (~m).

Figure 7: (left) DGPS base station antenna deployed in the test environment.

5. TEST ENVIRONMENT

The test site chosen for the experiments was a sand quarry situated in Oxfordshire UK, an aerial image of which is shown in Figure 8. The various experiments were conducted in the sandy area which covers main the left side of the site as viewed in the image. The site had a variety of sand materials as well as several different types of aggregate. Depending on the trajectory taken through the site the traverse of ~300m are possible.

Figure 8: Aerial image of the Tubney quarry test site

Experiments

To thoroughly test the localisation and VO system requirements a variety of traverse types were identified and conducted in the quarry;

- Traverse in a straight line over 100m distance.
- Traverse with 360 degree rotation along the path.
- Traverse into Direct sunlight.
- Traverse containing all previous components.
- Traverse in a “Snake” motion.
- Traverse two large loops.
- Approach a gully face.
- Astronaut follow.
- Shadows changing in the field of view.

In the sequence of graphs that follow the top graphs shows the X-Y position of the vehicle, the middle graph shows the X-Z position and the bottom graph shows the Y-Z position.

100m straight line: The 100m straight line traverse was used as the initial baseline for the system, as it would be the most frequent. Figure 9 shows the results from this traverse when compared to the ground truth DGPS information.

Due to the velocity limitations of the platforms we had available the requirement for the 2m/s velocity was checked by removing images from the sequence to simulate travelling faster but capturing the images at the same frequency. From the results of the pose drift (also shown in Figure 9) it was possible to calculate that as long as the image translation was less than 35cm the system could compute a pose that was within 1% of the distance travelled.

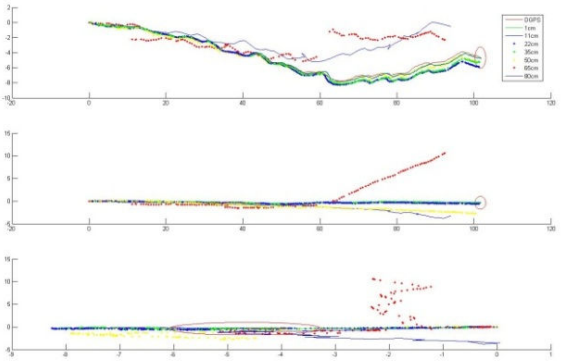

Figure 9: Trajectory comparisons over the 100m traverse. The red circle represents a 1% error margin and the red line shows the DGPS positions.

Traverse with 360 degree rotation: The second experiment was used to test the system's ability to cope with rotations as this is seen as a large potential source of error. Figure 10 shows the measured trajectory for the DGPS and the localisation system, again the final pose estimate is within the 1% distance travelled.

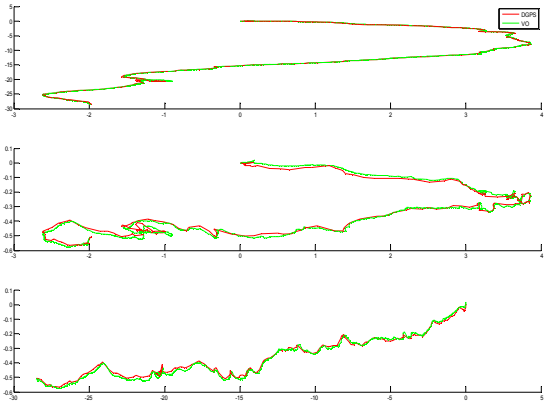

Figure 10: 37m traverse with 360 rotation half way along.

The DGPS provides accurate positional information, and although it can be integrated to obtain the orientation it was felt that the best way to demonstrate the system's robustness to rotation was to continue the traverse. This section after the rotation would easily highlight any errors due to the divergence of the two trajectories. Figure 10 shows that at the end of the traverse (bottom left) the difference in the end points is approximately 0.1% of the distance travelled.

Direct Sunlight: As the VO relies solely on the quality of the images being sufficient for feature extraction, the system was tested for its robustness in image and scene lighting condition changes. A trajectory was chosen so that the INDIE platform would navigate directly towards the sun. Figure 11 shows the trajectories from the systems.

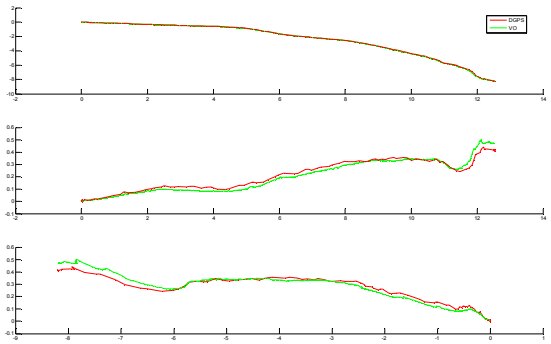

Figure 11: Results of the traverse directly towards the sun.

During the traverse the images underwent large changes in exposure from the surface reflection, sun glare and lens flare, again the system performed very robustly as the RMS error was approximately 0.4% of the distance travelled.

Long Range Traverse: Once the individual runs were captured, a traverse that navigated through all the previous trajectories was performed. Figure 12 shows the trajectories captured during a long 256m traverse. Note that the DGPS data contains several spikes - this was because the capture took place during high solar flare activity. The top of the graph shows the 100 metre traverse before traversing around a large mound (curve top right), then a 180degree point turn was performed before navigating back into the sun and then round several other obstacles in the quarry.

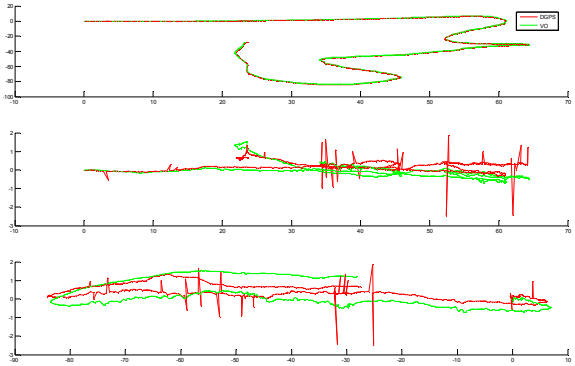

Figure 12: 256m trajectory that consisted of all the previous trajectories in a continuous capture.

Snake Trajectory: Again to show the systems robustness towards rotations a "snake" trajectory was performed. Here the vehicle traverse 30m

continually swerving before turning back and swerving toward the starting point, as shown in Figure 13.

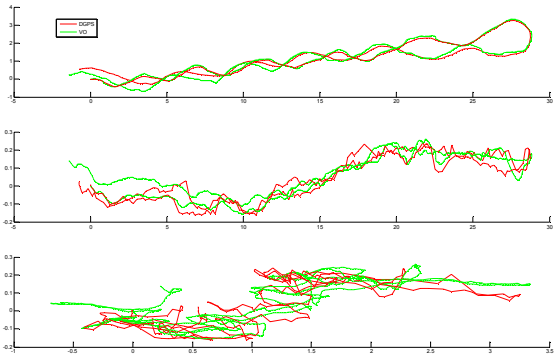

Figure 13: Snake traverse over 30m

Due to the roughness of the surface and the constant rotations the system (localisation and DGPS) experiences more noise than on previous runs, this is shown in the lower two graphs of Figure 13.

Astronaut Follow: With the requirements for EGP there will be times when the robot will be required to work alongside the astronaut and follow them to various locations. This test was used to determine how much of an effect the movement in the scene would have on the pose estimate. Figure 14 shows the ~30m traverse with the operator walking in the field of view, where they appear from the chest/waist down in the centre of the images.

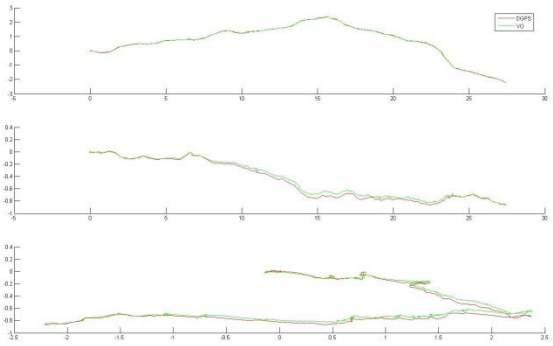

Figure 14: Trajectories generate when an astronaut was moving in front of the vehicle.

Traverse Summary

During the traverses the localisation system and VO, performed within the desired operational constraints.

Table **2** shows the numerical comparisons of the DGPS pose estimate and the localisation estimate for the various test runs that were performed. From this table it's possible to see that the tests that caused large rapid movements were the ones that were subject to higher errors, although general the system performed to approximately 0.3% of the distance travelled.

Table 2: Summary of the traverse distance, the measured RMS between the final DGPS and localisation pose and the percentage error in the measurement

Description	**Total Distance (m)**	**RMS error (m)**	**Percentage**
100m straight line traverse	106.39	0.33	0.31
Traverse with rotation	37.36	0.017	0.1
Direct sunlight	16.57	0.06	0.4
Slope	7.68	0.01	0.13
Around the quarry	256.13	0.99	0.38
Snake	64.85	0.72	1.1
High Speed	50	1.65	3.3
Loops	94.96	0.18	0.2
Gully	60.87	0.17	0.3
Approach	10.08	0.03	0.3

Astronaut follow	31.59	0.04	0.12
Long left	51.48	0.14	0.27
Shadows	0	0.44	N/A
Return to tent	5.3	0.26	0.47
Total	**1843.26 (total)**	**44.94 (total)**	**0.6 (average)**

6. FURTHER WORK

Further work has been performed to evaluate the suitability of the OVO and localisation architecture on space qualified hardware using the ExoMars breadboard (75MIPS Leon2) as a baseline[8]. Testing on flight-representative hardware using image data captured during a field trial to the Atacama Desert showed a mean time to process image frames below the required 10s per image frame, providing less than 1% RMS error.

This localisation system has been used alongside other SCISYS components in a project called GFreeNav for autonomous navigation of terrestrial vehicles in GPS denied environments, where a landrover vehicle was manoeuvred around a test facility at speeds of up to 2m/s.

The localisation is also being used on the Sample Acquisition Field Experiment with a Rover (SAFER)[2] ESA project as the main input for the SCISYS GNC software to control the Astrium Bridget platform in a Mars analogue environment.

7. CONCLUSION

This study showed that the use of visual odometry techniques can improve a vehicles localisation estimate to within the desired error bound for long range navigation. In addition to the local quarry trials carried out in the UK and at TAS-I a complementary version of the component has been tested in the most representative of conditions as part of a long-range (several autonomous km per day) navigation experiment (SEEKER[3]) in the Atacama Desert. This provides a high degree of confidence in the proposed technology as this environment offered the complete range of vision/terrain conditions expected on a Mars flight mission - something which cannot be tested locally in Europe. Given the results of this activity where the component was integrated as part of a complete GNC solution it indicates that the technology is at TRL 6 – "System/subsystem model or prototype demonstration in a relevant environment (ground or space)".

[2] ESA extended contract 4000104031/11/NL/NA

[3] ESA contract 4000104031/11/NL/NA

8. REFERENCES

1, G Sibley, C Mei, I D Reid and P M Newman, (2010). Planes, Trains and Automobiles -- Autonomy for the Modern Robot, Proc IEEE Int Conf on Robotics and Automation.

2, P.H.M. Schoonejans, (2004). EUROBOT System Requirements Document, ESA Doc. n. MSME-RQ-EB-0001-ESA, Issue 1 Rev. 1.

3, Mars Exploration Rovers, (2013), http://marsrovers.jpl.nasa.gov/home/index.html.

4, Allouis, E., Jordan, T., Patel, N., Ratcliffe, A., (2011), Sample Fetching Rover, In Proceeding ASTRA 2011.

5, M. Calonder and V. Lepetit and C. Strecha and P. Fua, (2010), BRIEF: Binary Robust Independent Elementary Features, Eurpean Conference on Computer Vision

6, Mei, C. and Benhimane, S. and Malis, E. and Rives, P., (2008), Efficient Homography-Based Tracking and 3-D Reconstruction for Single-Viewpoint Sensors, Journal IEEE transactions on Robotics

7, Martin A. Fischler and Robert C. Bolles, (1981), Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography, Journal of Comm. of the ACM

8, Hult, T., Petersén, A., Dean, B., Winton, A., (2010), The ExoMars Rover Vehicle OBC, In Proceedings of DASIA 2010