

SENSITIVITY ANALYSIS OF VIRTUAL TERRAIN ACCURACY FOR VISION BASED ALGORITHMS

Master's Thesis

RÓBERT MARC, BENG

ESA/ESTEC - EUROPEAN SPACE AGENCY

JUNE 5, 2012

Supervisors



Tomás Pajdla, PhD.

Assistant Professor
Center for Machine Perception, Department of Cybernetics
Faculty of Electrical Engineering
Czech Technical University in Prague
České vysoké učení technické v Praze



Luc Joudrier

Senior Robot System Engineer
Automation and Robotics Section, Mechatronics and Optics Division
Directorate of Technical Management and Quality
European Space Research and Technology Centre, The Netherlands
European Space Agency



Anita Enmark, PhD.

University Lecturer (*Universitetslektor*)
Space Technology Division
Department of Computer Science, Electrical and Space Engineering
Luleå University of Technology
Luleå tekniska universitet

Programme

Erasmus Mundus: Joint European Master in Space Science and Technology



To my parents: Ildikó and László

"The journey is what brings us happiness, not the destination."

Dan Millman, *Way of the Peaceful Warrior*

Declaration of Authorship

I, Róbert MARC, graduate student of the Czech Technical University in Prague and Luleå University of Technology, hereby declare that the ideas, analysis, design, development, results and conclusions within this Master's thesis represent my own effort except for those elements which are highlighted accordingly in this document. All references have been distinguished, quoted in enclosed reference list and all sources of information have been specifically acknowledged.

I also declare that, from my knowledge, this Master's thesis is an original work and was never presented before in any institution. This thesis is handed in to both universities in order to obtain the joint double degree - *MSc Elektrotechnika a informatika - Technická kybernetika* and *Teknologie masterexamen med huvudområde rymdteknik*

Noordwijk, The Netherlands

June 5, 2012

Abstract

A number of three-dimensional virtual environments are available to develop vision-based robotic capabilities. The virtual environments have the advantage of repeated trials at low cost compared to field testing. However, they still suffer from a lack of realism and credibility for validation and verification, especially when considering the immense variety of terrain types.

This work consists of the creation and validation of state of the art virtual terrains for research in Martian rover vision-based navigation algorithms. The present Master's thesis focuses on the creation of virtual environments, which are the exact imitations of the planetary terrain testbed at the European Space Agency's ESTEC site. Two different techniques are used to recreate the Martian-like site in a simulator. The first method uses a novel multi-view stereo reconstruction technique. The second method uses a high precision laser scanning system to accurately map the terrain. Later, a textured digital elevation map is obtained by image projection.

With the help of a precise positioning system a comparison of real environment to the virtual environments is done at exact same locations by making use of captured stereo camera images. Ultimately, the differences will be characterized by the main known feature detectors (e.g. Harris, SURF, and SIFT).

It has been demonstrated that the multi-view stereo reconstruction technique developed at the Czech Technical University in Prague attains a degree of accuracy which is very close to a laser scan reconstruction. Depending on the level of details of the virtual terrain, the captured images are both quantitatively and qualitatively analyzed.

This Master's Thesis work led to the creation and validation of a database containing highly realistic virtual terrains which can be found on Mars for the purpose of vision-based control algorithms verification.

Acknowledgements

First and foremost I would like to express my gratitude to my supervisor from the Czech Technical University, Tomáš Pajdla. He kindly accepted to supervise my work, although the topic changed at the last moment. I am heartily thankful to my direct supervisor at ESA, Mr. Luc Joudrier, whose encouragement, guidance and support from the initial to the final level enabled me to develop an understanding of the subject. Moreover, I would like to thank to Kapellos Konstantinos from Trasys for his support and collaboration during the project. Regarding 3DROV, his consultancy was greatly appreciated. Also, I would like to mention the help of Michal Jančošek, from the CMP research group.

A big thanks goes to the people with whom I am working together at the Automation and Robotics section at ESTEC, who allowed me to fit in into a very international group of researchers. So thanks Gianfranco, Michel, Luc, Pantelis, Kjetil, and Martin.

The help of Stephan from Teleobotics section is also appreciated. Discussing image processing with him provided me a good insight in the field. Kjetil was also giving good ideas during the analysis. So thanks guys!

Also I would like to mention the names of people who read the draft and provided useful comments. Thanks Maike, Felix, Marc, and Zainab.

All my work carried out at ESTEC would be not possible without the help of Victoria Barabash. She was the one who signed a cooperation agreement between SpaceMaster consortia and the European Space Agency, allowing me to carry out my Master's thesis in this framework. Also I am specially grateful to Martin Hromčík who was teaching and mentoring the SpaceMasters at the Czech Technical University in the third semester. But more than this, he was the one who helped me obtain the CTU scholarship for international students. Thanks also goes to Jana Nováková for her patience and help in the various administrative tasks in Prague.

During our stay in Sweden, Maria Winnebäck and Anette Snällfot - Brändström were taking care of the SpaceMaster students. Personally, I would like to thank them for all the excursions above the Arctic circle, and the kindness which all felt during our stay in Kiruna. Thanks also goes to all the people who were involved in the first academic semester in Würzburg.

A big thank you goes to all present and former SpaceMaster students. I have met amazing and unique people from all over the world, giving me the chance to explore these cultural differences. The programme offered me not just the possibility to study in three different universities, but much more: visiting picturesque places, living exciting moments and last but not least getting involved in several projects.

My deepest gratitude goes to my family for their unflagging love and support throughout my life, especially during my student years. Without their support, my ambition to study abroad could not be realized.

No acknowledgment or thanks would be meaningful without a special thanks to my girlfriend Maike, who never once faltered in her encouragement during this thesis. My thanks and my love goes to her.

The financial support from the different institutions is greatly appreciated: namely the Czech Technical University in Prague, through the Scholarship for International Students; Erasmus mobility grant from Luleå Technical University and last but not least the support coming from the European Space Agency during the internship.

Lastly, I offer my regards to all of those who supported me in any respect during the completion of the thesis.

DIPLOMA THESIS ASSIGNMENT

Student: **Róbert Marc**

Study programme: Cybernetics and Robotics
Specialisation: Systems and Control

Title of Diploma Thesis: **Sensitivity Analysis of Virtual Terrain Accuracy for Vision Based Algorithms**

Guidelines:

The thesis focuses on the creation of a virtual environment as accurately as possible by projecting texture onto a DEM (Digital Elevation Map). With the help of a small terrain, RIEGL laser scanner, accurate positioning system called VICON, and CMP SFM Service (ptak.felk.cvut.cz/sfmservice) we will compare the real environment and the virtual environment at the same locations. We will characterize the differences in feature selection from the main known feature (e.g. Harris, SURF, and SWIFT) depending on the level of details of the virtual terrain.

1. Literature survey.
2. Creation the digital version of the terrain with various accuracy.
3. Sensitivity analysis based on feature selection of DEMs in order to support visual odometry of rovers.
4. Implementation of various image analysis techniques and algorithms
5. Creation of image comparison tool (virtual against real) for various features like Harris or Surf
6. Analysis, verification and validation of the results.

Bibliography/Sources:

- [1] S. Chhaniyara, C. Brunskill, B. Yeomans, M. Matthews, C. Saaj, S. Ransom, and L. Richter. Terrain trafficability analysis and soil mechanical property identification for planetary rovers: A survey. Journal of Terramechanics, vol. 49, pp. 115–128, 2011.
- [2] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool. Speeded-up robust features (SURF). Computer Vision and Image Understanding, pp. 346–359, 2008.
- [3] M. Jancosek and T. Pajdla. Multi-view reconstruction preserving weakly-supported surfaces. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3121–3128, 2011.
- [4] C. Strecha, W. von Hansen, L. Van Gool, P. Fua, U. Thoennessen On Benchmarking Camera Calibration and Multi-View Stereo for High Resolution Imagery CVPR 2008
http://cvlab.epfl.ch/~strecha/publications/strecha_cvpr_2008.pdf

Diploma Thesis Supervisor: Ing. Tomáš Pajdla, Ph.D.


prof. Ing. Michael Šebek, DrSc.
Head of Department

Valid until the summer semester 2012/2013




prof. Ing. Pavel Ripka, CSc.
Dean

Prague, May 11, 2012

Table of Contents

Declaration of Authorship	i
Abstract	ii
Acknowledgements	iii
1 Introduction	1
1.1 Motivation	3
1.2 Planetary rover missions	3
1.2.1 Planetary rovers	3
1.2.2 Short history of lunar and planetary missions	4
1.2.3 Martian exploration	5
2 Martian and Martian-like Terrains	11
2.1 Characterization	11
2.1.1 Overview	11
2.1.2 Methods	13
2.2 Surface Hazards for Rover Navigation	13
2.3 Visual Odometry of Rovers	14
2.4 Testing on the Earth	15
2.4.1 Planetary Terrain Testbed	16
2.4.2 Overview of 3DROV	17
2.4.3 EXOMADER	18
3 Digitalization of the Planetary Testbed	19
3.1 Active Technique	19

3.1.1	Laser Scanning	19
3.1.2	Motion Tracking	23
3.2	Passive Technique	25
3.2.1	Classification of Multi-View Reconstruction Algorithms	25
3.2.2	Theoretical Background	26
3.3	Lens Correction	28
4	Image Processing	30
4.1	Optical Methods of Terrain Characterization	30
4.1.1	Stereoscopy	30
4.1.2	3D Imaging	31
4.1.3	LIDAR	31
4.2	Feature Points	32
4.2.1	Local features	32
4.2.2	Feature Detectors	33
4.2.3	Corner Detectors	34
4.2.4	Local Feature Detectors	37
4.2.5	Region detectors	41
4.2.6	Overview of Feature Detectors	42
5	Measurements and Simulation	44
5.1	Coordinate Systems	44
5.2	Active Technique	45
5.2.1	Measurements	45
5.2.2	Digital Elevation Map	47
5.3	Passive Technique: Multi-View Stereo Reconstruction of the Testbed	51
5.4	Cameras	51
5.4.1	Calibration	51
5.4.2	Stereo Cameras	54
5.5	Simulations Using 3DROV	56
5.5.1	Importing Components	57

5.5.2	Spatial Co-Registration and Image Acquisition	58
5.5.3	Experimental Scenarios	60
6	Analysis and Results	63
6.1	Laser Scan and Multi-View Stereo Reconstruction Comparison	63
6.1.1	Height Maps	63
6.1.2	Triangle-Meshes	64
6.2	Comparison of Stereo Camera Images	66
6.2.1	Contrasting Images	67
6.2.2	Preliminary Results	68
6.2.3	FFT analysis	70
6.2.4	Aligning Images	71
6.2.5	Features Location	74
7	Conclusions and Future Work	76
	Appendix	78
	Acronyms list	81
	Nomenclature	82
	References	83

List of Figures

1.1	Short graphical summary of the present work	2
1.2	Current action radius of planetary/cometary exploration (credits: DLR/Spohn)	5
1.3	Mars Earth size comparison (credits: NASA)	6
1.4	Three generation of NASA rovers (in order of increasing size): Sojourner, Spirit and Curiosity (credits: NASA/JPL-Caltech)	7
1.5	Illustration showing the elements that were to make up the original ESA-U.S. ExoMars Program (credits: ESA)	8
1.6	Former ESA and NASA collaboration plan (credits: NASA/MEPAG)	9
1.7	Artistic impression about ExoMars during rock drilling process (credits: ESA)	9
1.8	Mars exploration overview (credits: Jason Davis [1])	10
2.1	Lahontan crater on Mars seen from Spirit rover (credits: NASA)	12
2.2	'Lookout Panorama' from Spirit taken on February 27th to March 2nd, 2005 (credits: NASA)	12
2.3	Classification of terrain characterization method (Chhaniyara et al. [2])	13
2.4	Typical rover control architecture	14
2.5	Martian terrain constituents (left to right): basalt rock with 'terre rouge'; basalt grains; Martian rock like gypsum	16
2.6	3DROV building blocks	17
2.7	ExoMader custom built rover in the ARL	18
3.1	Terrestrial laser scanning equipment set-up	20
3.2	The electromagnetic spectrum (source: [3])	21
3.3	Time-of-flight laser scanner principle	22
3.4	Dimensional drawing of the LMS-Z210i TLS (Source: [4])	23
3.5	Retro reflective markers	23

3.6	Vicon MX13+ and specification	24
3.7	Spatial distribution of the Vicon MX13+ cameras	24
3.8	Voronoi diagram (gray edges) of points and its associated Delaunay triangulation (black edges) (showed in [5])	26
3.9	Weakly-supported surfaces (credits: M. Jancosek [6])	27
4.1	Composite image of the Exomader rover (red - left image, cyan - right image)	31
4.2	Importance of corners and junctions in object recognition (Source: [7])	33
4.3	Illustration of the components of the second moment matrix and Harris cornerness measure (Source: [8])	35
4.4	FAST Feature detection in an image patch (Source: [9])	36
4.5	Integral image (Source: [8])	38
4.6	Gaussian partial derivative comparison with SURF's box filter (y, x-y dir.) (Source: [8])	39
4.7	Feature detection with SURF: the 50 strongest points (left); the 150 strongest points (right)	39
4.8	Storing SIFT keys in sample images and identify matching keys from new images (Source: [10])	41
5.1	Polydata made by 6 laser measurements showing the scanner positions	46
5.2	Reference system origin and calibration wand	47
5.3	12 markers seen by the two systems	48
5.4	Overlapped poly-data from six measurements (height representation from a side view)	49
5.5	Top view of the different triangulation results (colored height representation)	50
5.6	Camera positions and orientation of the 113 images taken around the testbed	51
5.7	CMP SfM reconstruction of the ARL testbed (perspective view)	52
5.8	Image comparison	53
5.9	Stereo camera: from CAD drawing to reality	54
5.10	Field of view depending on focal length and sensor size	55
5.11	Aerial views: transformation from 3D reconstruction to 3DROV supporting format	57
5.12	ExoMader model on the laboratory testbed simulated in 3DROV	58
5.13	BW stereo camera pictures	59
5.14	BW stereo camera pictures	59
5.15	Origin in 3DROV on the multi-view reconstructed terrain	60
5.16	Results of automatic correction of lens distortion by Matlab toolbox	61
5.17	Texture projection	61

5.18 Precision correction method by using manual tuning	62
6.1 Comparison of simulated environments in 3DROV	63
6.2 Histogram comparison of the height maps	64
6.3 Matched 3D meshes	65
6.4 Side view of the matched surfaces with ICP (green: laser scan DEM; red: multi-view reconstruction DEM) . . .	65
6.5 Error between the two meshes (red/green lines) on a multi-view reconstructed surface (8.9 m x 8.9 m)	66
6.6 Error convergence of ICP algorithm	67
6.7 Contrast enhancement techniques comparison	67
6.8 Before-after comparison of contrasting images	68
6.9 SURF points detected on unaligned images	69
6.10 FFT representation of images from location three (logarithmic)	70
6.11 Filtering the 'real' image	70
6.12 Re-aligned image: reality (lower layer), simulation (upper layer)	71
6.13 Correctly matched SIFT points for the purpose of image alignment	72
6.14 Comparison of real and simulated environments at location 3 in ARL respectively in 3DROV	73
6.15 MSER region detection comparison	74
6.16 SURF features comparison	75

List of Tables

2.1	EXOMADER chassis specifications	18
4.1	Feature detectors: Overview (Source: [8])	43
5.1	Measurement angles	46
5.2	Marker positions on the testbed (in meters)	47
5.3	Triangulated DEMs with different precision	49
5.4	Calculated calibration parameters (in pixels) on 1024 x 768 pixel images	53
5.5	Calibration parameters (in pixels) on 3264 x 2448 pixel images	54
5.6	uEye camera specifications	55
5.7	Measured parameters of the stereo cameras in pixels [11]	55
5.8	Calculated FOV parameters of the stereo cameras	56
6.1	Feature points matching with SIFT on 640 x 512 pixel images at location 3 (see Figure 6.13)	74
6.2	Parameters of the applied feature detectors	74

Chapter 1

Introduction

This introductory chapter gives general information of the work which will be presented, along with the background necessary to understand the multidisciplinary field of robotics. The combination of different fields gives a unique insight to the current planetary rover challenges.

Robotic technologies play an important role in the exploration of planetary surfaces and deep space. Acting as proxies for human exploration, robots provide the key basic functionalities for the scientific investigation of planetary surfaces, mobility between investigation sites, execution of specific tasks such as the deployment of scientific instruments, and data transmission to Earth.

Chapter 1 discusses the motivation of the current research. Since this work focuses on vision-based analysis of virtual terrain accuracy, certain aspects of planetary testing must be discussed. Moreover, it is a must to briefly present the planetary explorations, both past and future ones, in particular the missions regarding the Red Planet¹. ExoMars will be briefly presented, since the present work is related to this visionary ESA project.

Chapter 2 introduces Martian-type terrains as they are known to mankind. In this part, different terrain characterization methods are shortly described. Also, the dangers regarding the exploration of the planetary surfaces are discussed, as are the hazards regarding surface missions. These missions are dependent on visual odometry of rovers, as will be shown. Taking all the factors into consideration, planetary surface missions need to be extensively tested on Earth. This can be achieved only during test campaigns which take place in similar conditions.

Chapter 3 introduces the tools and methods used for creating the exact copy of the planetary testbed found at ESTEC², in the Automation and Robotics Laboratory. The chapter deals with the theoretical background of these techniques.

Chapter 4 contains a brief introduction to a few optical methods for terrain characterization (like stereoscopy). The major part of this chapter is concerned with the different feature matching algorithms found in the literature. Although it is a vast area in image processing, we will describe just a few important ones which will be used in Chapter 6 as a method of analysis.

¹Red Planet is a nickname for the planet Mars, due to its surface color

²European Space Research and Technology Centre

The next part, Chapter 5, describes the different measurements used to create simulated Martian-like terrains. These can be categorized as active (laser scan reconstruction), or passive (multi-view stereo reconstruction), which are comparable in terms of accuracy. This chapter deals with the different scenario set-ups in both real world and simulations.

Chapter 6 provides the results which were found during the analysis by using different algorithms and techniques. Preliminary conclusions are drawn by comparing the simulations to the real test terrain.

In Chapter 7, final conclusions are presented regarding the virtual accuracy of a highly unique planetary simulation environment, in relationship with the actual reality. Regarding the direction of current and future work, we would like to try to establish some directions for the upcoming research. A more detailed quantitative analysis will be necessary to improve the results. Also, suggestions and recommendations are discussed for the Automation and Robotics section in ESTEC to further improve the accuracy of the planetary simulation environment.

A short summary of the present Master's thesis is shown in Figure 1.1. In the initial phase, the lifelike copy of the terrain was created with two different methods (bottom). The validation of these digital elevation maps were done through stereo camera images in both real world (left column) and simulation (right column of images). Finally a comparison will be carried out using different low level image processing techniques.

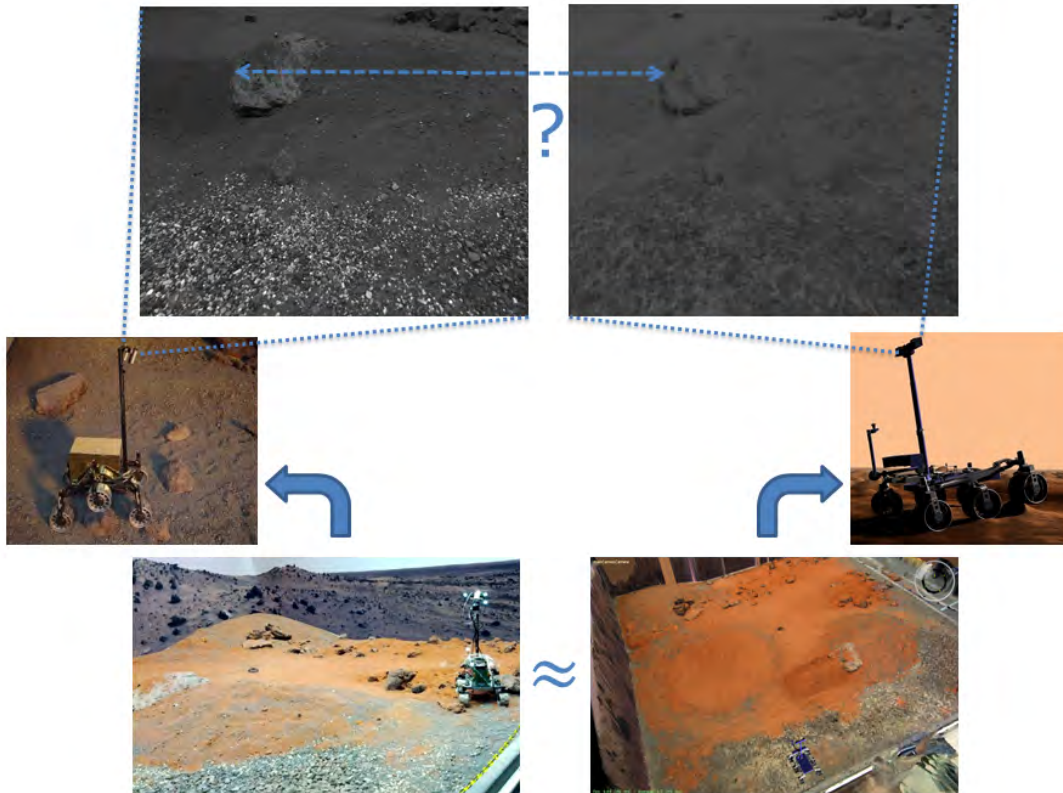


Figure 1.1: Short graphical summary of the present work

1.1 Motivation

Given the complexity of rovers and the fact that testing in operationally equivalent environments is either very expensive or impossible, simulations are a prominent tool both for the development and the validation process.

A number of 3D virtual environments are available to develop vision-based robotic capabilities. They have the advantage of repeatability at low cost compared to real testing. However, they still suffer from a lack of realism and credibility for validation and verification, especially when considering the immense variety of terrain types. This work explores the existing void in this direction.

Nowadays simulations are a necessity in almost all fields of science. The motivation is simple: they can accurately depict reality. By varying the accuracy, reality can be approximated. Of course the ideal situation would be a 100 % simulation accuracy to model the exact reality.

The difficulty of designing and testing a planetary rover arises from the lack of knowledge and reproduction capabilities of the target environment. By means of simulation software this can be surmounted. Although it sounds easy, it is not always that simple to overcome certain problems. The work is carried out at the European Space Agency's Automation and Robotics Section, which is active in the support of the two main European planetary exploration missions (ExoMars³ and MoonNEXT⁴) as well as in research and development activities for the development of technology building blocks in the field of space robotics and automation. Within this scope the in-house development activities include: testing and integration of rover robotic components and technologies at system level, prototyping concepts and selective shadow-engineering of industrial activities.

The present work will be used in the vision-based control algorithms verification phase for planetary rovers. The objective is to demonstrate that virtual modeling can be sufficiently accurate for proper verification and validation of rover GNC⁵ flight software.

A future outcome may be to avoid the need of building a large Mars Yard to investigate long traveling and path planning algorithms.

1.2 Planetary rover missions

1.2.1 Planetary rovers

Rovers have some advantages over stationary landers: they examine more territory and therefore more interesting features can be detected on a remote planet. Rovers can advance the knowledge of how to perform remote robotic vehicle control which is necessarily semi-autonomous due to the finite speed of light.

Moreover planetary rovers have the advantages over orbiting spacecraft that they can make observations to a microscopic

³Exobiology on Mars

⁴European Lunar Lander proposed by the European Space Agency

⁵Guidance, Navigation and Control

level and can conduct physical experimentation. Disadvantages of rovers compared to orbiters are the higher chance of failure, due to landing and other risks, and that they are limited to a small area around a landing site which itself is only approximately anticipated.

Because of these risks rovers need to be highly reliable, they have to withstand high levels of acceleration, high and low temperatures, pressure, dust, corrosion and cosmic rays, while remaining functional without repair for a needed period of time.

Another issue with planetary rovers is that they cannot be controlled in real-time since the speed at which radio signals travel (299792 km/s) is far too slow for real-time or near real-time communication. The communications delay between Earth and Mars can vary between five and twenty minutes depending upon the relative positions of the two planets. As a consequence of this, rovers need to be fully autonomous on Mars.

Another issue is the power for transmitting over such a huge distance. Continuous and direct data transmission to Earth is already a considerable challenge, but overcoming the problem of receiving a faint signal from Mars is a bigger one. One must use very large antennas for such situations (e.g. 70m). The US rovers are designed in such way that they transmit limited data back to Earth from time to time. It is mainly used to upload the daily activities at the start of the rover day. This is why ESA is developing the concept of UHF communication toward a relay orbiter and X-Band from Orbiter to Earth.

1.2.2 Short history of lunar and planetary missions

The Lunokhod 1 rover landed on the Moon in November 1970 [12]. It was the first roving remote-controlled robot to land on any celestial body. The rover drove 197 m, and during 10 communication sessions returned 14 close-up pictures of the Moon and 12 panoramic views. It also analyzed the lunar soil during its 11 months of operation.

NASA included Lunar Roving Vehicles in three Apollo missions: Apollo 15 (which landed on the Moon July 30, 1971), Apollo 16 (which landed April 21, 1972) and Apollo 17 (which landed December 11, 1972).

The Lunokhod 2 was the second of two unmanned lunar rovers landed on the Moon by the Soviet Union as part of the Lunokhod program. It was the second roving remote-controlled robot to land on any celestial body. Lunokhod 2 operated for about 4 months, covered 37 km of terrain sending back 86 panoramic images and over 80,000 TV pictures [13]. There were other ambitious Russian plans to reach Mars and Moon but due to funding and other problems projects like Soviet Mars 2 and Mars 3, Lunokhod 3, and Marsokhod never flew.

The Mars Pathfinder mission included Sojourner, the first rover to successfully reach another planet. NASA, the space agency of the United States, launched Mars Pathfinder on December 4, 1996; it landed on Mars in a region called Chryse Planitia on July 4, 1997. From its landing until the final data transmission, Mars Pathfinder returned 16,500 images from the lander and 550 images from Sojourner, as well as data from more than 15 chemical analyses of rocks and soil and extensive data on winds and other weather factors [14].

1.2.3 Martian exploration

Realistically and objectively speaking, current and near future planetary manned missions are limited by our relatively small action radius (see Figure 1.2). This includes low Earth orbit (LEO), the Moon, Lagrange points, near Earth asteroids (NEAs), Venus and Mars. Currently the most fascinating destination for humans is Mars. This is mainly due to the relative closeness to the Earth, which makes it reachable in a 'short' time period. Martian exploration would help to understand life, and search for the origin of life. The exploration is indeed necessary in this direction, and although we cannot predict the impact on society, it will certainly have positive outcomes in the future.

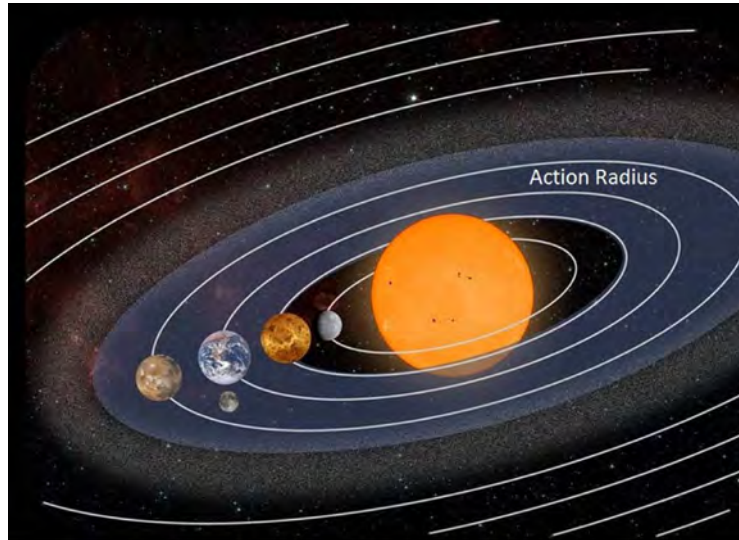


Figure 1.2: Current action radius of planetary/cometary exploration (credits: DLR/Spohn)

One can debate for years about the rationality of such explorations, but in principle this is what mankind has always been doing. Exploring without really considering it, or seeing the near- or long-term consequences. If such a Martian human exploration mission is feasible, one day it will be done. The search for life on another planet is really trying to answer a fundamental question: is there life only on Earth? And as a consequence, are we alone in the Universe?

Life on Mars

Mars⁶ is the fourth planet from the Sun in our Solar System, often described as the 'Red Planet', the color given by the iron oxide prevalent on its surface. Mars is one of the terrestrial planets and has a thin atmosphere. Its surface shows features that are reminiscent of the impact craters on Earth's Moon on the one hand, and to volcanoes, valleys, deserts and polar ice caps found on Earth on the other hand [15]. Due to its tilted rotational axis, Mars has seasons. Combined with the rotational period of Mars, the seasonal cycles resemble those on Earth. On Mars the highest known mountain within the Solar System can be found. *Olympus Mons*⁷ is almost three times as tall as Mount Everest's height above sea level. Also one of the largest known

⁶Named after the Roman god of war.

⁷Height of nearly 22 km

canyons of the Solar System, *Valles Marineris*⁸, has been localized on the surface of Mars by Mariner 9.

It was believed, but not proved, that the surface of Mars contains water. Only in 2005, radar data revealed the presence of large quantities of water ice at the poles [16] and at mid-latitudes. The Mars rover Spirit sampled chemical compounds containing water molecules in March 2007. The Phoenix lander directly sampled water ice in shallow Martian soil on July 31, 2008 [17].

It is known that water is the key factor for growing organisms and for proof of life inside our Solar System. The current understanding of planetary habitability - the ability of a world to develop and sustain life - favors planets that have liquid water on their surface. This most often requires that the orbit of a planet lies within the habitable zone, which for the Sun currently extends from just beyond Venus to about the semi-major axis of Mars. In other words this makes Mars as the number one candidate for life, besides Earth.

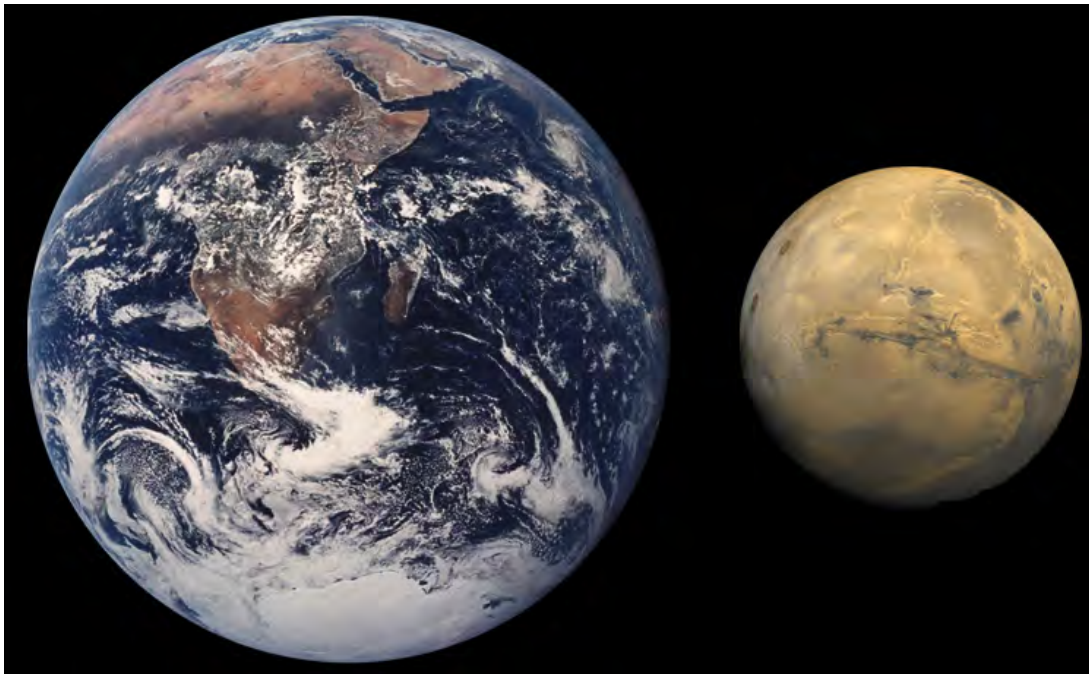


Figure 1.3: Mars Earth size comparison (credits: NASA)

The lack of a magnetosphere and extremely thin atmosphere are a challenge for life. Mars has a very small amount of heat transfer across its surface. Moreover there is insufficient Martian atmospheric pressure to retain water in a liquid form (it sublimates to a gaseous state). And also there is poor insulation against bombardment of the solar wind ions.

Evidence suggests that the planet was once significantly more habitable than it is today, but whether living organisms ever existed there remains unknown. The soil nutrients found during NASA missions (like magnesium, sodium, potassium and chloride) suggest that may be able to support life, but life would still have to be shielded from the intense radiation [15].

⁸More than 4,000 km long, 200 km wide and up to 7 km deep

Current and future missions on Mars

In the beginning, Martian exploration was a dream. It goes back to the time when Wernher von Braun created a plan in 1970 to land humans on Mars by 1981. It may sound like a dream today, but it was the first step. The next big milestone was established by NASA's Space Exploration Initiative in 1989: the US government at that time intended to land humans on Mars in 2019. In Europe, ESA's Aurora Programme in 2001 aimed for 2030 to send manned missions. In 2004, under the presidency of G.W. Bush, there was a presidential 'vision' in the US: return to the Moon by 2020, on the way to Mars (2035). As history suggests nobody can tell when humans will actually land on Mars, but robotic missions are working on this complex problem.

An overview of the Martian missions can be consulted in Figure 1.8 on page 10. There were numerous attempts of several space agencies to reach the Red Planet, either by orbiting satellites or rovers.

Right now, there exists only one active rover mission, namely Mars Exploration Rover B: Opportunity. Another one, Curiosity, is in transit to Mars: it was launched November 26, 2011. Opportunity is active since 2004. It is the remaining rover in NASA's ongoing Mars Exploration Rover Mission. Launched from Earth on 7 July 2003, it landed on the Martian *Meridiani Planum* on 25 January 2004 at 05:05 Ground UTC, three weeks after its twin Spirit (MER-A) touched down on the other side of the planet. Please refer to Figure 1.4 for the sizes of different NASA rovers.

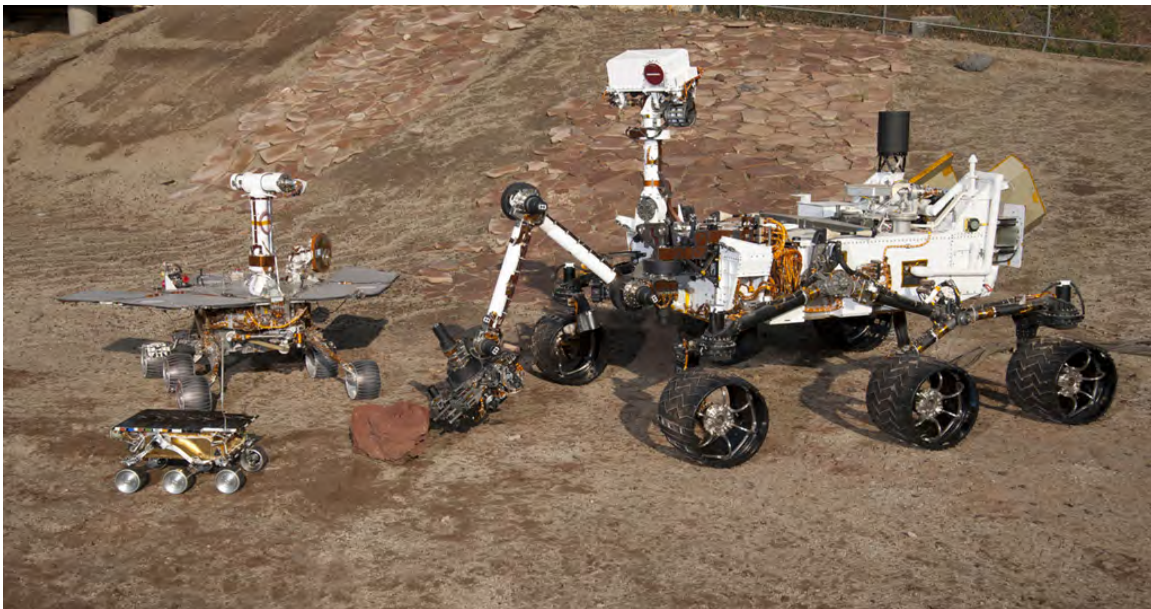


Figure 1.4: Three generation of NASA rovers (in order of increasing size): Sojourner, Spirit and Curiosity (credits: NASA/JPL-Caltech)

There is another mission from NASA which is right now en route to Mars, called the Mars Science Laboratory. The mission is scheduled to land 'Curiosity' rover on the surface of Mars in August 2012, where Curiosity will search for evidence of past or present life on Mars [18].

There are also few planned rover missions: Chang'e 3 [19], the first Chinese lunar rover, and Chandrayaan-II, which is a joint venture between India and Russia, consisting of a lunar orbiter and a lunar lander [20].

The exploration programme of the European Space Agency, is also known as 'Aurora'. In early versions of this ambitious plan, focus will be on planetary bodies that can ultimately be reached by humans, although the first steps of the 30-year programme can only be robotic. An ultimate goal for European astronauts is the participation in the first international mission that will land humans on the planet Mars. To achieve this, ExoMars and NEXT missions are building the base for key technology demonstrations, such as an Earth re-entry vehicle/capsule and Mars aero-capture demonstrator.

ExoMars

ExoMars is the first mission of the 'Aurora' Programme of the European Space Agency. This mission to Mars is currently under development by ESA with collaboration from the Russian Federal Space Agency (Roscosmos) [21]. Since its inception, ExoMars has gone through several phases of planning with various proposals for landers, orbiters, launch vehicles, and international cooperation planning, such as the defunct 2009 Mars Exploration Joint Initiative (MEJI) with the United States. In Figure 1.5 one can see this possible ESA-NASA cooperation plan.



Figure 1.5: Illustration showing the elements that were to make up the original ESA-U.S. ExoMars Program (credits: ESA)

Formerly there were ambitious plans in an international context for planetary explorations, as one can see in Figure 1.6 on page 9. But due to budget cuts and other reasons these are today just history.

The on-surface mission is performed by a near-autonomous mobile robotic vehicle (also referred to as the rover) with a mission design life of 180 sols [22]. The core part of the mission is a Mars Rover capable of acquiring subsurface soil samples from a 2 meter depth. Scientists are trying to answer the ultimate question: whether life is, or had ever arisen, on Mars.

The ExoMars mission's scientific objectives, in order of priority, are [23]:

1. To search for signs of past and present life on Mars;
2. To characterize the water/geochemical environment as a function of depth in the shallow subsurface;
3. To study the surface environment and identify hazards to future human missions;

4. To investigate the planet's subsurface and deep interior to better understand the evolution and habitability of Mars.

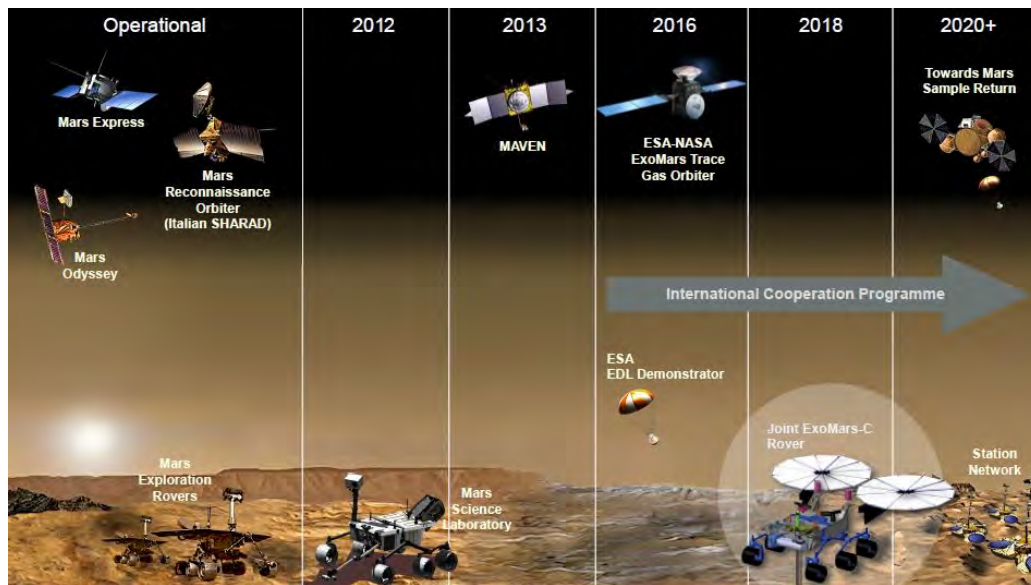


Figure 1.6: Former ESA and NASA collaboration plan (credits: NASA/MEPAG)

ExoMars will combine mobility and access to subsurface locations, where organic molecules may be still alive; thus allowing for the first time to investigate Mars's third dimension: depth. This will be done with drilling and sub-surface sample-taking (Figure 1.7).



Figure 1.7: Artistic impression about ExoMars during rock drilling process (credits: ESA)

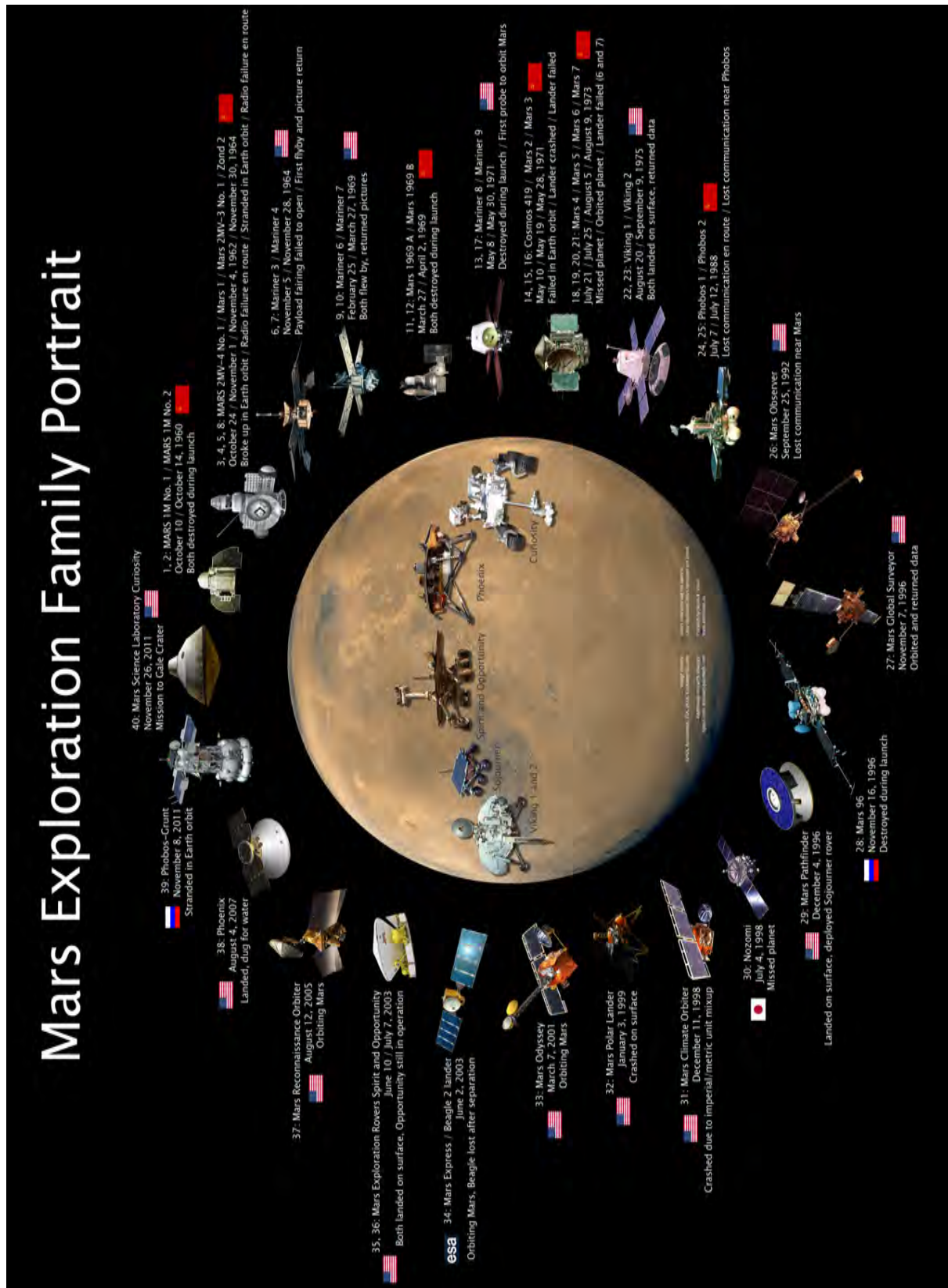


Figure 1.8: Mars exploration overview (credits: Jason Davis [1])

Chapter 2

Martian and Martian-like Terrains

2.1 Characterization

2.1.1 Overview

Mars is a dry planet with surface temperatures well below the freezing point of water. There are no known major bodies of surface water (rivers, lakes, oceans), leaving the terrain barren and similar to deserts on Earth. Conditions are analogous to several locations on the Earth's surface: the dryness, the deep and loose sand, rocky outcrops and plateaus of volcanic lava rock fields.

Thanks to the numerous lander and rover missions sent to explore the Martian surface, data is available on the material composition of the Martian soil at various sites across the planet [22]. Moreover the analysis of the Martian terrain and topography has formed a major part of these missions. Characterization of the composition of the Martian surface is crucial for future mission development, whether in determining the functionality of new instruments and experiments or, more practically, in the design of lander, rover or drill hardware for use on the extreme conditions presented by the terrain types.

The different terrain types have a direct consequence on the rover's path planning. To study, analyze and apply efficient path-planning algorithms it is a must to know these. In Figure 2.1 one can see a monochrome image taken by Spirit, one of the rovers of NASA's Mars Exploration Programme. A bad example of such a path-planning algorithm could follow the edge of the dune in the left side of Figure 2.1. Clearly this could be life threatening to the rover: it could tip over on the edge and cause serious damage.

Martian top surface soil can be characterized by parameters like color, density and compaction which can be used identifying different soil types (e.g. sandy clay, sandy silt and gravels).

Many topographical and terrain analogues can be found in numerous regions on the Earth, including rocky plains similar



Figure 2.1: Lahontan crater on Mars seen from Spirit rover (credits: NASA)

to those observed by the Viking 1 Lander¹ [24]. Similarities include large scale formations, such as tabular lava flows and meteorite impact-like craters, much more localized examples, like rocky till deposits in smaller sloped regions.

On a smaller scale, the discussion in Golombek et al. identifies numerous soil types on the Martian surface (e.g. aeolian deposits) based on Viking, Pathfinder/Sojourner and MER² data [25]. These include drift deposits, having low friction angles³ in the region of 15-21°. These are likely to be atmospheric dust particles of 2-4 μm in diameter, which are found in deposits thick enough to envelope an entire footpad of Viking 1, sinking it 16.5 cm deep. The terrain surroundings of the MER rovers have also been analyzed, measurements of fine grained dust have indicated sizes of up to 45 μm . When observing less fine sand deposits comprising the dunes and bedforms, Opportunity has found them to be relatively dust free, with particle sizes of 130-160 μm [22]. Spirit has also observed similar sand bedforms with particle sizes of approximately 60-160 μm .

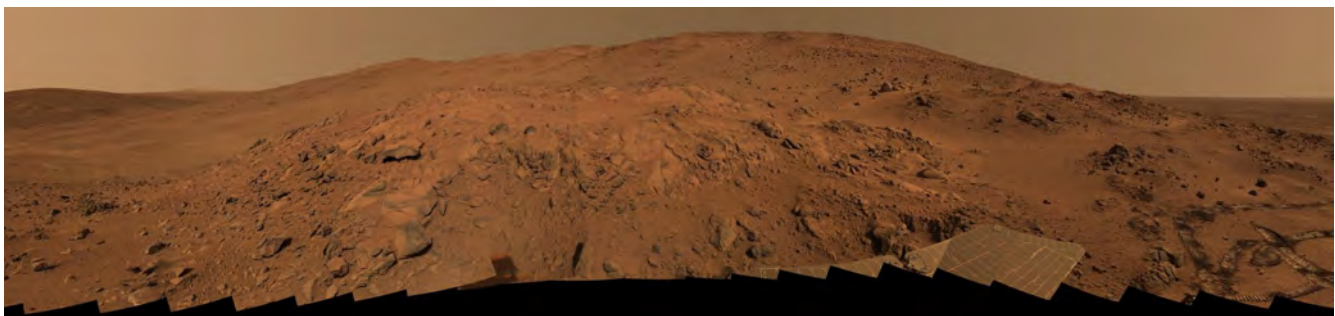


Figure 2.2: 'Lookout Panorama' from Spirit taken on February 27th to March 2nd, 2005 (credits: NASA)

¹First spacecraft to successfully land on Mars and perform its mission

²Mars Exploration Rover

³Critical angle of repose of a granular material

2.1.2 Methods

Terrain characterization is the science of understanding the topography and physical properties of the terrain under investigation. In the context of planetary rovers, terrain characterization provides information about the surface environment in which the rover needs to navigate.

Terrain characterization methods can be classified into two main groups, remote sensing and in-situ sensing methods, as shown in Figure 2.3. In situ sensing can be further subdivided based on whether the measurements are obtained directly or indirectly.

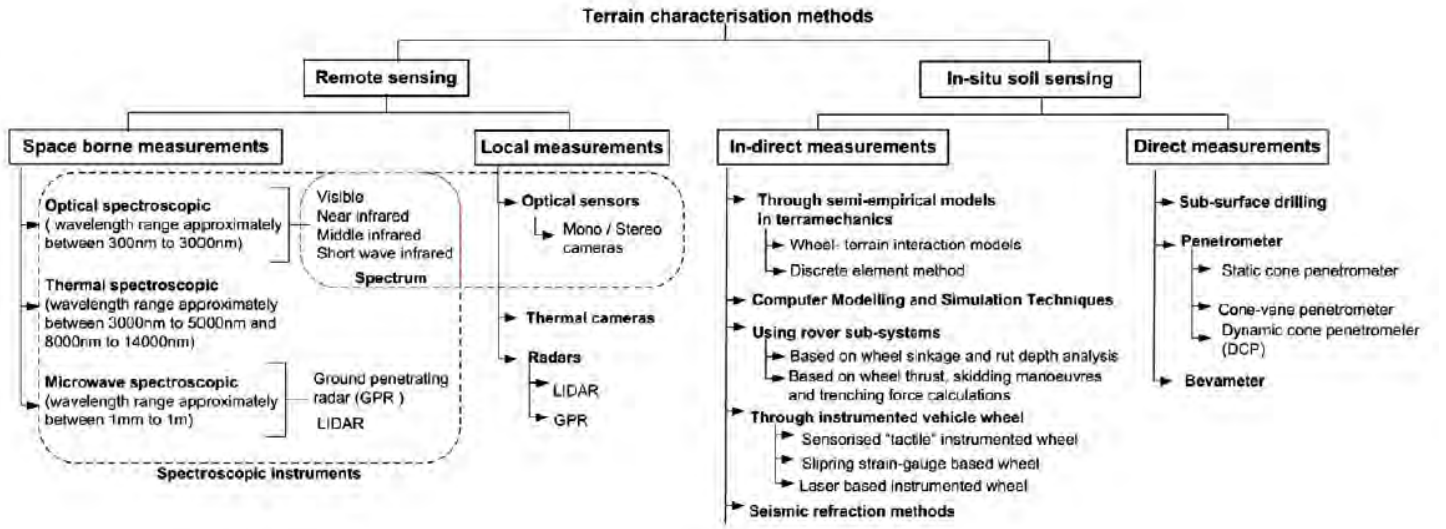


Figure 2.3: Classification of terrain characterization method (Chhaniyara et al. [2])

To date, remote sensing methods have been the primary source of data on the properties of planetary terrains. Since the work completed in the framework of this Master's thesis is a vision based comparison, a particular interest is given to local measurements by means of remote sensing. Optical sensors, like stereo cameras and LIDAR⁴ instruments are used in order to compare the precision of the simulation.

There are some practical drawbacks of optical methods. Firstly, by its very nature, image processing is highly computationally intensive. Secondly, it is well known and commonly observed that the mechanical behavior of sands can change drastically over a very short distance despite the material having a very similar visual appearance [2].

2.2 Surface Hazards for Rover Navigation

Planetary rovers require information about the terrain surface and near surface properties as well as a geometric description of their environment in order to navigate safely. A number of hazardous elements are summarized in the survey written by Chhaniyara et al. [2].

⁴Light Detection And Ranging

Rock-strewn terrain may exceed a rover chassis height and capability to manoeuvre. This hazard can be largely excluded by proper selection of landing sites in combination with on-board autonomous obstacle avoidance.

Soils may exhibit high compressibility and low shear strength as a result of having a low packing density and/or lack of appreciable cohesion. Loose soils can be particularly problematic on sloping ground. This could result in loss of traction as well as the occurrence of excessive sinkage.

Lava flows often produce very irregular surfaces that are extremely difficult to traverse. Moreover these irregularities can become masked through aeolian deposition of sand. Such infill materials are likely to be loose and hence may cause a significant hazard to rover mobility. A similar hazard is likely in terrains characterized by cobble to boulder size rocks which have been covered, partially or completely by wind-blown sand and silt. Moreover, on Mars, aeolian bedforms such as dunes and ripples often consist of low strength soils, as it has been discovered by both MER rovers.

As part of landing site selection for ExoMars, the thermal inertia⁵ measured from orbit provides a good understanding of the strength of the soil [2]. It is believed that loose soil can accumulate over several meters at some places, making landing on Mars impossible.

In order to safely navigate on planetary surface it is necessary that these hazards are identified during path planning. This is why several methods of terrain characterization methods have been developed to address this issue.

2.3 Visual Odometry of Rovers

In this section the concept of visual odometry will be briefly introduced. Rovers need a closed loop trajectory control which is based on rover localization. Figure 2.4 shows a general architecture of a typical rover control loop.

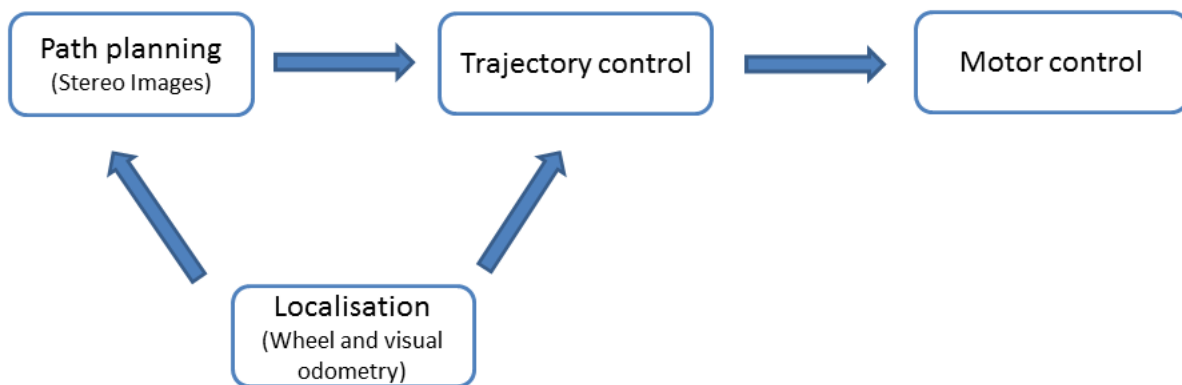


Figure 2.4: Typical rover control architecture

The correct path planning of an autonomous rover system needs to investigate the region to be traversed. The system must be able to perform well and identify rocks or other objects of interest. Correct path planning must be investigated, which then

⁵A measure of that surface's resistance to a change in temperature

gives the way-points to the trajectory control. Finally by means of motor control the locomotion subsystem is taking care of the robot movement.

Visual odometry alone can at most yield a two dimensional path estimate. So in principle it is not sufficient when navigating on a non-planar surface, it must be complemented with other inputs [26]. On the other hand, vision based path estimation is able to transmit accurate results while being intrinsically 3D.

A typical rover visual odometry procedure would look like this [26]:

1. **Extract Features from Images:** Detects objects of interest and provides a characterization of their properties. This both locates rocks in the images and extracts rock properties (features) including shape, texture and albedo.
2. **Analyze and Prioritize Data:** Uses the extracted features to assess the scientific value of the planetary scene and to generate new science objectives that will further contribute to this assessment.
3. **Plan and Schedule New Command Sequence:** Enables dynamic modification of the current rover command sequence (or plan) to accommodate new science requests from the data analysis and prioritization module. A continuous planning approach is used to iteratively adjust the plan as new goals occur, while ensuring that resource and other operation constraints are met.

There are a many important aspects in the field of visual rover odometry. Certainly a few are very important regarding navigation, like low matching error, repeatability of the feature points, and finally high number of detected features. These help in an accurate location estimate of planetary rovers.

It is clear that the key issue to get reliable and accurate results for estimation is the way image features are detected and represented.

2.4 Testing on the Earth

Testing is one of the key aspects in space applications. This is especially true in the field of space robotics, where robotic instruments can only be tested on Earth. This implies that testing phases always face a challenge, no matter if it is a well known space application or a novel approach to a problem. Analogues may be selected to replicate properties including chemical, magnetic, thermal, physical, mechanical, or organic behavior simulations [27].

Mechanical soil simulants for Martian testing have been primarily sourced by NASA JPL from two locations, the Hawaiian volcanic regions and the Mojave Desert. The data provided by the Martian rover missions informed of a subsequent simulant with mechanical parameters closer to the MMS⁶ [22]. Other simulants are in use in the various laboratories situated at JPL. These include [?]:

- crushed volcanic rock, used in the MER yard;
- a decomposed granite and brick dust mixture, used in the more general purpose Mars Yard;
- a dust free and washed silica sand in the JPL Lab;
- a dust free garnet mix in JPL Lab.

⁶Mojave Mars Simulant

ExoMars rover traction testing has been performed on various dust-free washed and dry quartz sand in the Astrium Mars Yard [28]. The wheel testing has been performed with several other types of simulant.

One needs to understand that for rover testing the use of unhealthy materials, i.e. fine dust, is undesirable. Critical points of simulating a Martian surface for testing on Earth are found to be:

- The uniformity of the surface material simulant. Such a uniform material distribution is not necessarily present on the Martian landing site and region to be investigated.
- The effect of lower gravity on Mars, which is difficult to reproduce in a terrestrial laboratory.
- The exact chemical reproduction of the soil constituents.

2.4.1 Planetary Terrain Testbed

The laboratory work of this thesis described in Chapter 3 was performed using the testbed at the Automation and Robotics Laboratory - ESTEC site. The size of this indoor testbed is 8.7 x 8.6 m. It contains several materials such as pozzolana soil, basalt grains of various sizes, gypsum and 'Terre Rouge du Royans'. This is to imitate as much as possible the Martian surface (see Figure 2.5).

Pozzolana, also known as pozzolanic ash⁷, is a fine volcanic ash. Pozzolana is a siliceous and aluminous material which reacts with calcium hydroxide in the presence of water. Pozzolanic reaction stands a simple acid-base reaction between calcium hydroxide ($Ca(OH)_2$), and silicic acid (H_4SiO_4 , or $Si(OH)_4$) [29].



Figure 2.5: Martian terrain constituents (left to right): basalt rock with 'terre rouge'; basalt grains; Martian rock like gypsum

Historically the testbed was used as a lunar type surface for lunar vehicle⁸ testing. In the present state it is more Martian like due to the reddish loam 'Terre rouge du Royans'. It contains 10% organic matter, 25% clay and silica. The fake rocks made by gypsum are covered partially by reddish dust. These are present in order to simulate Martian basalt rocks. Basalt is a common extrusive volcanic rock on the surface of planets. Defined as an aphanitic igneous rock that contains, by volume, less than 20% quartz and less than 10% feldspathoid. It is usually grey to black and fine-grained due to the fast cooling of lava at the surface of a moon or planet.

⁷pulvis puteolanus in Lat.

⁸xLuna

2.4.2 Overview of 3DROV

With the growth of computing capability over the past years, employing numerical modeling methods to analyze engineering problems has become a standard practice. The benefits of the numerical modeling are multiple: the representation of reality by means of complex scenarios, the interaction between different systems, or an initial virtual modeling at system or subsystem level gives a great understanding.

A newly-completed software tool, known as the 3DROV developed by TRASYS⁹, for planetary rover simulation environment allows early-stage virtual modeling and 'shakedown' testing of mobile surface missions for Mars.

The driving philosophy behind the 3DROV project is to build a tool for rover system design and simulation that supports the system design process and provides virtual verification [30]. Moreover it allows ESA with the means to effectively support the development of European planetary exploration missions - the planetary rover missions are fairly new to both ESA and to the space industry in Europe.

3DROV is an end-to-end rover mission simulation environment offering the necessary virtual infrastructure to test and evaluate robotic systems for planetary exploration and generate early design feedback. A modular simulation environment has been developed incorporating all the essential elements as it can be seen in Figure 2.6.

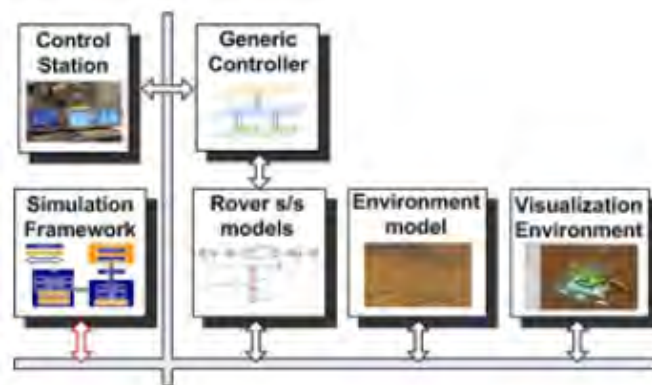


Figure 2.6: 3DROV building blocks

3DROV is built-up by several building blocks. The Control Station serves as the rover's ground control station; the Generic Controller assumes the role of the onboard flight software; the Environment component is responsible for ephemeris and time-keeping of the system, for generating the target environment atmospheric conditions and implements GIS for geographic, terrain and spatial data; the Visualization Environment provides real-time visualization of the simulation run. Finally, the Simulation Framework, based on the ESA-EGOS¹⁰ SimSat¹¹, is responsible for the execution of all the models and the scheduling of the simulation run.

3DROV's visualization environment renders in a photo-realistic virtual scene the simulated rover and its surrounding environment. This will be the tool to test the accuracy of the digitized terrain versions in accordance with the ARL testbed.

⁹Belgian IT consultancy and services company

¹⁰ESA Ground Operation System

¹¹SimSat is developed by ESA/ESOC in cooperation with European industry

2.4.3 EXOMADER

The EXOMADER is a rover built at the ARL in order to support the R&D activities for different projects (e.g. ExoMars). The rover features the so-called model-D chassis which exhibits extremely advanced cross-country ability, built by the Russian company RCL - Transmash, and fitted with avionics developed in-house designated for the transportation of an experimental 5 kg payload.

The chassis is a six-wheeled driven type, with four steerable wheels (6 x 6 x 4). The wheels are individually driven by ten motors, six of which are for driving and are equipped with optical encoders, whereas the other four are used for steering, and are provided with angular potentiometers. Please refer to Figure 2.7(a), to see the actual ExoMader on the testbed.

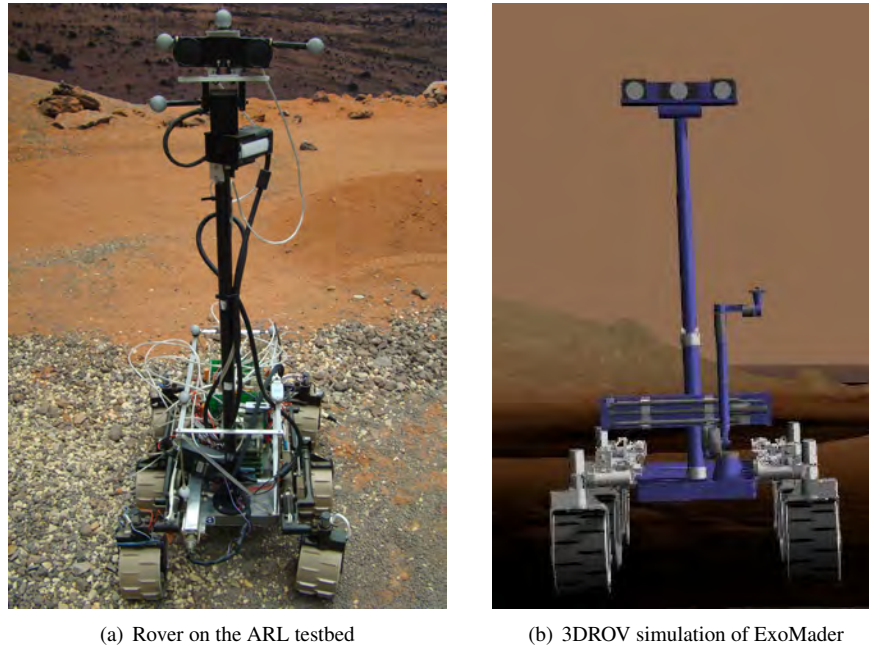


Figure 2.7: ExoMader custom built rover in the ARL

The EXOMADER control system, structured according to the MORCA¹² architecture, is implemented on a Linux-RTAI¹³ operating system. The Navigation Layer is a modified version of the CNES¹⁴ rover navigation system.

Characteristics	Value	Units (SI)
Mass	9,5	kg
Payload mass	5	kg
Size	0,652 x 0,575 x 0,255	m x m x m
Velocity with payload	0,05	m/s
Angle of static stability	45	°
Max. climbing slope (without/with payload)	20 / 15	°

Table 2.1: EXOMADER chassis specifications

¹²MOBILE Robot Control Architecture

¹³Real-Time Application Interface

¹⁴Centre National d'Études Spatiales

Chapter 3

Digitalization of the Planetary Testbed

Because of recent developments in computer vision and sensor technology, light has been used in a number of ways to measure objects. These measuring techniques can be divided into two categories: active and passive techniques. Both techniques will be used in the framework of this Master's Thesis.

Active measurement techniques incorporate scanners. These emit controlled radiation and detect its reflection in an environment. LIDAR is an optical remote sensing technology that can measure the distance by illuminating the target with light, using pulses from a LASER¹. LIDAR² technology has application in geomatics, archaeology, geography, seismology, remote sensing and atmospheric physics.

Passive techniques do not emit any kind of radiation themselves, but instead rely on detecting reflected ambient radiation. Passive methods can be very low-cost, because in most cases they do not need any particular hardware other than a digital camera. The problem with these techniques is that they rely on finding correspondences between 2D images, which do not always have unique solutions. For example, repetitive patterns tend to 'fool' the measurement method. The accuracy of these methods depends mostly on the resolution of the imaging system and the density of identifiable features in the image. Although one can obtain nice results with this method, as it will be shown in Chapter 6.

3.1 Active Technique

3.1.1 Laser Scanning

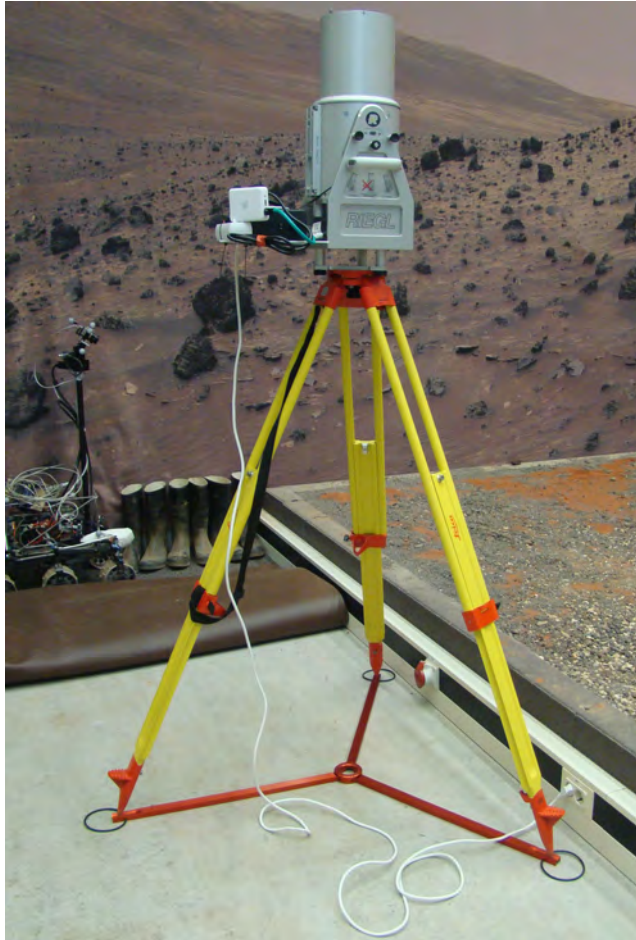
The major part of current applications of terrestrial laser scanning systems (TLS) require the geo-referencing of each 3D-view, which is normally performed by means of targets (or natural features) as ground control points (GCPs). In literature, some alternative methods are proposed to accomplish this task. For accurate data acquisition the number of GCPs must be in

¹Light Amplification by Stimulated Emission of Radiation

²Light Detection And Ranging

accordance with the size of the terrain, in order to insert the different point clouds into the ground reference system.

During the work performed a TLS (Figure 6.8(a)) was used in order to map in three dimension the Martian like robotic testbed (see Section 5.2.1). The system provides wide field-of-view, high accuracy, and fast data acquisition. The benefit of the current TLS is that the software (RiSCAN PRO) can instantly acquire a three dimensional point cloud and post-process the data.



(a) RIEGL Laser Measurement System on a Leica tripod, secured with the tripod star



(b) FISCO Solatronic Inclinometer

Figure 3.1: Terrestrial laser scanning equipment set-up

Background Information

Electromagnetic waves are typically described by any of the following three physical properties: the frequency, wavelength, or photon energy. Waves in the electromagnetic spectrum vary in size from very long radio waves to very short gamma-rays (see Figure 3.3). Wavelength is inversely proportional to the wave frequency, so gamma rays have very short wavelengths that are fractions of the size of atoms, whereas wavelengths can be as long as the universe. Photon energy is directly proportional to the wave frequency, so gamma rays have the highest energy (approximately a billion electron volts) and radio waves have very low energy (around a femto electron volts).

$$c = f \cdot \lambda; \quad E = h \cdot f = \frac{h \cdot c}{\lambda} \quad (3.1)$$

Where $c = 299,792,458$ m/s is the speed of light in vacuum and $h = 4.13566733(10) \cdot 10^{-15}$ eV is Planck's constant.

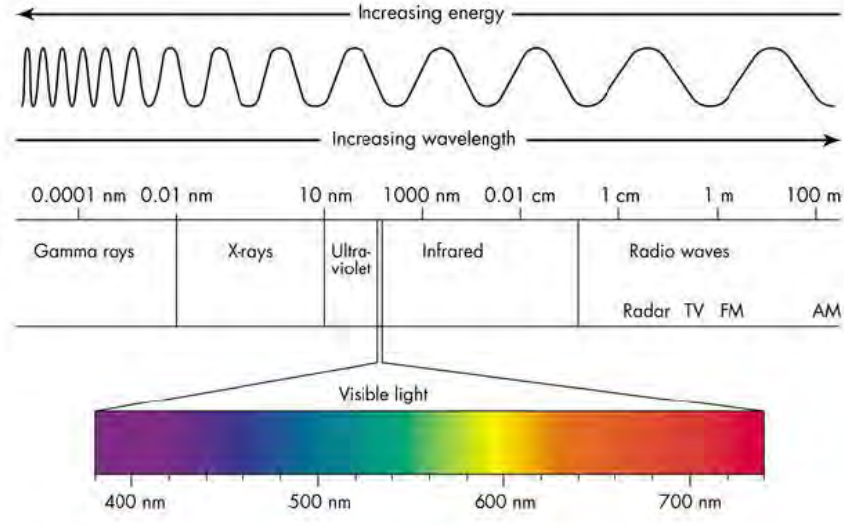


Figure 3.2: The electromagnetic spectrum (source: [3])

A device that is able to generate a wave of light using only a very narrow band of the spectrum is called a laser. A typical laser emits light in a narrow, low-divergence beam with a well-defined wavelength. The beam has a high degree of spatial coherence, therefore it propagates dominantly in a certain direction [31].

Incoherent Detection

The technology used for the laboratory work as a part of this Master's thesis work is a surface imaging one. This is based upon accurate distance knowledge by means of electro-optical measurements and a two axis beam scanning system [32]. 3D images are based on a number of independent laser distance measurements, in different, but well defined angular directions.

Light waves travel with a finite and constant velocity in a certain medium. Therefore, the time delay created by light traveling from a source to a target surface and back to the source can be measured (round trip). The distance to the target surface that shall be investigated can be evaluated using the formula:

$$D = c \cdot \frac{t}{2} \quad (3.2)$$

Where D is distance, c the speed of light and t the time of travel.

The principle of the distance meter is the time-of-flight measurement of short infrared laser pulses. A laser source emits pulses in IR³ region, afterwards the echo signal is reflected back from the target and hits a photodiode. This generates a signal in

³Infrared

a photo-detector, which is used in the calculation of the time interval between the transmitted and received pulse. To accurately know the time period, a quartz-stabilized clock frequency is necessary.

The 3D images are composed by a frame (number of lines), and each line is composed of a number of measurements.

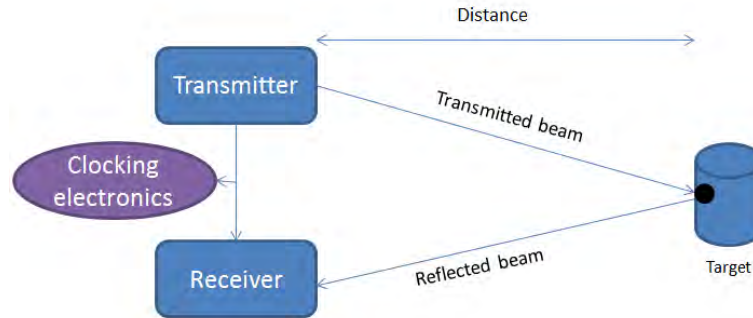


Figure 3.3: Time-of-flight laser scanner principle

The use of pulses for laser ranging constitutes an advantage by means of high power transmittance. This power makes it possible to achieve the required SNR (signal-to-noise ratio) needed for high accuracy measurements. Generally speaking, the disadvantage is the challenge of detecting the exact arrival time of the backscattered laser pulse due to the changeable nature of optical threshold and atmospheric attenuation. This was not the case, since the ARL is located inside the building, moreover constant temperature and light sources are present.

Equipment

During the laser measurement campaign the following instruments were used:

- RIEGL Laser Measurement System Z210i
- Leica wooden tripod and tripod star
- FISCO Solatronic Inclinometer EN17

Although this specific LMS-Z210i device (see Figure 3.4) is only a Class I laser scanner, emitting at near infrared wavelength, safety regulations were kept. According to specifications: *No individual, regardless of exposure conditions to the eyes or skin, would be expected to be injured by a Class I laser. No safety requirements are needed to use Class I laser devices. Although it is not recommended that the instrument is directly aimed at people's eyes.*

In order to implement the safety rules precautions were taken (e.g. restrictive access in the lab while using the equipment). Also direct eye contact with the device was avoided.

As one can see in Figure 6.8(a), the scanner is attached to the tripod, and secured with the tripod star, this was necessary due to the slippery surface. The top part, where the scanner was attached, was set to be horizontal with the help of the inclinometer (see Figure 6.8(b)). A custom-made metallic plate is screwed to the laser base in order to hold a power supply and a socket adapter. At the same time, an Apple Airport Express is connected to the system to transfer the data via Wi-Fi to the workstation.

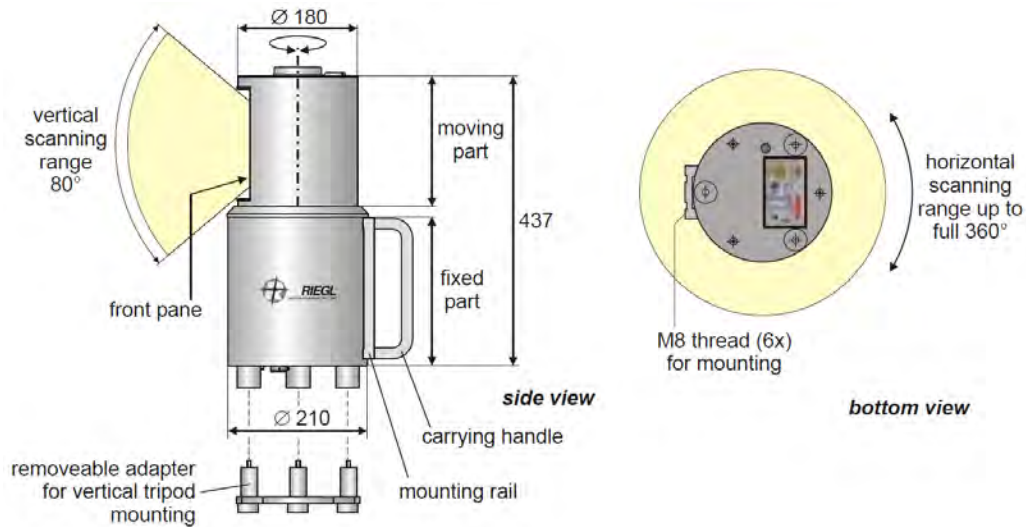
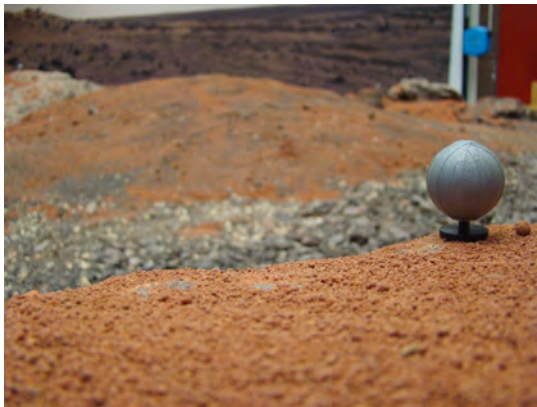


Figure 3.4: Dimensional drawing of the LMS-Z210i TLS (Source: [4])

The Z210i laser system emits light at $0.9 \mu\text{m}$, which corresponds to IR range. The beam laser beam divergence is 3 mrad^4 . The laser scanner accuracy is typically around $\pm 2,5 \text{ cm}$ for a single shot [4]. For the purpose of a precise digital reconstruction of the terrain this gives sufficient precision.

3.1.2 Motion Tracking

The Vicon system measures the 3D position of retro-reflective markers, which are spheres made of reflective material. The diameter of such a sphere is approximately 26 mm. These objects scatter back the light to the source, unlike normal objects which produce incident reflections. Once the coordinates of these markers from multiple cameras are obtained, a three-dimensional positioning can be performed. Each circle represents a one dimensional solution that emanates from a camera location projected out into the captured volume. This line of solution is referred to as the camera ray.



(a) A marker



(b) The position of the markers seen in flashlight

Figure 3.5: Retro reflective markers

⁴1 mrad corresponds to 100 mm beam width per 100 m distance

In the system setup, at least two cameras must see the marker in order to locate in 3D. To accurately find the 3D coordinates, it is necessary to know the location and orientation of the cameras. Without this, the system can not precisely triangulate the location.

A higher number of cameras offers better visibility of the tie points spread on the surface and therefore better accuracy in the position determination.

Equipment

The Vicon MX architecture in the Automation and Robotics Laboratory consists of several MX cameras, an MX Ultramet, and a workstation. During the measurement campaign a number of 8 MX13+ Vicon cameras were used (Figure 3.6). Please refer to table below for full performance specification [33].

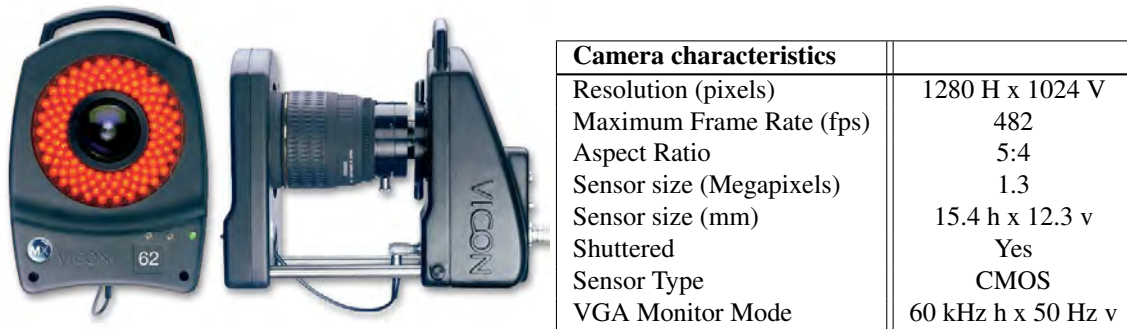


Figure 3.6: Vicon MX13+ and specification

The distribution of the Vicon cameras can be seen in Figure 3.7. The active ones are green, although one can observe more than eight devices. Those extra ones were present but not used, since eight cameras are sufficiently covering the testbed.

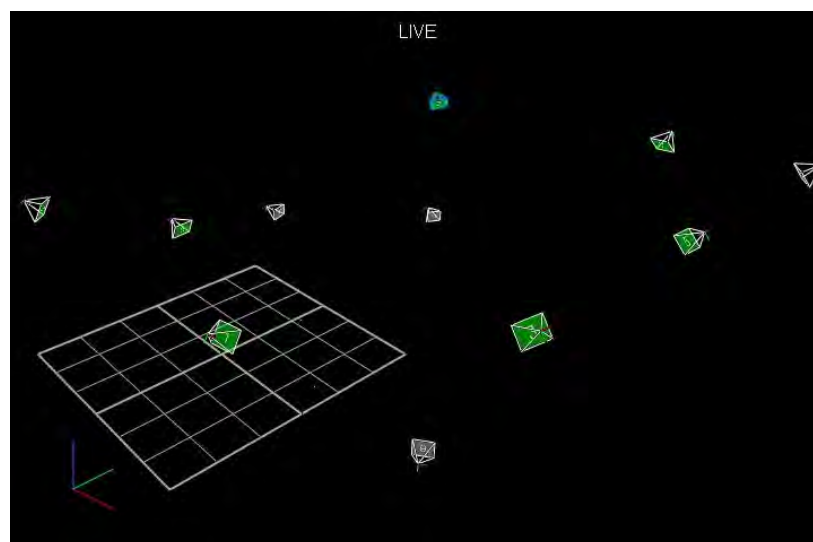


Figure 3.7: Spatial distribution of the Vicon MX13+ cameras

3.2 Passive Technique

In this section a state of the art 3D reconstruction technique will be presented, developed at the Center for Machine Perception (CMP), in the Czech Technical University in Prague. This was used in order to create the DEM by means of multi-view stereo reconstruction and later import it into the simulation software, 3DROV.

According to recent studies approaches to multi-view reconstruction attain a degree of accuracy and completeness comparable to laser scans. This will be analyzed and discussed in Chapters 6. The degree of accuracy based on the simulation and reality will be investigated.

3.2.1 Classification of Multi-View Reconstruction Algorithms

Multi-view stereo algorithms can be roughly categorized into four classes [34]:

- The first class operates by first computing a cost function on a 3D volume, and then extracting a surface from this volume (e.g. voxel coloring algorithm). Others differ in the definition of the cost function and the surface extraction method.
- The second class of techniques works by iteratively evolving a surface to decrease or minimize a cost function. These techniques use methods based on voxels, level sets, and surface meshes. Level-set techniques minimize a set of partial differential equations defined on a volume. Other approaches represent the scene as an evolving mesh that moves as a function of internal and external forces.
- In the third class are image-space methods that compute a set of depth maps. To ensure a single consistent 3D scene interpretation, these methods enforce consistency constraints between depth maps.
- The final class consists of algorithms that first extracts and matches a set of feature points and then fits a surface to the reconstructed features.

Advantage of Multi-View Reconstruction

Producing complete reconstructions of outdoor and complicated scenes is still a challenge. Most of the state-of-the art multi-view reconstruction methods produce a 3D point cloud, which is later used to compute a mesh representing the 3D world. There are several approaches to computing the point cloud, such as plane-sweeping based, stereo based, or growing based methods.

The novel method proposed in [6], delivers a 3D reconstruction of objects purely from images. The multi-view dense 3D reconstruction tool implemented at CMP can be used to create colored or textured surface models of the scenes from perspective data sets once the cameras are calibrated.

Automatic reconstruction from multiple view imagery already is a low-cost alternative to laser systems as expressed by the authors, but could become a total replacement once the geometrical accuracy of the results can be proven. This is one of the many aspects of the present master thesis work.

3.2.2 Theoretical Background

In this subsection, the Delaunay triangulation will be introduced in order to provide better understanding of the theory of multi-view reconstruction. Also an overview is provided of the reconstruction technique used in [6]. The following is taken from [35].

Let $\mathcal{P} = p_1, p_2, \dots, p_n$ be a set of points in \mathbb{R}^d . The Voronoi cell associated to a point p_i , denoted by $V(p_i)$, is the region of space that is closer from p_i than from all other points in \mathcal{P} :

$$V(p_i) = \{p \in \mathbb{R}^d : \forall j \neq i, \|p - p_i\| \leq \|p - p_j\|\} \quad (3.3)$$

$V(p_i)$ is the intersection of $n - 1$ half-spaces bounded by the bisector planes of segments, $V(p_i)$ is therefore a convex polytope, possibly unbounded. The Voronoi diagram of \mathcal{P} , denoted by $V(\mathcal{P})$, is the partition of space induced by the Voronoi cells $V(p_i)$.

The Delaunay triangulation is defined as the geometric dual of the Voronoi diagram. There is an edge between two points p_i and p_j in the Delaunay triangulation if and only if their Voronoi cells $V(p_i)$ and $V(p_j)$ have a non-empty intersection. It yields a triangulation of \mathcal{P} , that is a partition of the convex hull of \mathcal{P} into d -dimensional simplices (i.e. into triangles in 2D, in tetrahedra in 3D). Figure 3.8 displays an example of a Voronoi diagram and its associated Delaunay triangulation in 2D.

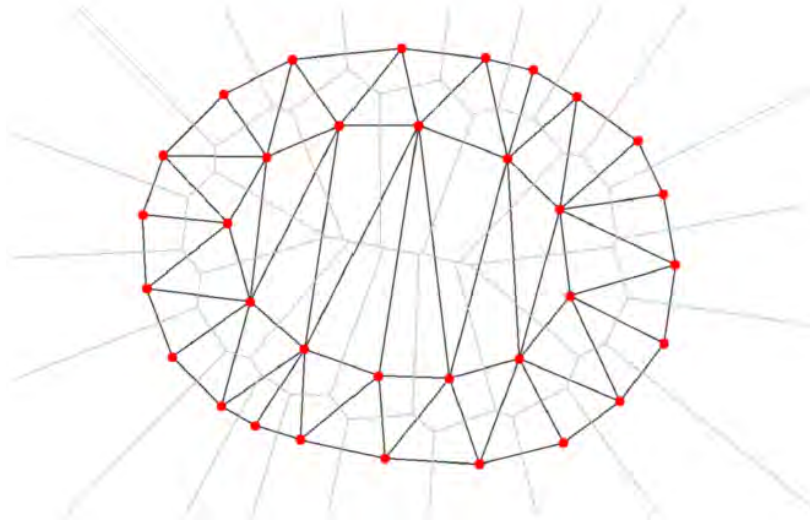


Figure 3.8: Voronoi diagram (gray edges) of points and its associated Delaunay triangulation (black edges) (showed in [5])

The work of [6] is based on the minimal s-t cut of a graph derived from the Delaunay tetrahedralization of the input 3D point cloud in order to label tetrahedra as inside or outside. Their approach casts the surface reconstruction as an energy minimization problem that can be globally optimized. The new novel element in [6] is in computing the free space support, then detecting the highly-supported-free-space boundary and finally changing the t-weights of the graph to reconstruct even the weakly supported surfaces.

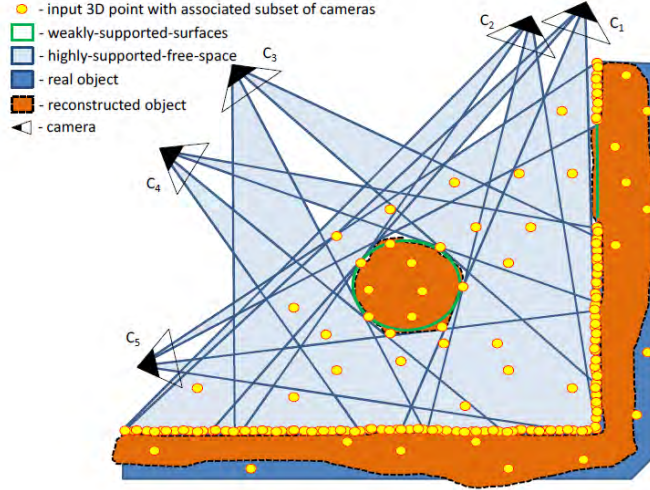


Figure 3.9: Weakly-supported surfaces (credits: M. Jancosek [6])

In Figure 3.9, a highly-supported-free-space is defined as part of the space which is between the surfaces densely sampled by the input 3D points and the associated cameras. Weakly-supported surfaces (green) are the surfaces of the real object which are weakly sampled by the input 3D points and are close to the border of the union of the highly-supported-free-space cones.

This method is built on what the authors call as *base-line method*. The free-space-support of the tetrahedron is the sum of weights of all incoming edges to the node corresponding to the tetrahedron. The free-space-support is related to the total number of pairs (camera, point) for which the point is occluded by the tetrahedron in the camera [6]. The authors point out that they use different α_{vis} value for each 3D point, although in the original base-line method this value is constant.

Let us denote the number of cameras associated with a point $x \in X$ from the input 3D point cloud as $N_c(x)$. In this method it is considered the grouping of points $p \in T$ of the tetrahedralization such that $S(p) \subset X$ within a distance γ surrounding of $p \neq \emptyset$ and $\forall p_i, p_j \in T; |p_i - p_j| > \gamma$. The $\alpha_{vis}(p)$ is defined in [6] as follows:

$$\alpha_{vis}(p) = \sum_{x \in S(p)} N_c(x) \quad (3.4)$$

The value of $\alpha_{vis}(p)$ must be small enough for a noisy point p , where the neighborhood γ contains no other points. Consequently a good situation arises when a point is supported from more camera positions (each point has a set of cameras and the $\alpha_{vis}(p)$ value associated).

When adding a new point x from the point cloud, the algorithm first checks if there is a 3D point p of the actual tetrahedralization within a distance γ from x . If there is such a point then we do not add x to the tetrahedralization but associate the camera with the nearest point p and increase $\alpha_{vis}(p)$ by 1. Otherwise add the point p to the tetrahedralization, associate the camera with it, and initialize an $\alpha_{vis}(p)$ to 1.

The multi-view reconstruction pipeline is divided into two modules. The first module creates a depth map for each camera.

A 3D point cloud can be computed for each depth-map using a related camera matrix. Later all 3D point clouds from all depth-maps are merged into one 3D point cloud, where the camera from which the point is computed is related to each 3D point. This point cloud is used as the input to the second module, which constructs a triangulated surface.

3.3 Lens Correction

Many inexpensive camera lenses, especially zoom lenses, suffer from distortion. In order to project images on the DEM created by laser scan, one should overcome this issue. One reason is that 3DROV in the current status does not model lens distortion. It will be shortly described the camera distortion effect, the solution and the outcome: the undistorted images.

The pinhole camera model (the most simplistic camera model, see Figure 5.8(a)) describes the mathematical relationship between the coordinates of a 3D point and its projection onto the image plane of an ideal pinhole camera, that is: $sm' = A[R|t]M'$. Or expanded [36]:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (3.5)$$

Where:

- (X, Y, Z) are the coordinates of the 3D point in the world coordinate space,
- (u, v) are the coordinates of the projection point in pixels,
- A is called camera matrix, or it is also common the notion of matrix of *intrinsic parameters*,
- (c_x, c_y) is the principal point (usually at the image center),
- (f_x, f_y) are the focal lengths expressed in pixel-related units.

Thus, if an image is scaled by some factor, all of these parameters should be scaled (multiplied, respectively divided) by the same factor. The camera matrix does not depend on the scene viewed and, once estimated, can be re-used (as long as the focal length is fixed). The joint rotation-translation matrix $[R|t]$ is called the matrix of *extrinsic parameters*. It is used to describe the camera motion around a static scene, or vice versa. Translates the coordinates of a point (X, Y, Z) to some coordinate system, fixed with respect to the camera. If $z \neq 0$ then the transformation is equivalent to:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = R \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + t \quad (3.6)$$

$$x' = x/z \quad (3.7)$$

$$y' = y/z \quad (3.8)$$

$$u = f_x \cdot x' + c_x \quad (3.9)$$

$$v = f_y \cdot y' + c_y \quad (3.10)$$

Real lenses usually have some distortion, mostly radial distortion and slight tangential distortion. So, the above model is extended with the equations [36]:

$$x'' = x'(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + 2p_1 x' y' + p_2 (r^2 + 2x'^2) \quad (3.11)$$

$$y'' = y'(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + p_1 (r^2 + 2y'^2) + 2p_2 x' y' \quad (3.12)$$

Where:

- $r^2 = x'^2 + y'^2$,
- k_1, k_2, k_3 are radial distortion coefficients,
- p_1, p_2 are tangential distortion coefficients.

Higher order coefficients than two will be not considered.

The lens mapping $L(r)$ is used to describe how a lens is distorting an image. In [37] it is defined the mapping for a fixed object plane distance as follows: a point in the object plane at distance r from the optical axis is imaged by the lens to a point at distance $L(r)$ from the axis in the film plane. A lens with no distortion would have $L(r) = r$.

Transverse magnification at a distance r from the optical axis is given by $M = dL/dr$. Barrel distortion results when the magnification decreases with the off-axis distance r , that is, $dM = dr < 0$. On the other hand pincushion distortion is the result of magnification increasing with the off-axis distance, $dM = dr > 0$. It is possible to have both types of distortions in a lens.

To overcome the distortion problem, the distortion function is introduced by Ojanen [37]: $U^{-1} = L$, conveniently $U(r) = r + \Delta(r)$.

If $\Delta(r) = c_2 r^2 + c_3 r^3 + \dots$, then we have $U(r - \Delta(r)) = r - c_2 r^2 + c_2 (r - \Delta(r))^2 + O(r^3) = r + O(r^3)$.

Thus we get the formulae in terms of $\Delta(r)$ for the lens function and magnification: $L(r) \approx r - \Delta(r)$ and $M(r) \approx 1 - \Delta'(r)$.

Chapter 4

Image Processing

In this chapter discussions on various image analysis techniques are presented which will be useful in the Analysis part of this Master's thesis.

4.1 Optical Methods of Terrain Characterization

The on-board navigation ability of the Martian rovers Spirit and Opportunity has been enhanced and fully used over the course of surface operations. Efforts were concentrated on the improvement of stereo visual odometry¹ to measure rover slippage [2]. Mars rovers carry improved image-based terrain characterizing algorithms, which can identify rocky and sandy areas, slopes and dangerous areas.

The optical methods of terrain characterization have, of course, drawbacks. By its very nature, image processing is highly computationally intensive, this is why even today, monochrome images with low resolution are used for navigation.

For planetary rover applications, the main challenges are high power consumption, the overall system mass requirements and the substantial data processing burden, making a sophisticated compression algorithm essential for real-time application.

4.1.1 Stereoscopy

Stereoscopy or stereo imaging refers to a technique for creating or enhancing the illusion of depth in an image by presenting two offset images separately to the left and right eye of the viewer. One can refer to Figure 4.1 as an example.

In practice, stereo imaging involves four steps when using two cameras [38]:

1. Mathematically remove radial and tangential lens distortion, this is called undistortion.
2. Adjust for the angles and distances between cameras, a process called rectification.
3. Find the same features in the left and right camera views, a process known as correspondence.

¹ Vision based path-estimation

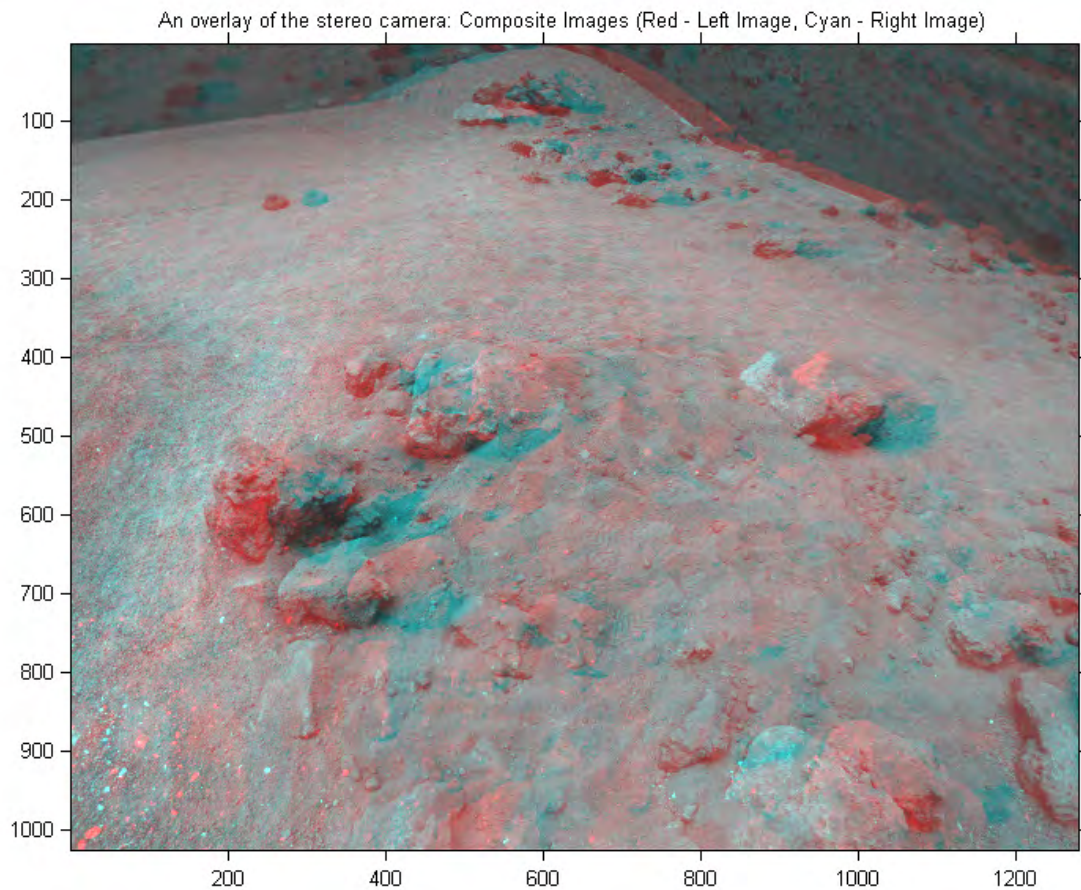


Figure 4.1: Composite image of the Exomander rover (red - left image, cyan - right image)

4. If the geometric arrangement of the cameras is known, it can be turned into a disparity map by triangulation. This step is called re-projection.

4.1.2 3D Imaging

As a matter of fact 3D imaging means that one can retrieve the depth coordinate which got lost by the projection of the object onto an image plane. There exists two major categories in the field of 3D imaging: in the first category one can retrieve only the depth of a surface in the three dimensional space, the second allows the full reconstruction of 3-D objects. Stereo imaging (e.g. human eye) falls into the first category. In latter category falls tomographic reconstruction, scanning light sheet techniques and confocal laser scanning microscopy.

4.1.3 LIDAR

LIDAR scanners, in contrast to optical stereo camera methods, collect considerably more data, are robust in harsh lighting conditions, and derive the 3D information directly, requiring no image processing to extract it. However, LIDAR-derived terrain models cannot provide information on the trafficability of the soil substrate of the terrain in question.

4.2 Feature Points

The current section will provide a basic understanding of feature points, with the description of the most popular feature detectors and descriptors.

In computer vision, the concept of feature points, also called key-points or interest points, has been largely used to solve many problems in object recognition, image registration, visual tracking, and 3D reconstruction. The idea behind it is that instead of looking at the image as a whole, it could be advantageous to select some special points in the image and perform a local analysis on these.

These approaches work well as long as a sufficient number of points is detected in the images of interest and these points are distinguishing and stable features that can be accurately localized. This subchapter will introduce the problem in detail.

4.2.1 Local features

A local feature is an image pattern which differs from its immediate neighborhood. It is usually associated with a change of one or more image properties, although it is not necessarily localized exactly on this change. The most common properties considered are intensity, color, and texture.

Generally speaking local features can be edges, but also points or small image patches. Typically, some measurements are taken from a region centered on a local feature and converted into descriptors. The descriptors can then be used for various applications. In Figure 4.2 one can see the importance of local features for humans: the left column shows the original intact versions of the objects; the middle column shows the still recoverable versions - contours have been deleted, and replaced through collinearity; the right column shows the nonrecoverable versions.

Properties of the Ideal Local Feature

Good features should have the following properties according to the survey made by Tuytelaars et al. [8]:

- *Repeatability*: Given two images of the same object or scene, taken under different viewing conditions, a high percentage of the features detected on the scene part visible in both images should be found in both images.
- *Distinctiveness/informativeness*: The intensity patterns underlying the detected features should show a lot of variation, such that features can be distinguished and matched.
- *Locality*: The features should be local, so as to reduce the probability of occlusion and to allow simple model approximations of the geometric and photometric deformations between two images taken under different viewing conditions (e.g., based on a local planarity assumption).
- *Quantity*: The number of detected features should be sufficiently large, such that a reasonable number of features are detected even on small objects. However, the optimal number of features depends on the application. Ideally, the number of detected features should be adaptable over a large range by a simple and intuitive threshold. The density of features should reflect the information content of the image to provide a compact image representation.
- *Accuracy*: The detected features should be accurately localized, both in image location, as with respect to scale and possibly shape.
- *Efficiency*: Preferably, the detection of features in a new image should allow for time-critical applications.



Figure 4.2: Importance of corners and junctions in object recognition (Source: [7])

4.2.2 Feature Detectors

The concept of feature detection refers to methods that aim at computing abstractions of image information and making local decisions at every image point whether there is an image feature of a given type at that point or not. The resulting features will be subsets of the image domain, often in the form of isolated points, continuous curves or connected regions.

Feature detection is a low-level image processing operation. That is, it is usually performed as the first operation on an image, and examines every pixel to see if there is a feature present at that pixel. If this is part of a larger algorithm, then the algorithm will typically only examine the image in the region of the features.

In general, (point) feature extraction consists of two steps [26]:

- **Detection** - aiming at the localization of visually salient points in the image, and at the determination of the apparent size (scale) of the visual feature.
- **Description** - aiming at representing in a compact form the image behavior in the vicinity of the detected point. This representation typically consists of a fixed number N of values, that can be interpreted as points in an Euclidean space E^N . This way, the similarity of features from different images can be easily computed from Euclidean distance of their representations.

Usually, detectors act by searching the image for the extrema of some local function of the smoothed luminance, such as the Harris operator or the Hessian (though other criteria may be used, e.g. edge line intersections). The size (scale) of the feature is then found at the scale-space extrema of some other operator (e.g. the Laplacian), possibly followed by an iterative refinement aiming at finding the affine shape of the feature.

4.2.3 Corner Detectors

Corner detection is an approach used within computer vision systems to extract certain kinds of features and infer the contents of an image. Corner detection is frequently used in motion detection, image registration, video tracking, image mosaicing, panorama stitching, 3D modeling and object recognition.

Harris Detector

Probably the most widely used corner detector is the one from Harris, proposed back in 1988. It is based on the eigenvalues of the second moment matrix, however this method is not scale invariant.

Harris proposed a formal mathematical definition for corners (or more generally, interest points) based on the rate of intensity changes in two perpendicular directions. Although this constitutes a sound definition, it requires the computation of the image derivatives which is a costly operation, especially considering the fact that interest point detection is often just the first step in a more complex algorithm.

To define the notion of corners in images, Harris looks at the average directional intensity change in a small window around a putative interest point [39]. If one considers a displacement vector (u, v) , the average intensity change is given by:

$$R = \sum (I(x + u, y + v) - I(x, y))^2 \quad (4.1)$$

The summation is defined over the neighborhood around the considered pixel. This average intensity change can then be computed in all possible directions which leads to the definition of a corner as a point. From this definition, the Harris test is performed as follows: it is first obtained the direction of the maximal average intensity change. Next, it is checked if the average intensity change in the orthogonal direction is also high. If it is the case, then a corner has been found.

Mathematically, this condition can be tested by using an approximation of the preceding formula using the Taylor expansion:

$$R \approx \sum \left(I(x, y) + \frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v - I(x, y) \right)^2 = \sum \left(\left(\frac{\partial I}{\partial x}u \right)^2 + \left(\frac{\partial I}{\partial y}v \right)^2 + 2 \frac{\partial I}{\partial x} \frac{\partial I}{\partial y}uv \right) \quad (4.2)$$

This can be rewritten in matrix form:

$$R \approx \begin{bmatrix} u & v \end{bmatrix} \begin{bmatrix} \sum \left(\frac{\delta I}{\delta x} \right)^2 & \sum \frac{\delta I}{\delta x} \frac{\delta I}{\delta y} \\ \sum \frac{\delta I}{\delta x} \frac{\delta I}{\delta y} & \sum \left(\frac{\delta I}{\delta y} \right)^2 \end{bmatrix} \quad (4.3)$$

This matrix is a covariance matrix that characterizes the rate of intensity change in all directions. This definition involves the image's first derivatives that are often computed using the Sobel operator. It can be shown that the two eigenvalues of the covariance matrix gives the maximal average intensity change and the average intensity change for the orthogonal direction [39].

It then follows that if these two eigenvalues are low, one is in a relatively homogenous region. If one eigenvalue is high and

the other is low, one must be on an edge. Finally, if both eigenvalues are high, then one is at a corner location. Therefore, the condition for a point to be accepted as a corner is to have the smallest eigenvalue of the covariance matrix higher than a given threshold.

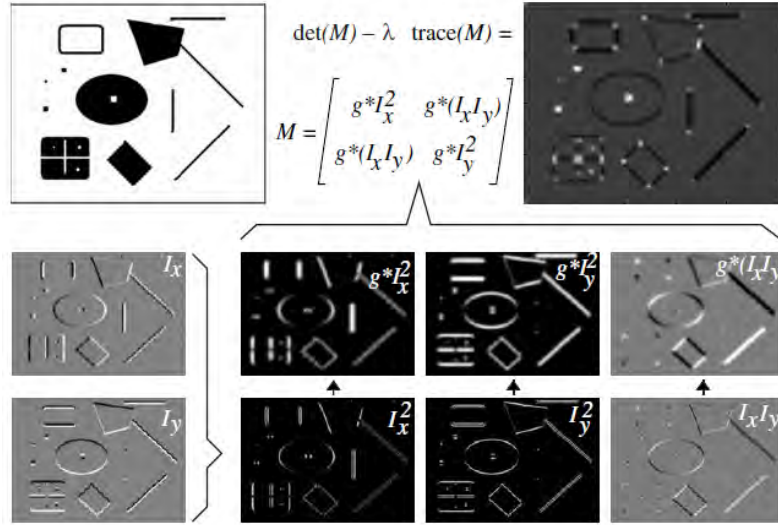


Figure 4.3: Illustration of the components of the second moment matrix and Harris cornerness measure (Source: [8])

Subsequent stages of the corner extraction process are illustrated in Figure 4.3. Given the original image I (upper left), the first step consists of computing the first-order derivatives (lower left). Next, one takes the product of these gradient images (lower right). Then, the images are smoothed with a Gaussian kernel. These images contain the different elements of the Hessian matrix, which are then in a final step combined into the cornerness measure (upper right).

Herbert is concluding in their study [40] that the Hessian-based detectors are more stable and repeatable than their Harris-based counterparts. All after studying the existing detectors and from the published comparisons. Additional improvements can be made to the original Harris corner algorithm.

Features from Accelerated Segment Test

It is the first corner detection algorithm based on the AST (Accelerated Segment Test) [9]. FAST has been specifically designed to allow quick detection of interest points in an image. The decision to accept or not to accept a key-point being based on only a few pixel comparisons.

The most promising advantage of FAST corner detector is its computational efficiency. Referring to its name, it is fast and indeed it is faster than many other well-known feature extraction methods, such as Difference of Gaussian (DoG) used by SIFT, SUSAN and Harris.

The FAST (Features from Accelerated Segment Test) feature algorithm derives from the definition of what constitutes a 'corner'. The definition is based on the image intensity around a putative feature point. The decision to accept a key-point is done by examining a circle of pixels centered at a candidate point. If an arc of contiguous points of length greater than 3/4 of

the circle perimeter is found in which all pixels significantly differ from the intensity of the center point, then a key-point is declared (see Figure 4.4).

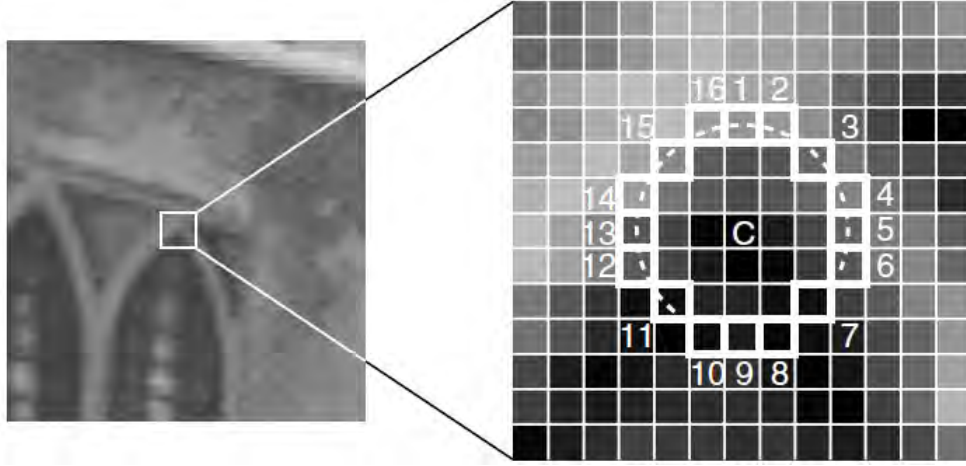


Figure 4.4: FAST Feature detection in an image patch (Source: [9])

The algorithm is explained below:

1. Select a pixel in the image (P). Assume the intensity of this pixel to be IP . This is the pixel which is to be identified as an interest point or not.
2. Set a threshold intensity value T
3. Consider a circle of 16 pixels surrounding the pixel (*Bresenham circle*)
4. N contiguous pixels out of the 16 need to be either above or below IP by the value T , if the pixel needs to be detected as an interest point.
5. To make the algorithm fast, first compare the intensity of pixels 1, 5, 9 and 13 of the circle with IP . As evident from the figure above, at least three of these four pixels should satisfy the threshold criterion so that the interest point will exist.
6. If at least three of the four pixel values $-I_1, I_5, I_9$ and I_{13} are not above or below $IP + T$, then P is not an interest point (corner). In this case reject the pixel P as a possible interest point. Else if at least three of the pixels are above or below $IP + T$, then check for all 16 pixels and check if contiguous pixels fall in the criterion.
7. Repeat the procedure for all the pixels in the image.

FAST is sufficiently fast that it allows on-line operation of the tracking system [9]. The algorithm uses an additional trick to further speed-up the process. Indeed, if we first test four points separated by 90° on the circle (for example, top, bottom, right, and left points) it can be easily shown that to satisfy the condition expressed above, at least three of these points must all be brighter or darker than the central pixel. If it is not the case, the point can immediately be rejected without inspecting additional points on the circumference. This is a very effective test since, in practice, most of the image points will be rejected by this simple 4-comparison test. In principle, the radius of the circle of examined pixels should be a parameter of the method. However, it has been found that, in practice, a radius of 3 gives both good results and high efficiency [39].

As for Harris features, it is often better to perform non-maxima suppression on the corners found. Therefore, a corner strength measure needs to be defined: the strength of a corner is given by the sum of absolute difference between the central pixel and the pixels on the identified contiguous arc.

This algorithm results in very fast interest point detection and should then be used when speed is a concern. For example, this is often the case in visual tracking applications where several points must be tracked in a video sequence with high frame rates.

4.2.4 Local Feature Detectors

The literature on local feature detection is vast and since it is not the purpose of this work only short presentation is necessary.

In the context of realtime face detection, Viola and Jones have proposed to use integral images [41], which allow for very fast computation of Haar wavelets or any box-type convolution filter. First, it will be described the basic idea of integral images. Then it will be show how this technique is useful to obtain a fast approximation of the Hessian matrix, as used in SURF (Speeded-Up Robust Features).

Detecting the Scale-Invariant

When trying to match features across different images, we are often faced with the problem of scale changes. That is, the different images to be analyzed can be taken at a different distance from the objects of interest, and consequently, these objects will be pictured at different sizes. If we try to match the same feature from two images using a fixed size neighborhood then, because of the scale change, their intensity patterns will not match. To solve this problem, the concept of scale-invariant features has been introduced in computer vision. The main idea here is to have a scale factor associated with each of the detected feature points. In recent years, several scale-invariant features have been proposed like Speeded Up Robust Features (SURF) and Scale-Invariant Feature Transform (SIFT). They are not only scale-invariant features, but they also offer the advantage of being computed very efficiently [39].

Integral Images

The entry of an integral image $I_{\Sigma}(x)$ at a location $x=(x, y)$ represents the sum of all pixels in the input image I of a rectangular region formed by the origin and x [8].

$$I_{\Sigma}(x) = \sum_{i=0}^{i \leq x} \sum_{j=0}^{j \leq y} I(i, j) \quad (4.4)$$

Using integral images, it takes only four operations to calculate the area of a rectangular region of any size:

Moreover, the calculation time is not depending of the size of the rectangular area.

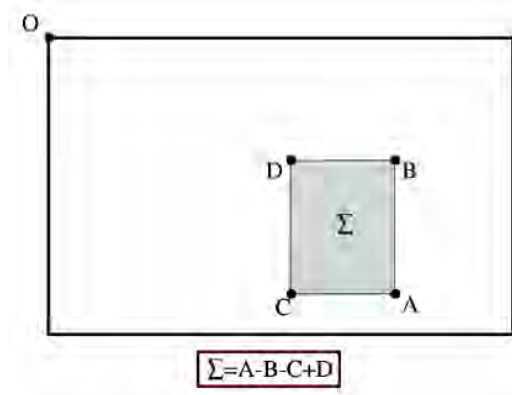


Figure 4.5: Integral image (Source: [8])

Speeded Up Robust Features

SURF or Speeded Up Robust Features have been proposed by Bay [40]. It is a scale-invariant feature detector based on the Hessian-matrix, as is, e.g., the Hessian-Laplace detector. However, rather than using a different measure for selecting the location and the scale, the determinant of the Hessian is used for both. The Hessian matrix is roughly approximated, using a set of box-type filters, and no smoothing is applied when going from one scale to the next [8].

Image derivatives of an image can be estimated using Gaussian filters. Those filters make use of a σ parameter defining the aperture (size) of the kernel. This σ corresponds to the variance of the Gaussian function used to construct the filter, and it then implicitly defines a scale at which the derivative is evaluated [39].

If one computes, the Laplacian of a given image point using Gaussian filters at different scales, then different values are obtained. Looking at the evolution of the filter response for different scale factors, we obtain a curve which eventually reaches a maximum value at some σ value. Thus, scale-invariant features should be detected as local maxima in both the spatial space (in the image) and the scale space (as obtained from the derivative filters applied at different scales).

SURF implements this idea by proceeding as follows. First, to detect the features, the Hessian matrix is computed at each pixel. This matrix measures the local curvature of a function and is defined as:

$$H(x, y) = \begin{bmatrix} \frac{\delta^2 I}{\delta x^2} & \frac{\delta^2 I}{\delta x \delta y} \\ \frac{\delta^2 I}{\delta x \delta y} & \frac{\delta^2 I}{\delta y^2} \end{bmatrix} \quad (4.5)$$

The determinant of this matrix gives the strength of this curvature. The idea behind is to define corners as image points with high local curvature (that is, high variation in more than one direction). Since it is composed of second-order derivatives, this matrix can be computed using Laplacian Gaussian kernels of different scale σ [40].

A scale-invariant feature is therefore declared when the determinant of this Hessian reaches a local maximum in both spatial and scale space. The calculation of all of these derivatives at different scales is computationally costly. The objective of the SURF algorithm is to make this process as efficient as possible. This is achieved by using approximated Gaussian kernels

involving only few integer additions (right part of Figure 4.6).

Gaussians are optimal for scale-space analysis, but in practice they have to be discretized which introduces artifacts, in particular in small Gaussian Kernels (1, -1, -2 in Figure 4.6). SURF pushes the approximation even further, using the box filters as shown below:

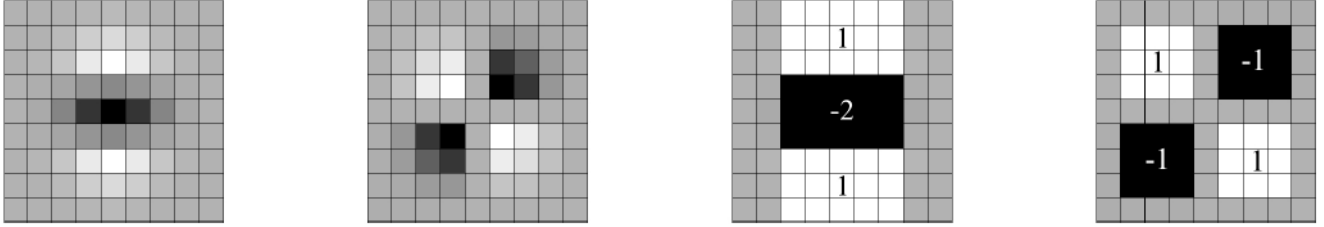


Figure 4.6: Gaussian partial derivative comparison with SURF's box filter (y, x-y dir.) (Source: [8])

Once the local maxima is identified, the precise position of each detected interest point is obtained through interpolation in both scale and image space. The result is then a set of feature points localized at sub-pixel accuracy and to which is associated a scale value.

The approximated second-order partial Gaussian derivatives can be evaluated very fast using integral images, independently of their size. Surprisingly, in spite of the rough approximations, the performance of the feature detector is comparable to the results obtained with the discretized Gaussians. Box filters can produce a sufficient approximation of the Gaussian derivatives. SURF has been reported to be more than five times faster than DoG.

In Figure 4.7 one can observe the detected features in a gray-scale camera images, using the SURF algorithm to find blob features.

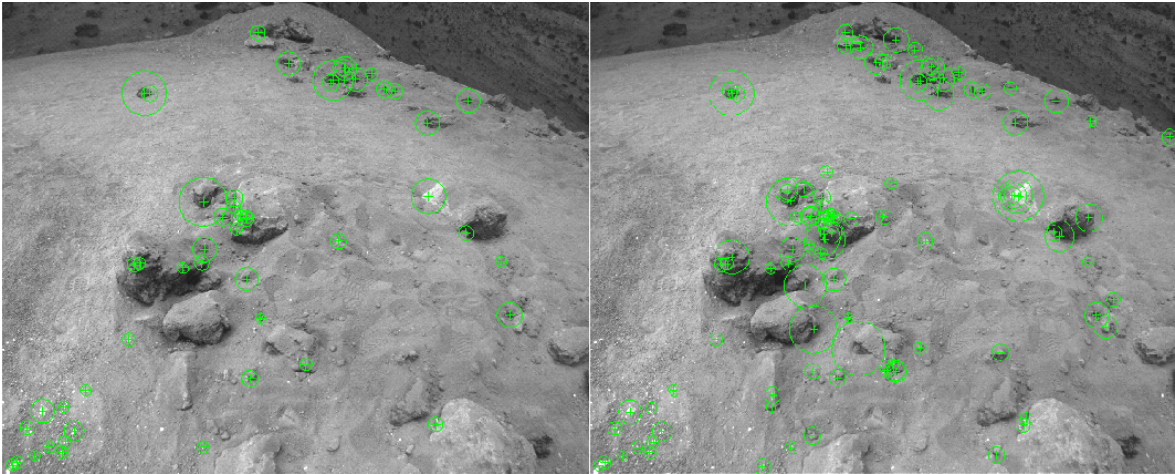


Figure 4.7: Feature detection with SURF: the 50 strongest points (left); the 150 strongest points (right)

Scale-Invariant Feature Transform

The algorithm was first published by Lowe in 1999 [10]. Applications include object recognition, robotic mapping and navigation, image stitching, 3D modeling, gesture recognition, video tracking, and match moving.

For any object in an image, interesting points on the object can be extracted to provide a 'feature description' of the object. This description can then be used to identify the object when attempting to locate the object in a test image containing many other objects. To perform reliable recognition, it is important that the features extracted from the training image be detectable even under changes in image scale, noise and illumination. Such points usually lie on high-contrast regions of the image, such as object edges.

Lowe's patented method can robustly identify objects even among clutter and under partial occlusion, because his SIFT feature descriptor is invariant to uniform scaling, orientation, and partially invariant to affine distortion and illumination changes [10].

The scale-invariant features are efficiently identified by using a staged filtering approach. The first stage identifies key locations in scale space by looking for locations that are maxima or minima of a difference-of-Gaussian function [10]. Each point is used to generate a feature vector that describes the local image region sampled relative to its scale-space coordinate frame. The features achieve partial invariance to local variations, such as affine or 3D projections, by blurring image gradient locations.

As the 2D Gaussian function is separable, its convolution with the input image can be efficiently computed by applying two passes of the 1D Gaussian function in the horizontal and vertical directions:

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2} \quad (4.6)$$

To characterize the image at each key location, the smoothed image A at each level of the pyramid is processed to extract image gradients and orientations. At each pixel, A_{ij} , the image gradient magnitude, M_{ij} , and orientation, R_{ij} , are computed using pixel differences [10]:

$$M_{ij} = \sqrt{(A_{ij} - A_{i+1,j})^2 + (A_{ij} - A_{i,j+1})^2} \quad (4.7)$$

$$R_{ij} = \text{atan2}(A_{ij} - A_{i+1,j}, A_{i,j+1} - A_{ij}) \quad (4.8)$$

The magnitude and direction calculations for the gradient are done for every pixel in a neighboring region around the key-point in the Gaussian-blurred image L . An orientation histogram with 36 bins is formed, with each bin covering 10 degrees. Each sample in the neighboring window added to a histogram bin is weighted by its gradient magnitude and by a Gaussian-weighted circular window

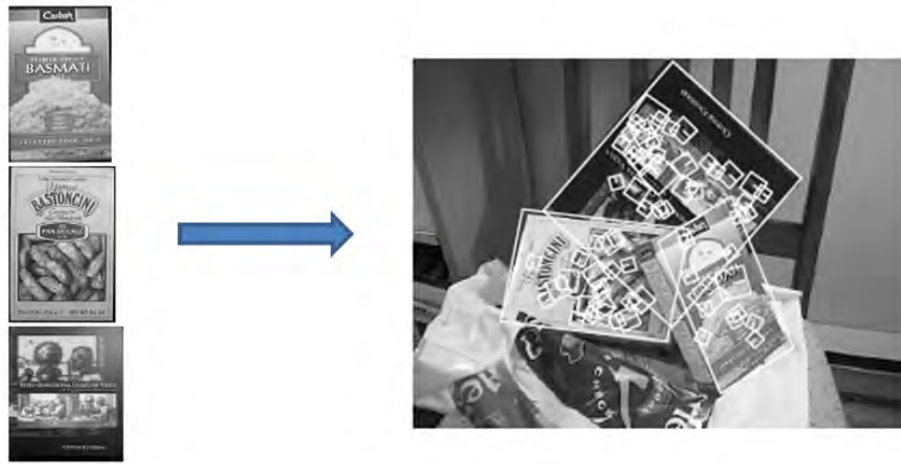


Figure 4.8: Storing SIFT keys in sample images and identify matching keys from new images (Source: [10])

Now we want to compute a descriptor vector for each key-point such that the descriptor is highly distinctive and partially invariant to the remaining variations such as illumination, 3D viewpoint. Outliers can be removed by checking for agreement between each image feature and the model, given the parameter solution.

First a set of orientation histograms are created on 4×4 pixel neighborhoods with 8 bins each. These histograms are computed from magnitude and orientation values of samples in a 16×16 region around the key-point such that each histogram contains samples from a 4×4 subregion of the original neighborhood region. The magnitudes are further weighted by a Gaussian function with equal to one half the width of the descriptor window. The descriptor then becomes a vector of all the values of these histograms. Since there are $4 \times 4 = 16$ histograms each with 8 bins the vector has 128 elements. This vector is then normalized to unit length in order to enhance invariance to affine changes in illumination. To reduce the effects of non-linear illumination a threshold of 0.2 is applied and the vector is again normalized [42].

SIFT features can essentially be applied to any task that requires identification of matching locations between images. Work has been done on applications such as recognition of particular object categories in 2D images, 3D reconstruction, motion tracking and segmentation, robot localization, image panorama stitching and epipolar calibration.

4.2.5 Region detectors

Often these detectors are used to find correspondences between image elements from two images with different viewpoints.

Maximally Stable Extremal Regions

Maximally Stable Extremal Regions (MSER) have been proposed by Matas et al. in [43]. A Maximally Stable Extremal Region is a connected component of an appropriately thresholded image. 'Extremal' refers to the property that all pixels inside the MSER have either higher (bright extremal regions) or lower (dark extremal regions) intensity than all the pixels on its outer boundary. 'Maximally stable' in MSER describes the property optimized in the threshold selection process.

Detection of MSER is related to thresholding, since every extremal region is a connected component of a thresholded image [8]. However, no global or 'optimal' threshold is sought, all thresholds are tested and the stability of the connected components evaluated. The output of the MSER detector is not a binarized image. For some parts of the image, multiple stable thresholds exist and a system of nested subsets is output in this case.

The enumeration of the extremal regions is very efficient, almost linear in the number of image pixels. The enumeration proceeds as follows [8]:

- First, pixels are sorted by intensity.
- After sorting, pixels are marked in the image (either in decreasing or increasing order)
- The list of growing and merging connected components and their areas is maintained using the union-find algorithm
- MSER are the parts of the image where local binarization is stable over a large range of thresholds.

Advantages of MSER include [44]:

- Invariance to affine transformation of image intensities.
- Covariance to adjacency preserving transformation.
- Stability: only regions whose support is nearly the same over a range of thresholds is selected.
- Multi-scale detection without any smoothing involved.

4.2.6 Overview of Feature Detectors

Local features are a popular tool for image description nowadays. They are the standard representation for wide baseline matching and object recognition, both for specific objects as well in most of the application, as well in robotics.

The choice of the feature detector very much depends on a problem. For example, if one is doing monocular SLAM ², FAST detector is weapon of choice, because it's fast and detects a lot of features. For image stitching, pattern recognition and other feature-descriptor related tasks, scale- and rotation-invariant detectors are preferable.

On the next page (see Table 4.1), an overview of the different feature detector is given. For our analysis Harris, MSER, SIFT and SURF were chosen from the list. Several reasons are behind Harris: it is one of the most common algorithms, also it is considered the most possible candidate for ExoMars navigation algorithms. SURF detectors were designed for efficiency which results in a very short computation time. MSER is also an efficient due to the use of a watershed segmentation algorithm. The results of the analysis are presented in the next Chapter.

One of the major drawbacks of all feature detectors is a weak robustness to various transformations in particular in estimation of local scale and shape. By far the most frequently used evaluation measure is the repeatability of features. The images will be analyzed according to these principles.

²Simultaneous Localization And Mapping

Feature Detector	Corner	Blob	Region	Rotation invariant	Scale invariant	Affine invariant	Repeatability	Localization accuracy	Robustness	Efficiency
Harris	✓			✓			+++	+++	+++	++
Hessian		✓		✓			++	++	++	+
SUSAN	✓			✓			++	++	++	+++
Harris-Laplace	✓	(✓)		✓	✓		+++	+++	++	+
Hessian-Laplace	(✓)	✓		✓	✓		+++	+++	+++	+
DoG	(✓)	✓		✓	✓		++	++	++	++
SURF	(✓)	✓		✓	✓		++	++	++	+++
Harris-Affine	✓	(✓)		✓	✓	✓	+++	+++	++	++
Hessian-Affine	(✓)	✓		✓	✓	✓	+++	+++	+++	++
Salient Regions	(✓)	✓		✓	✓	(✓)	+	+	++	+
Edge-based	✓			✓	✓	✓	+++	+++	+	+
MSER			✓	✓	✓	✓	+++	+++	++	+++
Intensity-based			✓	✓	✓	✓	++	++	++	++
Superpixels			✓	✓	(✓)	(✓)	+	+	+	+

Table 4.1: Feature detectors: Overview (Source: [8])

Chapter 5

Measurements and Simulation

Two different methods were used to create the lifelike copy of the Martian terrain found at the Automation and Robotics Laboratory at ESTEC. With the so-called *active* technique, a highly precise DEM was created using a terrestrial laser scanning system. The drawback of this method was the lack of texture. To overcome this problem, undistorted high-resolution images were projected on the simulated environment.

The other method was the so-called multi-view stereo reconstruction technique (*passive* one) which turned out to be surprisingly precise. This is comparable to the previous one, having the advantage of containing the texture information.

This Chapter will show all the measurements done, the outcome and the different scenario set-ups in both real world and simulations.

5.1 Coordinate Systems

Obtaining precise data sets for analysis and validation (Chapter 6) is not possible without an accurate coordinate determination. Coordinates uniquely determine the position of a point on a manifold such as Euclidean space. These coordinates and coordinate systems used in both reality and simulations must have been precisely matched. In the following a short introduction is presented to define the different coordinate systems.

RiSCAN PRO uses a number of coordinate system: SOCS¹ is the coordinate system in which the scanner delivers the raw data. For every laser measurement the data contains geometric information. This is embedded into the GLCS² for the ease of data manipulation and visualization. PRCS³ is defined by the user at an already existing coordinate system at the scan site.

A scan position is characterized the position and orientation of the scanner within the project coordinate system. Position and orientation can generally be described by 6 parameters (3 for position, 3 for rotation) or by a transformation matrix. The

¹Scanner's Own Coordinate System

²Global Coordinate System

³Project Coordinate System

software uses a 4 x 4 matrix for describing the sensor's orientation and position.

The GRS⁴ is shared between more than one scan. To transform each scan from its own coordinate system into a GCS a 3D roto-translation needs to be computed on the basis of common control points (or features). This operation is called scan co-registration [45]. Given the vector X storing coordinates of a point in the GCS, and the vector x with the corresponding coordinates, the transformation between both reference systems can be expressed by introducing the rotation matrix R and the vector O_i expressing the origin of the SOCS with respect to the GRS:

$$X = Rx + O_i \quad (5.1)$$

The rotation matrix can be parameterized by angles as commonly done in photogrammetry. Concerning materialization of a GRS, this can be done by a set of control points with known coordinates (see later Section 5.2.1).

One of the difficulties of the present work is to match 3D positions and rotations from real life with the simulation. Considerable efforts were needed to overcome this problem as it will be seen in Section 5.5.2.

5.2 Active Technique

In order to obtain a high precision point cloud of the testbed in the ARL, high accuracy laser scans were conducted. With the help of the lab equipment, namely the laser scanning system (RIEGL) and the tracking system (Vicon) several measurements were conducted and post-processed. These mapped a digital copy of the martian terrain which was imported in the simulation software.

In order to achieve a highly detailed representation of our Mars-like testbed, several scans from different positions have been performed. Furthermore, several reflectors ('markers' or 'tie points') were placed on the terrain in order to properly correlate all point clouds acquired from different scan positions.

The comparison and validation of the created simulation environments will be discussed in Chapter 6. First the 3D reconstruction from the laser scan (active technique) will be presented, later the multi-view stereo reconstruction.

5.2.1 Measurements

In almost all applications data acquisition is based on taking scans from different locations in order to get a complete data set of the object's surface without gaps or 'scan shadows'. The different scan locations are referred as scan positions.

Six measurements were conducted and post processed in RiSCAN PRO, as seen in Figure 5.1. The accuracy of the laser scan needed to be high enough in order to have a smooth terrain representation. Ergo the resolution was selected to be 0.030° in both vertical and horizontal directions. The final data set contains 14.07 million points. These measurements were conducted

⁴Ground Reference System

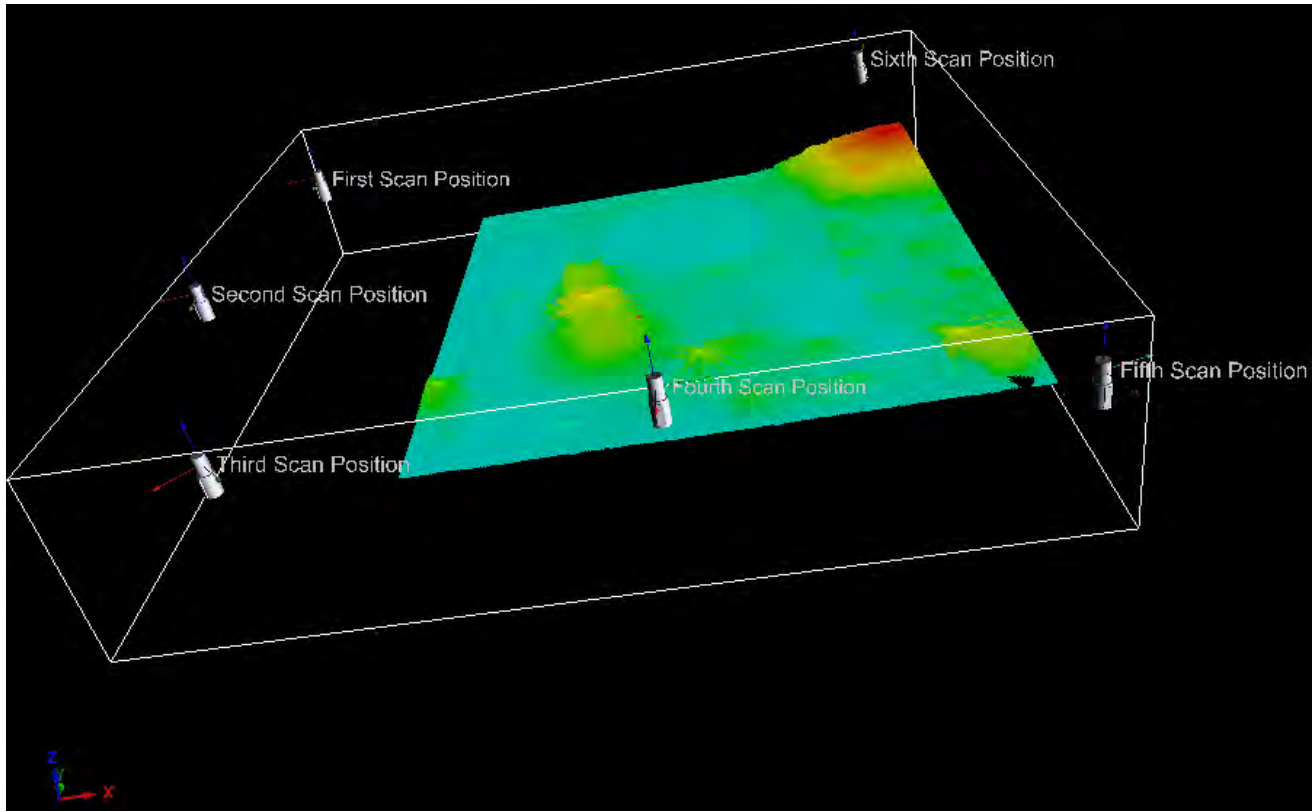


Figure 5.1: Polydata made by 6 laser measurements showing the scanner positions

with the conditions listed in Table 5.1, showing the separate data sets for each scan position.

Measurement nr.	Nr. of data points	Vertical angle FOV (θ)	Horizontal angle FOV (φ)	Resolution
Measurement 1	5 658 585	$90^\circ - 129.96^\circ$	$103^\circ - 230.32^\circ$	0.030°
Measurement 2	4 140 861	$90^\circ - 129.96^\circ$	$125^\circ - 237.77^\circ$	0.030°
Measurement 3	3 369 824	$90^\circ - 129.96^\circ$	$145^\circ - 220.84^\circ$	0.030°
Measurement 4	5 644 600	$90^\circ - 129.96^\circ$	$110^\circ - 240.20^\circ$	0.030°
Measurement 5	3 829 709	$90^\circ - 129.96^\circ$	$155^\circ - 241.19^\circ$	0.030°
Measurement 6	6 675 664	$90^\circ - 129.96^\circ$	$180^\circ - 330.20^\circ$	0.030°

Table 5.1: Measurement angles

The different data sets need to be combined in order to get a final DEM of the terrain. For this a motion tracking system was used to measure the retro-reflective ball's position in three dimension. These markers were placed on the terrain, and acted as precise way-points in the post processing step. In order to obtain the location of these markers a reference system and a proper camera calibration was necessary.

By default the origin is located at the position where camera number one is. But the VICON system has been configured to use one of the corners of the testbed as the origin, that is, the (0,0,0) coordinate. This was done using the ErgoCal L-Frame 25 mm (see Figure 5.2(a)) which contains four markers. The origin is detected by the Vicon system knowing the type of the L-Frame.

In order to have the correct positions of the marker, the Vicon system must be calibrated. The calibration process involves

identifying both internal (focal length and image distortion) and external (camera position and orientation) camera parameters. Although, the most important factor is related to the spatial distribution of the infra cameras. Obviously if one has not a calibrated system, the measurement will not be precise. A 3-marker calibration wand (see Figure 5.2(b)) is used to calibrate the MX cameras. Based on the distance between the markers, the software can triangulate the camera position. To have a sufficient calibration accuracy the system needs to have between 1000-1300 wand marker counts.

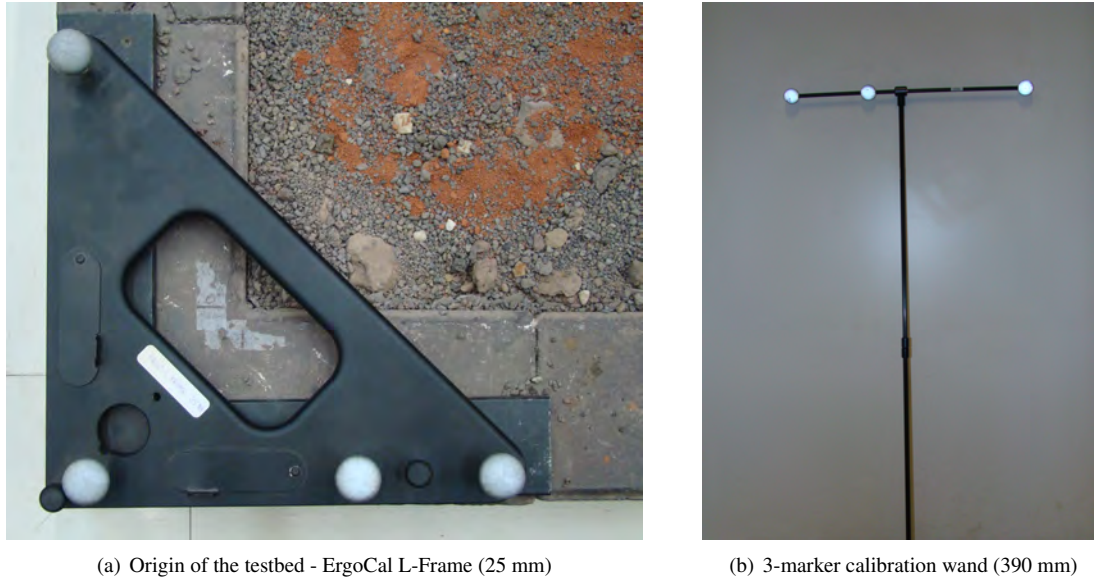


Figure 5.2: Reference system origin and calibration wand

In Table 5.2 the coordinates of the 12 markers placed on the testbed are summarized.

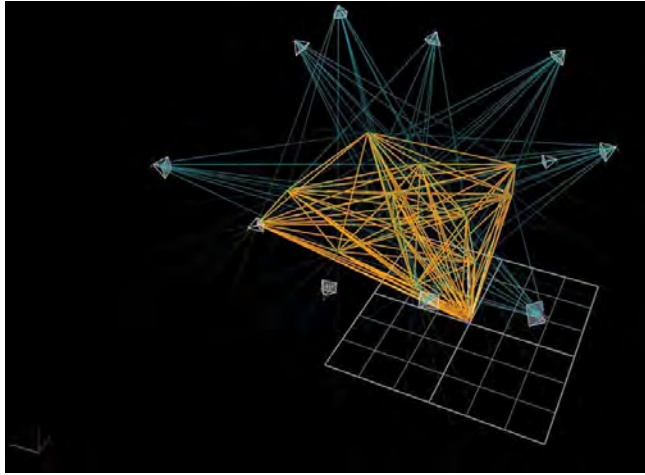
Marker nr.	X coordinate	Y coordinate	Z coordinate
Marker1	0.490766	1.931042	0.277539
Marker2	2.505144	0.736717	0.035237
Marker3	4.321479	1.712909	0.308652
Marker4	2.673920	3.609800	0.485498
Marker5	1.198011	4.878947	0.061847
Marker6	4.670933	4.559106	0.141789
Marker7	6.313473	0.743458	0.151778
Marker8	6.805769	4.432093	0.175847
Marker9	7.797544	1.090224	0.397679
Marker10	7.371297	6.884144	0.608227
Marker11	3.432952	8.108245	0.083629
Marker12	1.414247	8.040763	0.029089

Table 5.2: Marker positions on the testbed (in meters)

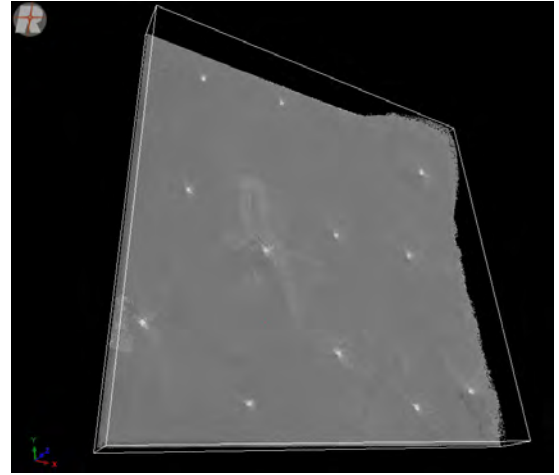
5.2.2 Digital Elevation Map

A digital elevation map is a digital model or 3-D representation of a terrain's surface. Henceforth it will be referred to as DEM. Such a map is created from terrain elevation data.

The six laser measurements of the ARL testbed were meshed together with the help of the the RiSCAN PRO software version 1.4.3sp3. The markers were used as a reference: the program can detect them automatically knowing the position from the reference system (the corner of the testbed). The markers, which were already presented before, can be seen in Figure 5.3(a) and Figure 5.3(b). Straight blue lines are indicating those markers which are in the FOV of a specific camera. The orange lines are connecting the origin with the markers and creating the terrain segment. Next, in Figure 5.3(b) one can observe the markers after the post-processing phase of the six terrestrial scans.



(a) Real-time marker detection by the Vicon system



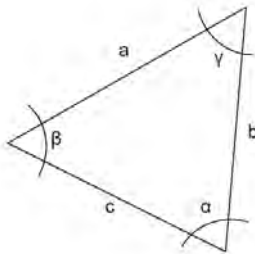
(b) Markers seen on the terrain with the help of RiEGL system

Figure 5.3: 12 markers seen by the two systems

Figure 5.4 represents the final point cloud of the robotic testbed at ARL. This model contains 14.07 million unique data points. The height is represented by colors: the highest point at 0.764 meters represented by red (the top of the hill in top left corner), and the lowest point at -0.025 meters represented by dark blue (the lowest point of the crater). This will be later used to create the volume of the testbed by triangulation, explained briefly in the next section.

Triangulation and DEM

Classical triangulation were already used in the ancient Greek time, it has to be credited to the Greek philosopher Thales (6th century BC). But the principle remains the same even today:



$$\frac{a}{\sin(\alpha)} = \frac{b}{\sin(\beta)} = \frac{c}{\sin(\gamma)} \quad (5.2)$$

$$a^2 = b^2 + c^2 - 2 \cdot b \cdot c \cdot \cos(\alpha) \quad (5.3)$$

$$c = a \cdot \cos(\beta) + b \cdot \cos(\alpha) \quad (5.4)$$

The raster (computed) DEM has been created by triangulation. Triangulation is the process to create a surface out of a point cloud where the surface is made up by triangles connecting the data points. Triangulated data (or 'mesh') gives a better

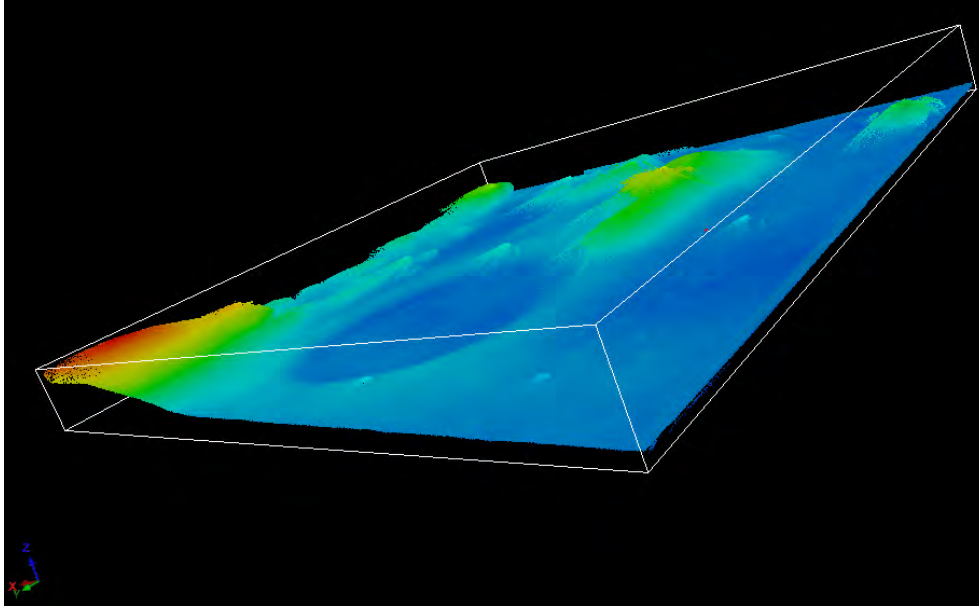


Figure 5.4: Overlapped poly-data from six measurements (height representation from a side view)

representation of the scanned object.

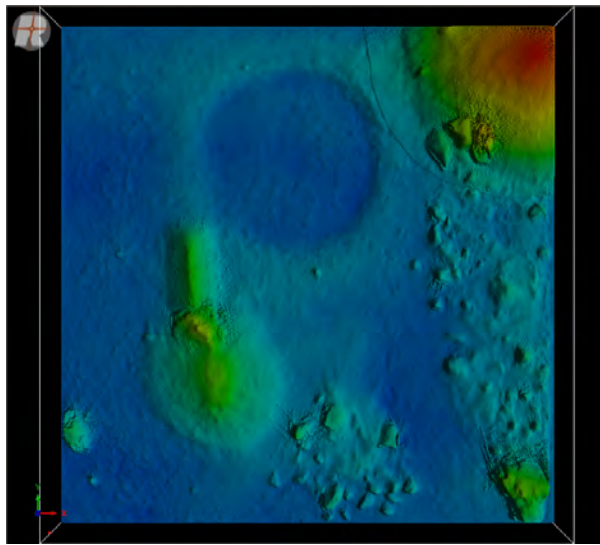
Having the point cloud representation, a coplanar plane was added in parallel with the x-y plane, using the position of -0.1 m in the z direction. This way a volume was created as a triangulated mesh. Moreover, there must be a trade-off when selecting the raster size, since this parameter defines the smoothness of the DEM. The surface triangulation has been carried out at the original data resolution, and after the number of triangles has been reduced by 'smooth and decimate' function of RiscanPro. This modifies the surface structure of the polydata object by optimizing the point data (smoothing), and by reducing the amount of triangles (decimating). Those areas where the density of the points is high enough, are very realistic. However, those areas with a low density are interpolated in an unrealistic way, showing long and sharp edges when joining the small triangles together. For this reason several DEMs were created using several raster size (please consult Table 5.3).

Raster size [m]	Surface [m^2]	Volume [m^3]	Nr. of points	Nr. of polys
0.030	80.688	13.248	84,400	168,796
0.020	81.586	13.225	188,065	376,126
0.015	82.570	13.213	333,249	666,494
0.010	85.117	13.202	746,999	1,493,994
0.005	104.466	13.186	2,971,834	5,943,664

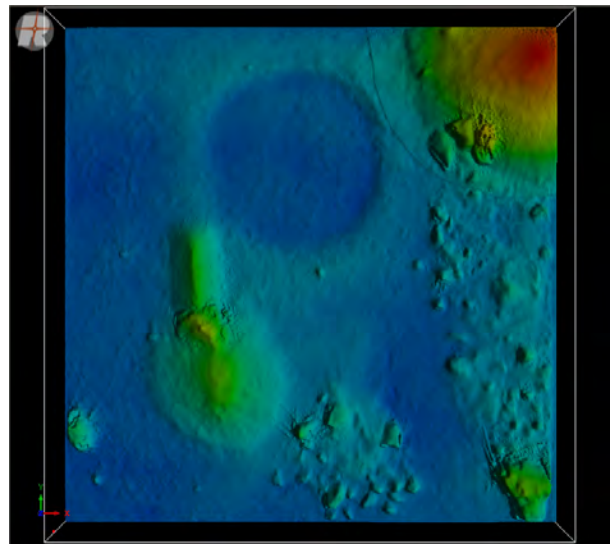
Table 5.3: Triangulated DEMs with different precision

According to the real-life measurement the actual size of the terrain is 8.69 m x 8.605 m, giving an area of 74.77 m^2 . Adding the corner stones 12 cm distance per side, gives the size of 8.93 m x 8.84 m, resulting in an area of 78.94 m^2 .

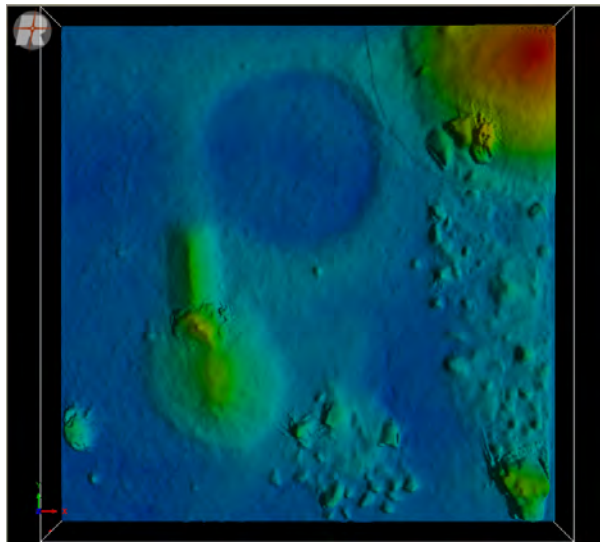
Clearly there is a very slight difference between the reality and the model, in the terms of the test bed area. The more precise the triangulation, the higher the area. Also, showing long and sharp edges is not permitted. Therefore, all these factors have been considered, and a trade-off solution implemented.



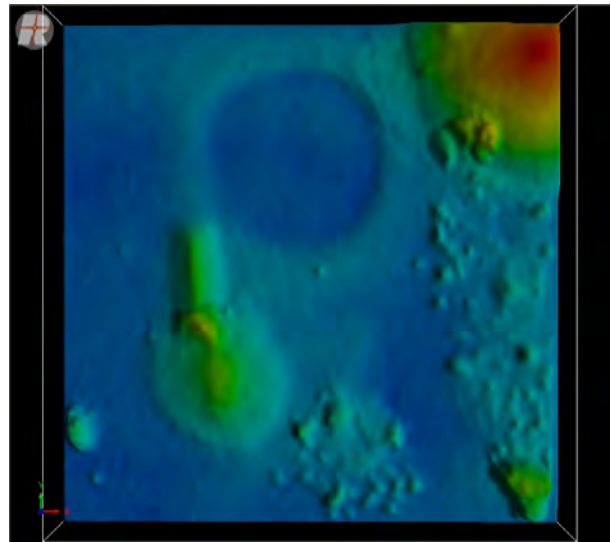
(a) Triangulation with a precision of 1 cm



(b) Triangulation with a precision of 1.5 cm



(c) Triangulation with a precision of 2 cm



(d) Triangulation with a precision of 5 cm

Figure 5.5: Top view of the different triangulation results (colored height representation)

The DEM model with a raster size of 1.5 cm was chosen after a throughout visual inspection. As in Figure 5.5 is visible, it differs the smoothness of the terrain according to the precision parameter.

Later this very precise DEM will be used as input to a simulation software called 3DROV. Although the model does not contain texture, high-resolution images will be projected on the terrain model. This will later help to validate the simulation and it is also meaningful for the sensitivity analysis.

5.3 Passive Technique: Multi-View Stereo Reconstruction of the Testbed

Distorted RGB-camera images are used as input for the multi-view reconstruction algorithm developed at CTU. It was tried to make several horizontally oriented pictures capturing the whole testbed (including the walls). The camera was aimed all the time to the middle of the playground. The pictures were taken approximately at every half a meter at a constant height. The goal was to make it in a such way, that the positions will be uniformly distributed around the area of interest.

During this experiment 113 pictures were captured with a Sony DSC-H7. The first set of images belongs to the first height level (around 1.4 m), the second set belongs to the second height level (around 2.3 m) and the third set is taken at an approximately 3.5 m height (see Figure 5.6). Because of the lack of space around the walls, there were no pictures taken at the third height level next to the walls. The camera positions from the terrain was approximately 1 m on front and right side, and 0 meters from left and back side (see Figure 5.6) since there was a wall.



Figure 5.6: Camera positions and orientation of the 113 images taken around the testbed

The 3D dense reconstruction procedure described in [6] was used to create the model with the help of the online CMP Sfm web-service [46]. In Figure 5.7 one can see the different steps which are necessary to create a reconstruction. First the algorithm is creating the 3D reconstruction only from points, next shaded surfaces are created without texture, finally the texture from the pictures is computed.

5.4 Cameras

5.4.1 Calibration

To correct lens distortion effect the parameters of the function $\Delta(r)$ must be determined (see Section 3.3). The function contains all the information that is needed to undistort images created with a particular lens. To find these parameters and later undistort

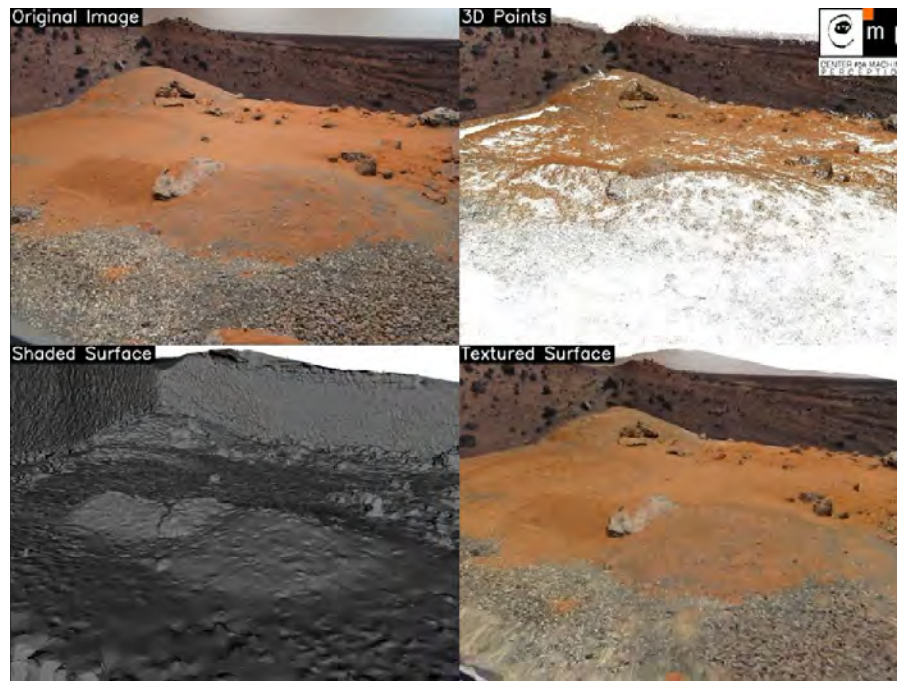
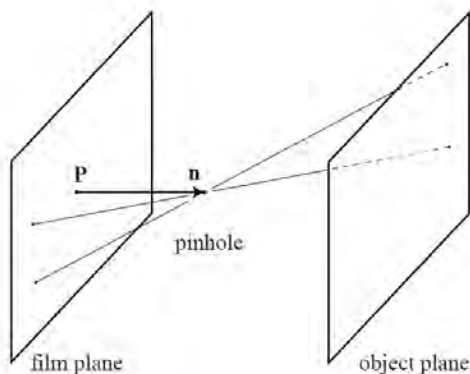


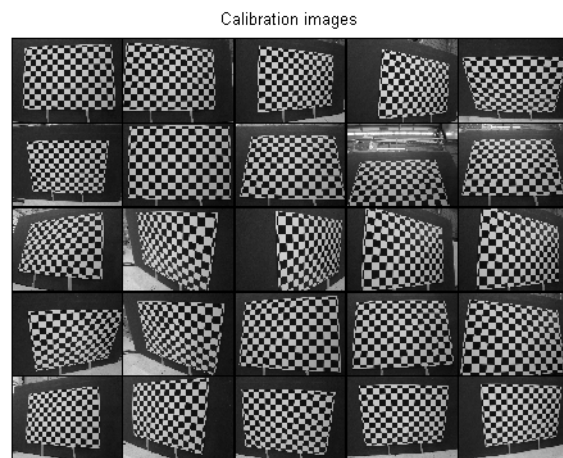
Figure 5.7: CMP SfM reconstruction of the ARL testbed (perspective view)

the projected images different values needed to be measured like focal length, skew, and distortion.

25 images of a planar checkerboard are used for calibration purposes (see Figure 5.8(b)). The checkerboard consists of alternating white and black squares (15 x 10, a total of 150). The reason is that the corner detection algorithms easily detect the edges of such pattern. The procedure is considered as a standard in the field of image processing.



(a) Diagram of a pinhole camera model



(b) Checkerboard calibration images

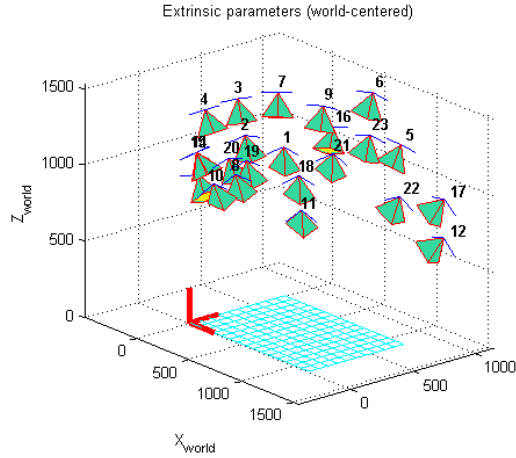
The image corners are then automatically extracted, by knowing the parameters like the width and height in mm of the white and black squares. The process is repeated on all the calibration images. Finally the calibration results are optimized and the parameters of Sony DSC-H7 can be found in Table 5.4.

Direction	Focal length	Error	Principal point	Error	Pixel error
Horizontal	970.31	2.25	509.42	2.45	0.85
Vertical	973.12	2.19	393.54	1.87	0.73

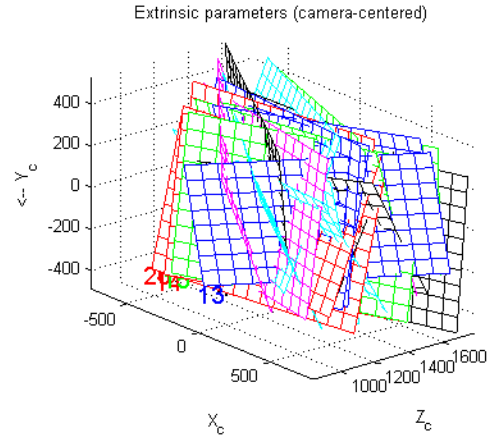
Table 5.4: Calculated calibration parameters (in pixels) on 1024 x 768 pixel images

Note that the numerical errors are approximately three times the standard deviations.

In Figure 5.8(c), every camera position and orientation is represented by a green pyramid. In Figure 5.8(b), the frame is the camera reference frame. The red pyramid corresponds to the effective field of view of the camera defined by the image plane.

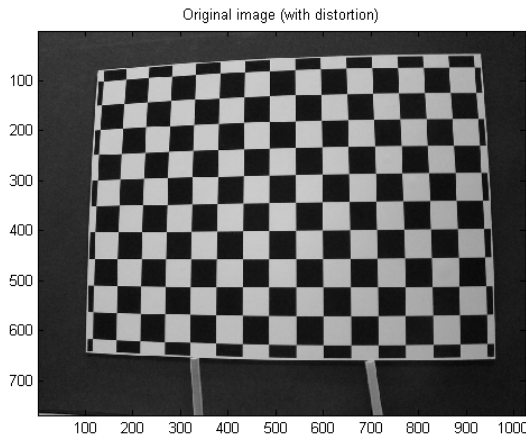


(c) Camera position and orientation of the calibration images

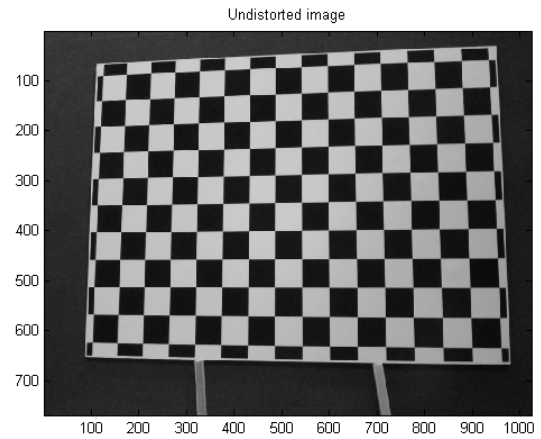


(d) Camera centered view of calibration images

The distortion function parameters $\Delta(r)$ are found out to be $c_1 = -0.18429$, $c_2 = 0.17718$, $c_3 = 0.00079$, $c_4 = -0.00056$ and $c_5 = 0$. For this the algorithm presented in [37] requires 11 integrations order to reach the minimum. The reason for that fast convergence is the quality of the initial guess for the parameters computed by the initialization procedure.



(e) Original image of the checkerboard



(f) Undistorted image after lens correction

Figure 5.8: Image comparison

Final Model

Since camera calibration in MATLAB was found to be computational intensive and not enough memory could be allocated, a new approach was necessary to undistort high resolution images. First, the camera parameters were scaled up. Later a numerical calculation was carried out by means of camera calibration using OpenCV. It was found out that the first result had a close approximation. The results of the new calibration parameters are shown in the following Table 5.5:

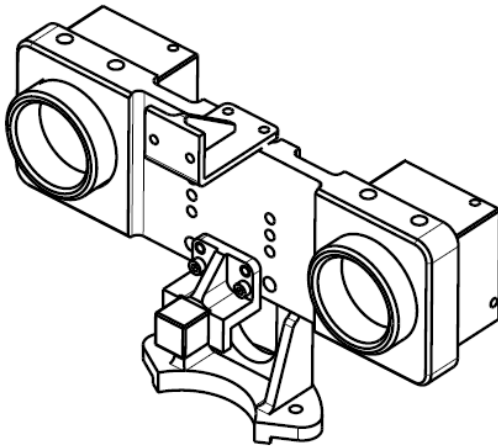
Direction	Focal length	FOV	Principal point	Error
Horizontal	3094.38	55.69	1611.85	1.50
Vertical	3106.03	43.09	1242.49	0.82

Table 5.5: Calibration parameters (in pixels) on 3264 x 2448 pixel images

The radial distortion coefficients ($k_1 = -0.1796$; $k_2 = 0.1450$) and the tangential distortion coefficients ($p_1 = -0.0002$; $p_2 = -0.0010$) are slightly different from the previous case. The explanation is simple: a lower number of calibration images were used (15 instead of 25). Theoretically these must be the same even if zooming is applied, or in the case of a change of lens. Absolutely the same distortion coefficients can be used for images of a different resolution from the same camera (while f_x , f_y , c_x and c_y need to be scaled appropriately).

5.4.2 Stereo Cameras

The NavCams (see camera specifications on Table 5.6) of the ExoMader are black and white stereo cameras with a 10 cm stereo base line, allowing to reconstruct the vicinity of the rover (from the wheels to approximately 5 m ahead) and to plan a path toward the assigned goal. NavCams will be used to capture reality by means of stereo images.



(a) Stereo camera CAD drawing



(b) Stereo cameras of ExoMader with retro-reflective balls

Figure 5.9: Stereo camera: from CAD drawing to reality

During the data acquisition campaign, a few positions on the testbed were recorded. Reflecting balls were attached to the stereo camera frame in order to obtain not just the distance from the origin, but also the rotation angles. This was possible

Characteristics	Value	Units in SI
Sensor model (type)	UI-1440-M (Mono)	N.A.
Resolution	1280 x 1024	pixels
Resolution	1.3	Mp
Sensor size	1/1,8	inch
Optical size	7.68 x 6.14 ⁵	mm
Optical diagonal	8.5	mm
max. fps in Freerun Mode	17	fps

Table 5.6: uEye camera specifications

by recording two points in every plane and measuring the angle between the line passing by this line and the norm in every direction. Consult Figure 5.9(b) for the retro-reflective sphere locations on the stereo camera platform.

For the stereo calibration, 37 images were taken with the cameras of ExoMader. A 5 coefficients distortion model was used [47]. The camera parameters were measured in [47]. Only the focal length is used to determine the actual FOV (consult Table 5.7).

Parameters measured [pixels]	Left-Cam	Right-Cam
Focal length (horizontal; vertical)	832.159; 832.693	831.471; 832.024
Focal length error (horizontal; vertical)	± 0.191 ; ± 0.158	± 0.198 ; ± 0.163
Principal point (horizontal; vertical)	672.502; 510.394	642.985; 506.628

Table 5.7: Measured parameters of the stereo cameras in pixels [11]

Field of view is defined as the area which can be focused by the camera lens. Depending on the focal length and sensor size, wider lens gives wider FOV (see Figure 5.10). Camera's FOV can be measured horizontally, vertically, or diagonally - in this case it will be measured in both x and y direction.

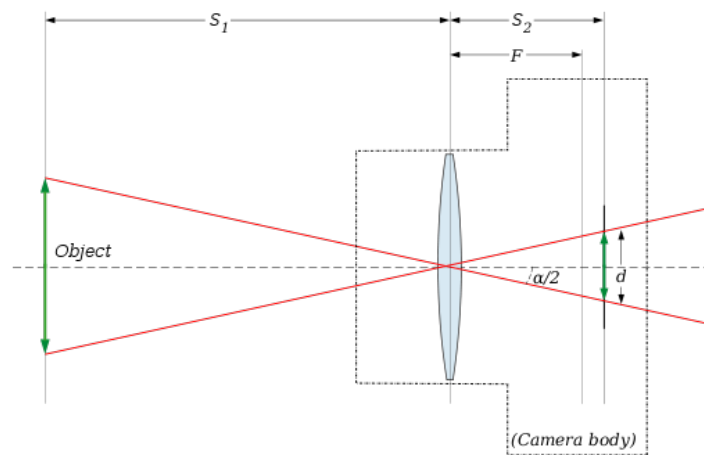


Figure 5.10: Field of view depending on focal length and sensor size

To determine the field of view (angle of view) Equation 5.5 is used:

$$\alpha = 2 \cdot \arctan \left(\frac{d_x \cdot N_x}{2 \cdot f} \right) \quad (5.5)$$

where d_x is the dimension of one pixel in the x direction, f being the focal length, and N_x is the number of pixels in the x direction. It is defined similarly for y direction.

It is known that the focal length measured in *mm* is directly proportional to the focal length measured in *pixels* - as the dimension of one pixel on the sensor is proportional to the number of pixels on the same direction.

Knowing this, the precise FOV of each camera can be determined (listed in Table 5.8).

FOV [°]	Left-Cam	Right-Cam
Horizontal	75.1266	75.1724
Vertical	63.1724	63.2135

Table 5.8: Calculated FOV parameters of the stereo cameras

5.5 Simulations Using 3DROV

The visualization part of the software was by far the most valuable during the present work. Its objectives include the support of those robotic operations that involves the motion of the robot, imaging and interaction between the robot and its environment. It visualizes the terrain with its shadows depending on the position of the sun in the real-time. With high performance computer systems it achieves a 25 fps visualization, which delivers a continuous framework for the eye. Moreover, thanks to the system running under the hood, the contact points between the rover and the terrain are highly accurate. In this way, the simulation is very realistic.

Currently in the Automation and Robotics Laboratory 3DROV version 1.44 is used. During the different experiments the one-to-one scaled ExoMader model and the stereo camera component is used. The stereo cameras (NavCams) can be moved in a 6 DOF in both rotational and translational directions, which gives the possibility to adjust the model exactly the same as the real ExoMader rover NavCams. The rover can be moved in two ways, via the joystick, or by specifying coordinates in the cartesian space (also in the joint space).

Visualization of the different cameras attached to the rover is also supported. It can be monitored and footprints can be used on the terrain to mark the FOV of the camera, moreover the software supports image generation of these attached stereo cameras (both 8-bit and 24-bit).

By means of rover stereo camera pictures from the simulated environment and the actual ExoMader NavCams it will be investigated the accuracy level of the simulations. Also a sensitivity analysis will be carried out in the next Chapter.

5.5.1 Importing Components

A set of terrains will be used during our experimental campaign:

- The very precise digital elevation map of the planetary testbed obtained from six laser scans - with a high precision and without the real texture, the colors being projected by means of the elevation;
- The DEM from laser scan, projecting pictures as texture;
- 3D reconstructed terrain only from digital images with the help of [46] - with texture see Figure 5.11(d).



(a) Delanuy vertices reconstructed with the help of CMP Sfm [46]



(b) 3D reconstruction with texture using CMP Sfm [46]



(c) Digital elevation map without texture



(d) DEM orthophoto with corrections and added texture from Fig. 5.11(b)

Figure 5.11: Aerial views: transformation from 3D reconstruction to 3DROV supporting format

The terrain reconstructed had small imperfections as one can see in Figure 5.11(b). This is due to the fact that those specific regions (seen white) were weakly supported: very few images covered those locations. Despite this fact, for our current work

this problem could be solved. The purpose of this work is leading to the 3D visual representation of the terrain, and stereo camera pictures will be acquired from certain locations. As in Figure 5.11(c) it was achieved to reconstruct the terrain without imperfections. Both Figure 5.11(c) and Figure 5.11(d) are geometrically corrected, orthorectified, such that the scale is uniform. In this way it can be used to measure true distances.

The laser scanned terrain was exported to 3DROV format (CSV \rightarrow RAW \rightarrow PNG); similarly the 3D scene reconstruction (VRML \rightarrow PNG). So far 3DROV supports both CSV, LCSV and PNG import, by means of scripts. The PNG precision needs to be 16bit and the height is encoded as value of the pixel. By knowing the size of the terrain (in x and y direction) and the height (in z direction) one can convert it to 3DROV input format. Also different parameters need to be set (like the resolution of the input height-field texture, the patch resolution etc.).

The camera model is based on the real ExoMader stereo camera. It has a baseline of 10 cm and the camera parameters are tunable (resolution, field of view, bit depth).



Figure 5.12: ExoMader model on the laboratory testbed simulated in 3DROV

After a successful import of the different components (environment, camera and robot model) into 3DROV one can start to set up different test scenarios (as in Figure 5.12)).

5.5.2 Spatial Co-Registration and Image Acquisition

While the acquisition of image data may not be seen as directly related to image analysis and processing, the design decisions involved in this stage (i.e. camera type, lighting and optical system design) have a fundamental effect. It is very important to understand that good results can only be achieved with precise planning in the data acquisition stage.

Reality and Simulation Seen Through Stereo Cameras

The model of the terrain was aligned, resized, and rotated to be situated in the exact same position as the real testbed with its frame of reference. In order to demonstrate the accuracy of the simulation a comparison of stereo pictures taken from the real

testbed and from the simulation software 3DROV is intended.



(a) Undistorted image of the reality by using the left stereo camera

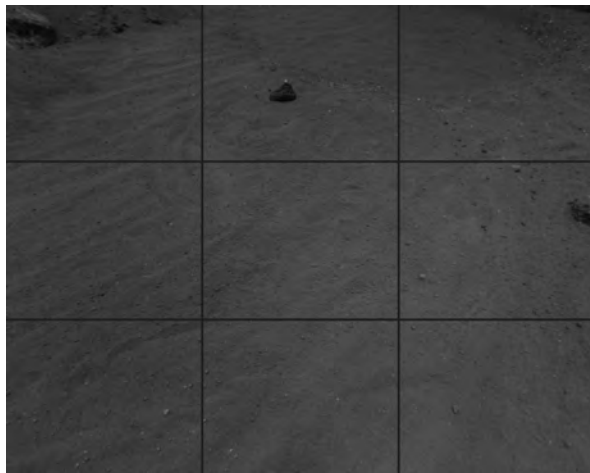


(b) Simulation as seen through the model of the left stereo camera

Figure 5.13: BW stereo camera pictures

The comparison was achieved with the following set-up:

1. Place the ExoMader on the terrain and record position
2. Obtain stereo pictures from ExoMader
3. Co-register the recorded position in 3DROV
4. Visual fine-tuning in 3DROV
5. Get the same images from the simulation software
6. Verification



(a) Rectangles used to ease fine tuning of the stereo camera orientation



(b) Projection of the grid in 3DROV seen from the main canvas camera

Figure 5.14: BW stereo camera pictures

Reflective balls were used to measure the position of NavCams from the reference system. Co-registration in 3DROV is done by placing a stereo camera model into the exact same location. Later, orientation was calculated by using classical

trigonometrical equations. The stereo camera model parameters were tuned according to the real ones. Visual fine tuning was necessary in order to obtain the very exact FOV, as seen in Figure 5.14. This was done by dividing the image seen by the camera into nine rectangles (see in Figure 5.14(a)). In simulation a grid was projected as texture to the surface, to ease this manual adjustment (see Figure 5.14(b)).

5.5.3 Experimental Scenarios

In this subsection the precision and the sensitivity of the terrains will be analyzed through the camera images in three different cases. The images from ExoMader on the testbed will be compared with two simulated environment captions taken from a camera model in 3DROV.

During the *first experiment*, the DEM created by multi-view reconstruction is used. This has the advantage of containing the texture. The size, height and orientation of the DEM in the simulator has been adapted in such a way that it became the exact replica of the real terrain. The origin has been selected also in the same corner, in the same direction and orientation as on the terrain at the ARL (see Figure 5.15). This will later help to place the stereo camera model in the exact same position as in reality.

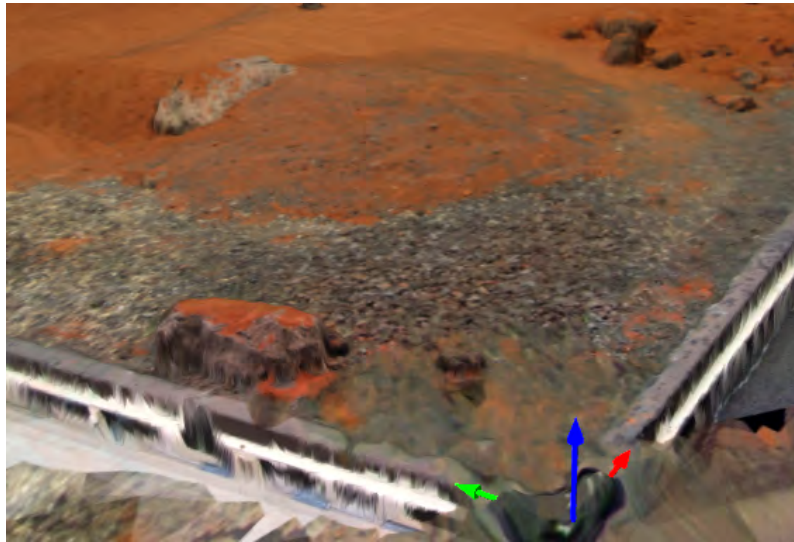


Figure 5.15: Origin in 3DROV on the multi-view reconstructed terrain

The reason behind using just the stereo camera model and not the whole ExoMader, which includes the stereo camera and the mast, is for the sake of simplicity. It is much more convenient to measure the spatial position and orientation of the ExoMader's stereo camera on the real terrain, and to transfer this information to 3DROV and adjust only a camera model.

During the *second experiment* the goal is to accurately project images of the real testbed on the DEM created by six laser scans. This is due to the fact that our laser scanning system is limited in the current set-up. It does not contain a high resolution DSLR camera, so it was needed to find other means for obtaining the texture. The solution adopted will combine undistorted images of the testbed (see the difference in Figure 5.16(a) and 5.16(b)). These will be projected on the DEM by a monoCam

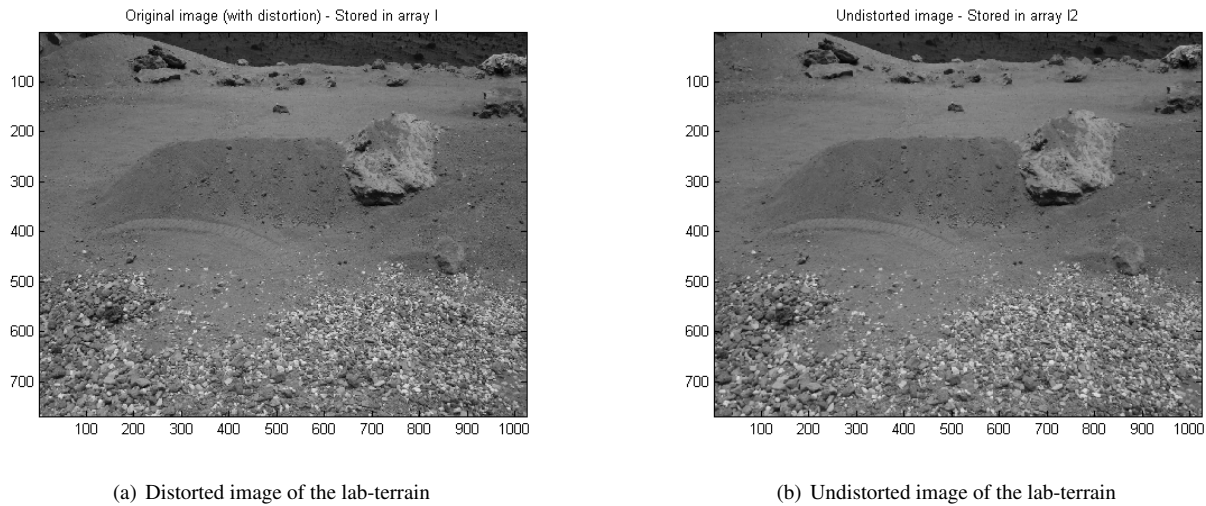


Figure 5.16: Results of automatic correction of lens distortion by Matlab toolbox

model in 3DROV having the same FOV.

Texture Projection

The precise DEM is shown as a monochrome height map in Figure 5.17. The colored image is projected on top of it, at an approximate same location as the image was taken. The stereoCam model will acquire again the same images from the same locations, as in reality. Because of the large number of parameters (two times 3 positions and 3 rotations) the data acquisition is not as precise as in the previous experiment. Although a precision method was used to place and orientate the different cameras, there is no perfect solution to this problem. This means that the simulated stereo camera FOV does not contain the exact same information as the real images. Further low level post-processing is necessary.

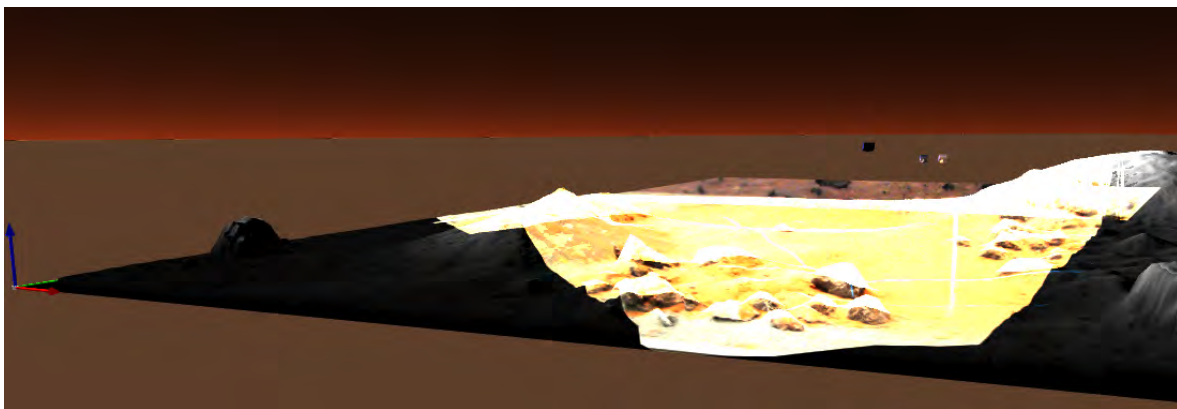


Figure 5.17: Texture projection

The resolution of a camera, or more precisely the pixel size, determines the image precision on the $X - Y$ plane. The closer the object to the sensor, the higher the number of pixels describing the object in the image. The information in the image decreases with an increasing distance between camera and object.

The point density will decrease with increasing distance of the object surface from the sensor. Considering the point density as the number of points per unit area, while the number of points remains constant, the area is proportional to the square distance from the sensor:

$$S \sim d^2 \quad (5.6)$$

Thus, the point density on the $X - Y$ plane is inversely proportional to squared distance from the sensor (inverse-square law):

$$\varrho \sim \frac{1}{d^2} \quad (5.7)$$

Knowing this and the fact that the laser scan method consists of a point cloud, where the individual points are measured separately, this particular method of projecting texture will suffer from precision on higher distances.

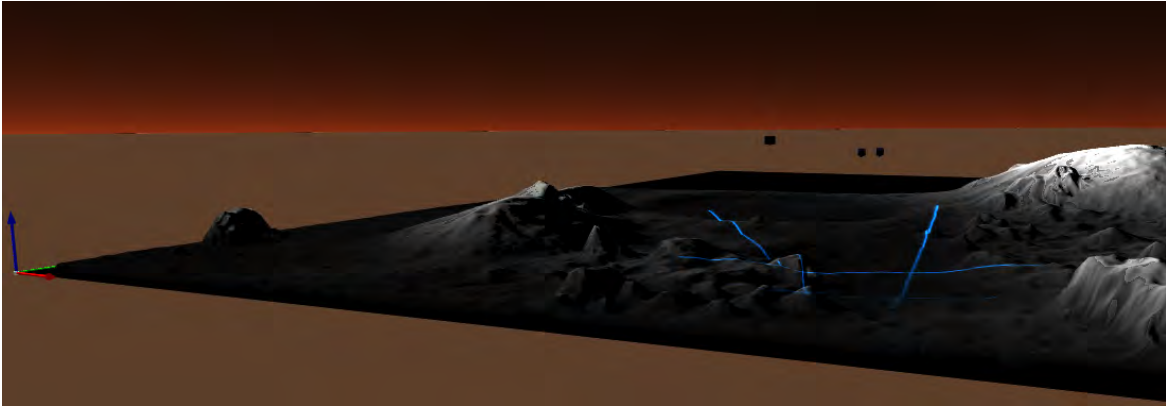


Figure 5.18: Precision correction method by using manual tuning

Another difficulty are the shadows shadows: it was not possible to fully remove them from the terrain, although the shadows of the camera models were disabled manually. Again, to obtain a higher precision, the image was divided into nine squares and manually tuned from the stereoCam orientation (see Figure 5.18).

In total eight different data sets were recorded which will be further analyzed by different feature searching algorithms. It is obvious that there is certain level of texture degradation in terms of precision. This will be analyzed qualitatively and quantitatively in Chapter 6. The three different data sets of several locations are found in Appendix.

Chapter 6

Analysis and Results

6.1 Laser Scan and Multi-View Stereo Reconstruction Comparison

In this section it will be demonstrated that recent approach to multi-view reconstruction [6] of the ARL testbed attain a degree of accuracy which is comparable to laser scan reconstruction. Furthermore, the former method contains texture information.

6.1.1 Height Maps

In the following, a qualitative comparison of the height-maps of the two DEMs will be presented. Figure 6.1 shows these in monochrome, 8-bit images, ranging from 0 to 255. As suspected, the DEM created by the laser scan (6.1(a)) shows a higher precision.

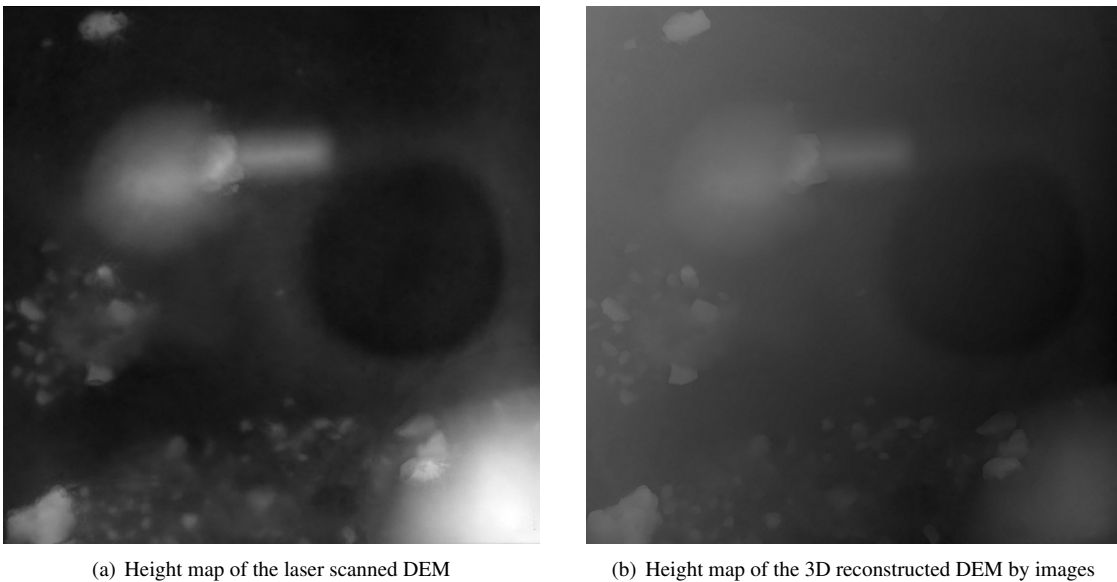


Figure 6.1: Comparison of simulated environments in 3DROV

One can clearly distinguish the better defined edges of the rocks, which are more defined in this height map. Nevertheless these edges contain much more information, being more accurate. This is the outcome of the method presented Section 3.1.1. It has only one major limitation: it does not contain the texture information due to equipment limitations.

A different approach was used in Section 3.2 to obtain textures. Although here is the proof that the multi-view stereo reconstruction lacks of very high precision, but also demonstrates that for certain applications it has a sufficient accuracy. The most important advantage of this method it is low cost.

Figure 6.2 shows the histograms of the two height maps presented before in Figure 6.1. It is plotted the number of pixels for each tonal value. From this, it is clear that both shows lack of highlights (right third of the histogram). Although a minimum of highlight information is found in the laser scan histogram. In the second case, the histogram is more uniform, suggesting that the accuracy of the map is lower. High peaks and tone variation on the laser scan histogram (left panel) suggests the opposite.

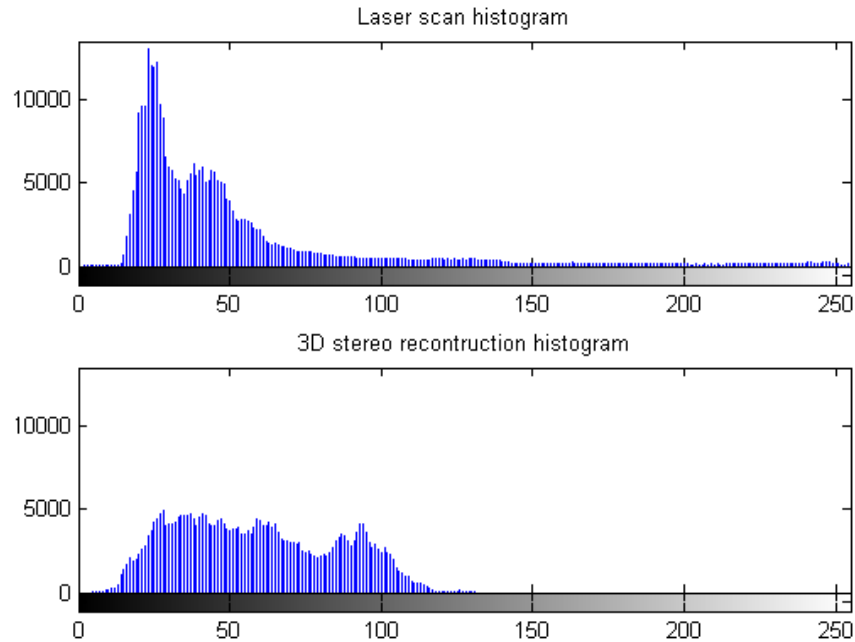


Figure 6.2: Histogram comparison of the height maps

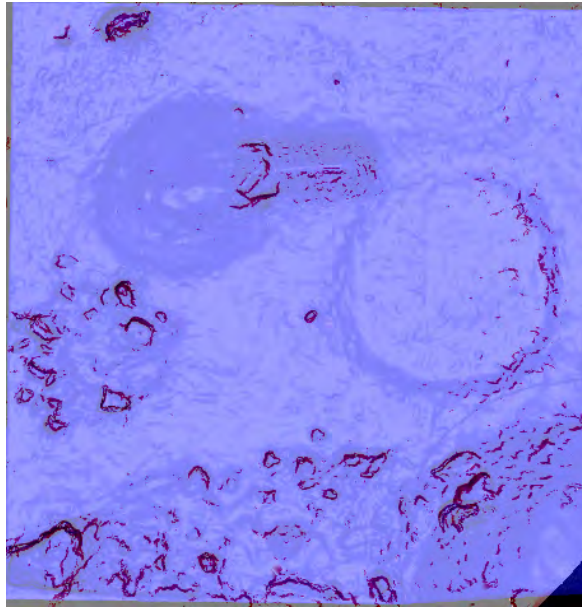
6.1.2 Triangle-Meshes

During this section the quantitative comparison of the two simulated digital elevation maps will be presented. It will be shown that multi-view stereo reconstruction technique is fairly precise compared to the laser scan.

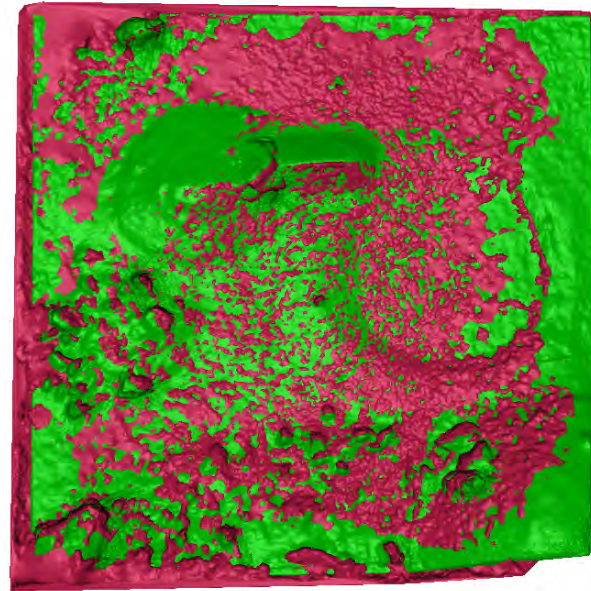
The point clouds in both cases were converted to Polygon File Format (PLY). PLY uses the description of the object as a list of nominally flat polygons. During the alignment of the two meshes several techniques were used (like manual corresponding points matching and automatically driven pairwise ICP¹).

¹Iterated Closest Points

In Scanalyze² the two were compared. Using different colors for the two meshes it is easier to visualize. It was necessary to scale down the multi-view reconstructed mesh by a factor of 0.68. This was done by the diagonal of the transformation matrix.



(a) Bird's eye view of manually matched meshes (purple: the edges of the multi-view reconstruction; blue: laser scan)



(b) Bird's eye view of ICP matched meshes (green: laser scan; red: multi-view stereo reconstruction)

Figure 6.3: Matched 3D meshes

Several surface matching methods has been analyzed. First manual matching was tried out around 20 points were matched on each surface. The result can be seen in Figure 6.3(a). It turned out that this method was un-accurate since the number of matching points are very limited.

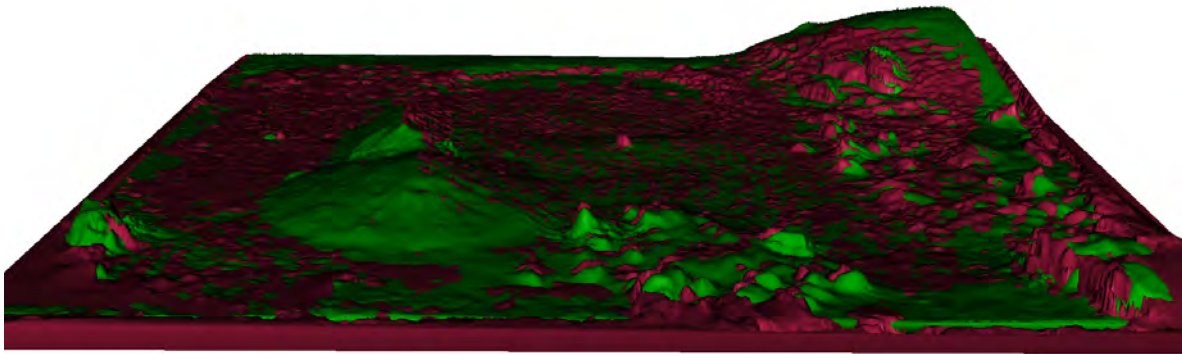


Figure 6.4: Side view of the matched surfaces with ICP (green: laser scan DEM; red: multi-view reconstruction DEM)

The built-in ICP function was used afterwards which offered robustness and higher matching accuracy. The ICP algorithm is widely used for geometric alignment of three-dimensional models when an initial estimate of the relative pose is known. If correct correspondences are known, it is possible to find correct relative rotation/translation.

²Stanford Computer Graphics Laboratory's research program

The method calculates the distance between points and planes and it minimizes the distance by iterations. The matching with ICP in Figure 6.3(b) needed 30 iterations. A sampling rate of 0.25 was used. The cull percentage was set to 3.

It is clear from Figure 6.4 that the weekly supported regions (mostly around the sides of the square) are not as accurate as the midmost parts. Surprisingly good surface matching is obtained at central part of the terrain (see the connecting lines in Figure 6.5).

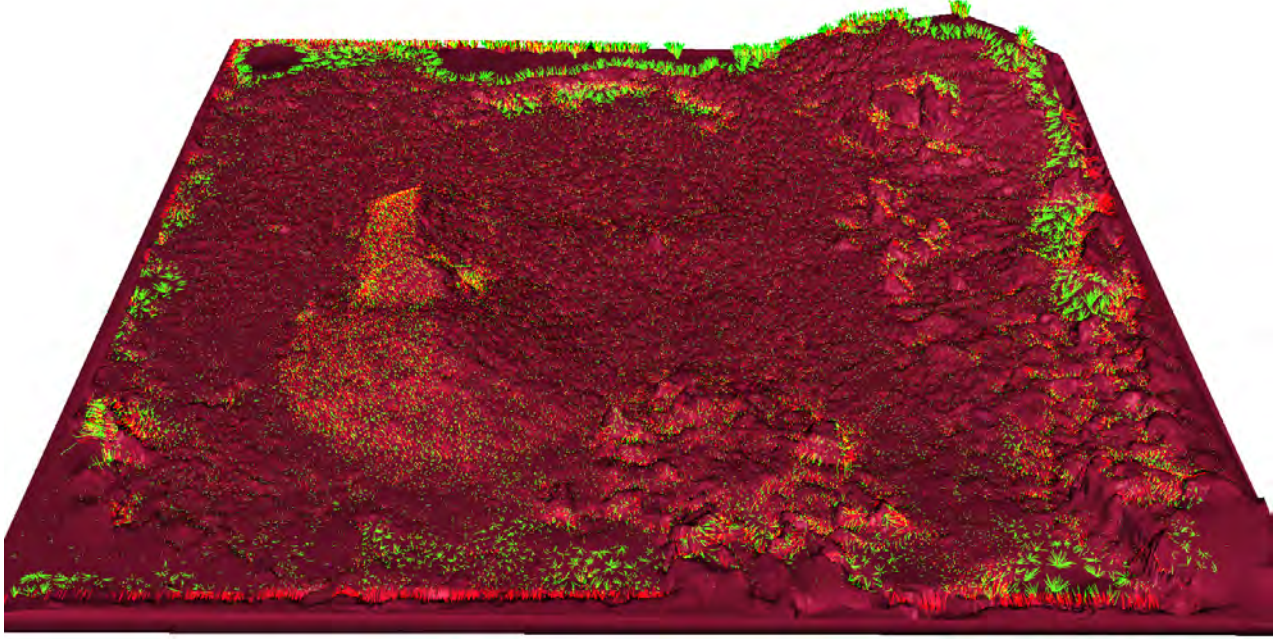


Figure 6.5: Error between the two meshes (red/green lines) on a multi-view reconstructed surface (8.9 m x 8.9 m)

In Section 5.2.2 the point cloud obtained by laser scan was converted to a mesh with 1.5 cm precision. That is why the comparison of the two meshes are done in the cm range. After 30 iterations the average point-to-point distance in x, y and z directions are $4.70499 \cdot 10^{-4}$, $3.1993 \cdot 10^{-4}$ respectively $2.6665 \cdot 10^{-4}$. The absolute error of the ICP algorithm was set to 0.5 cm. Figure 6.4 shows in false colors the automatically matched DEMs.

Since the terrain is approximately 76 m^2 , the error is acceptable in the cm range. The error convergence of the ICP method is shown in Figure 6.6 on page 67. The starting iteration shows a 2.2 m error, which is unacceptable. The method is proving robustness since the error is converging to approximately 3 cm after 30 iterations. This shows that the DEM created with the method of [6] is surprisingly precise when digitizing large objects (e.g. terrains).

6.2 Comparison of Stereo Camera Images

In this section it will be analyzed, compared and contrasted the rover navigation images. As a basis it will be considered the undistorted images taken by ExoMader stereo cameras. It will be compared with the simulated environment in 3DROV seen through a stereo camera model with the same properties as the real one.

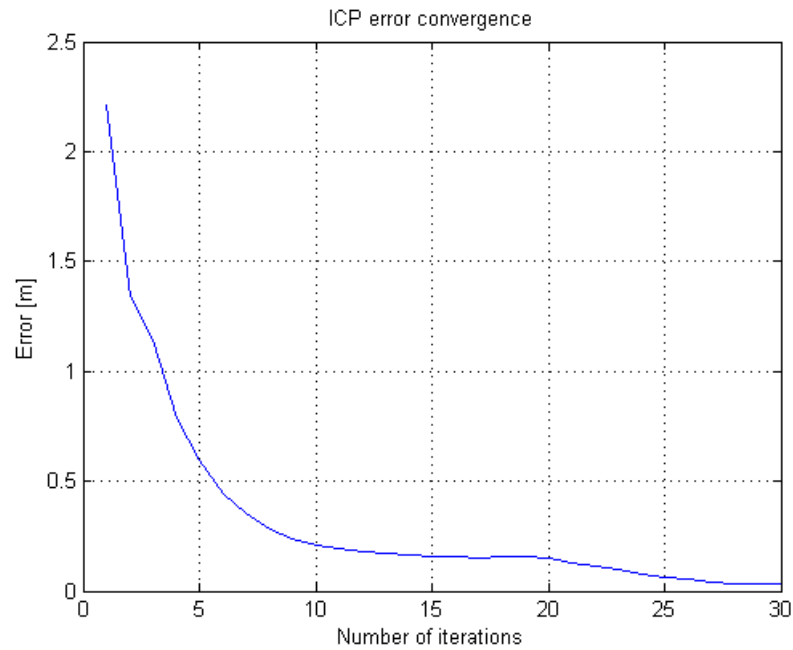


Figure 6.6: Error convergence of ICP algorithm

6.2.1 Contrasting Images

In order to start the image analysis with different feature detectors (Harris, SURF, SIFT) it was necessary to auto-contrast the data sets. MATLAB has certain auto-contrasting functions for doing such kind of operations, but it was considered to use a custom built program.

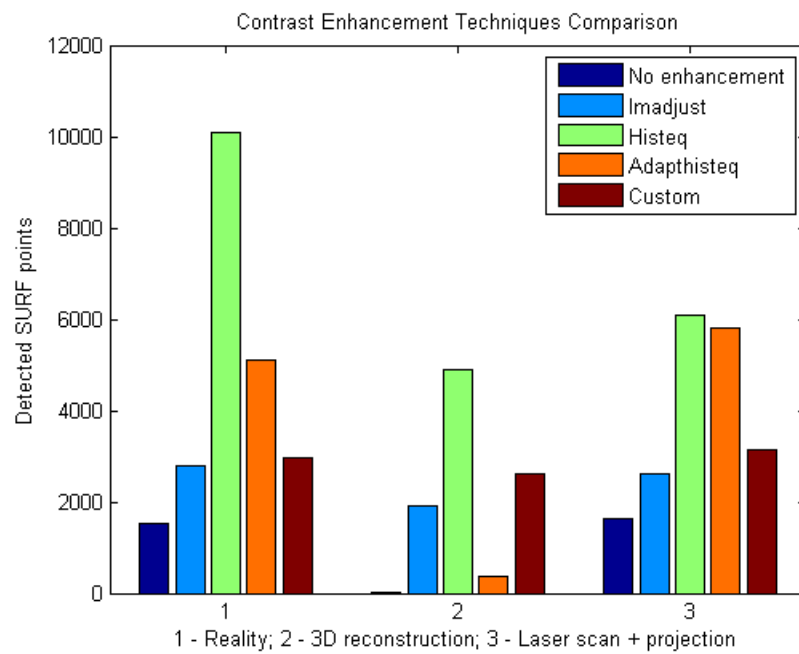


Figure 6.7: Contrast enhancement techniques comparison

A comparison of the different methods can be found in Figure 6.7 on page 67. The images from *Location 1* were used, since this is one of most accurate data set in terms of FOV. The number of SURF points are plotted in three different cases: the analysis is done on the real image (leftmost), on multi-view reconstruction (center) and finally on the laser scan and projection (right). It was found out that in the second case, without any contrast enhancement only one SURF point is detected (dark blue in the middle group in Figure 6.7). This led to the necessity of the use of image contrasting.

The idea was to select a method which delivers uniform images. Uniform in the sense that feature detectors can be applied. If one is interested in the overall picture: 3DROV capabilities must be demonstrated by means of rover navigation, which has not been done so far. Even today rover navigation algorithms are using stereo camera black and white images, so it came naturally the idea to compare navigation images from different spots in three different environments: one real and two simulated terrains.

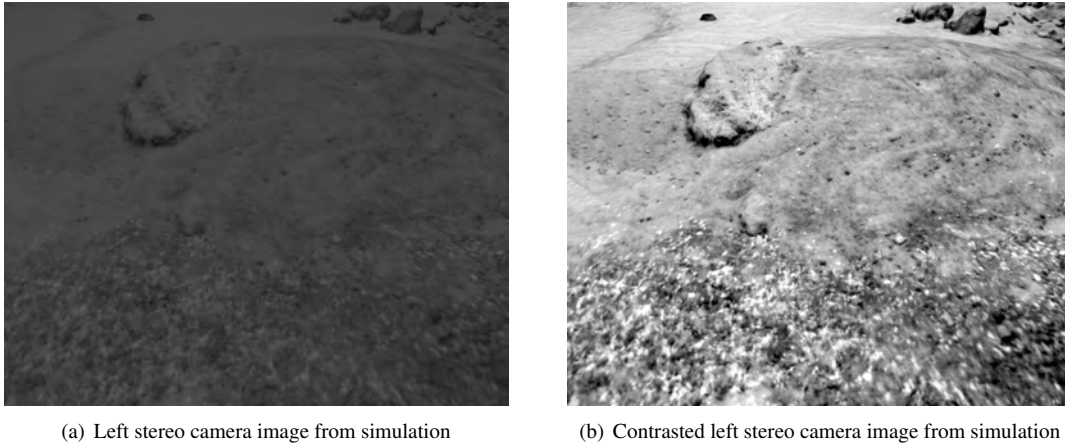


Figure 6.8: Before-after comparison of contrasting images

Imadjust function gives nice results by means of contrast but in the framework of this thesis, this technique was further enhanced by using a custom built function³. In Figure 6.7 one can compare all the methods which were tested. *Histeq* (green) gives a huge enhancement on the real image, but is not comparable with the simulated ones. *Adapthisteq* (orange) does not affect the multi-view reconstructed terrain (middle).

6.2.2 Preliminary Results

Since SURF is one of the most often used feature detector in literature, naturally came the idea to test the data with it. Another good reason why it was chosen is the high robustness and low computation time.

In Figure 6.9 one can consult the results after applying the SURF algorithm on the contrasted images. The number of SURF points are plotted in three cases for eight different locations:

- Images captured with Exomader's stereo camera on the ARL testbed (blue);

³Based on the work of Divakar R.

- Images taken with the model of the stereo camera in 3DROV on a multi-view reconstructed terrain (green);
- Pictures captured with the model of the stereo camera in 3DROV on a laser scanned DEM where high resolution undistorted images were projected as texture (red);

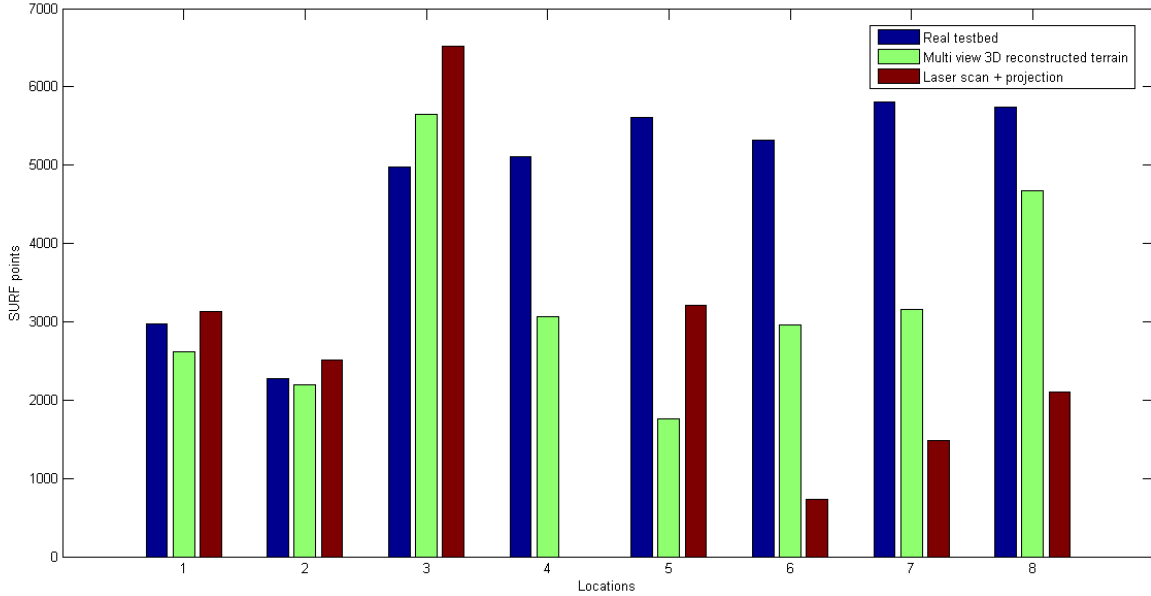


Figure 6.9: SURF points detected on unaligned images

A good correlation can be found in the first three cases (Location 1, 2 and 3). In the fourth case there is a lack of the last image set. Also in the last three cases (Location 6, 7 and 8) a significant variation of the numbers of SURF features are present (specifically in comparison of red bar to blue). This is due to the shadowing presented in Section 5.5.3. This method of image projection could have been replaced by the use of an advanced version of the current laser scanning system.

Generally speaking, less SURF points are detected on the images in the first simulation environment (green) than in reality (blue). This proves that the method of multi-view stereo reconstruction presented in Section 3.2 lacks of very high precision texture. The data set can be improved by feeding a higher number of images to the SFM pipeline [46] with a better image capturing technique.

Representative conclusions can be stated, since an average of 68.95 % of SURF points can be re-detected in the simulated DEM created by multi-view stereo reconstruction (Figure 6.9). More than two third of the features from the testbed are present in this simulation.

The best data sets are the first three. This if one considers the accuracy of the projection method. Comparing only the first three locations, there are more SURF features in the second simulated DEM (red) compared to real images. In average there is a higher SURF concentration in the simulation images with 19.04 % more. Although the same camera model is used with the same FOV and pixel-size.

6.2.3 FFT analysis

The use of the 2D Fast Fourier Transform (FFT) is a common method to compute the discrete Fourier transform of two images for comparison purposes in the field of image processing. Figure 6.10 shows this for both real and simulated cases. For comparison it is used the image from Exomader's camera looking at the terrain in the ARL from location three, and the simulated terrain created by multi-view reconstruction technique (see Appendix for the images).

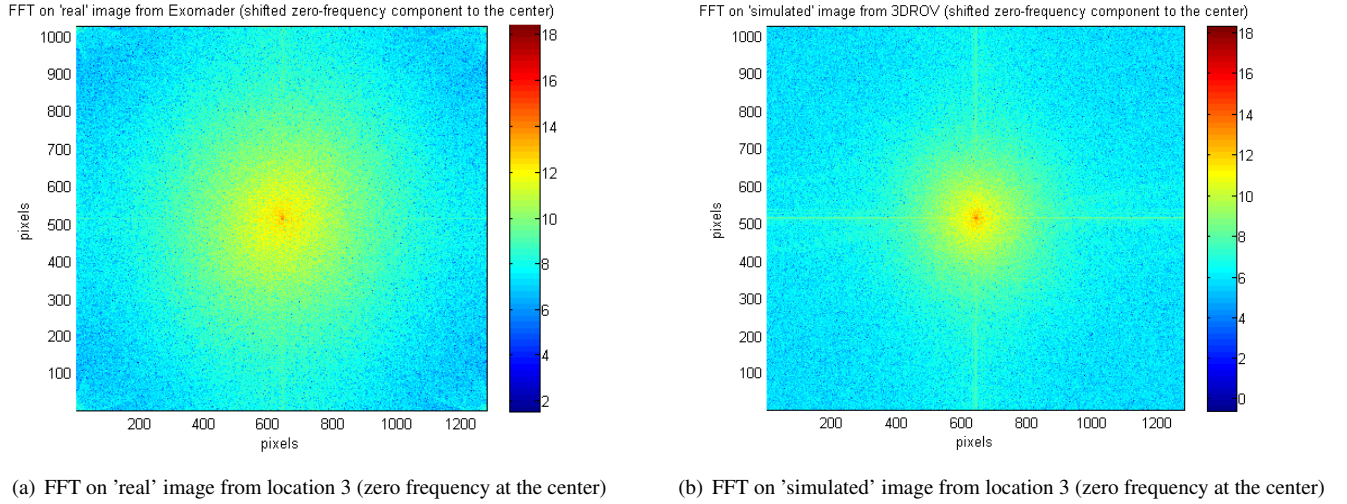


Figure 6.10: FFT representation of images from location three (logarithmic)

It is clear that the high frequencies in the simulation are damped which was obviously expected. Figure 6.10 shows the shifted zero-frequency component to center, in this way it is easier to look at the two. Obviously the simulated image consists of a smaller 'circle' and the 'cross' component is present in both.

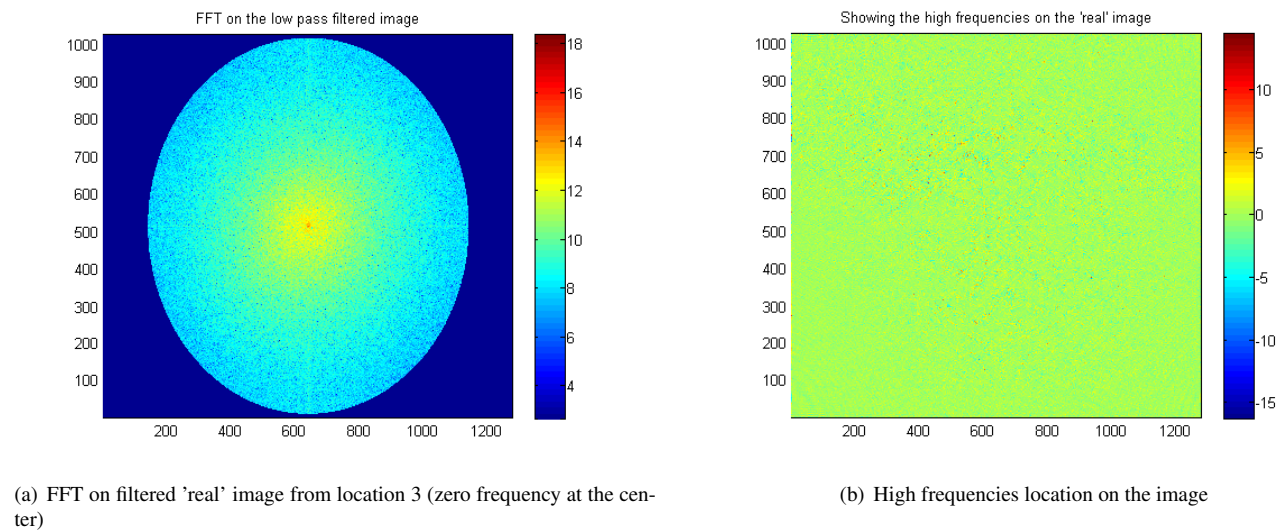


Figure 6.11: Filtering the 'real' image

Applying a low pass filter on the 'real' image is shown in Figure 6.11(a). This looks very similar to the simulation image

showed in Figure 6.10(b). The high frequencies location on the original image are plotted in Figure 6.11(b). As a conclusion it can be stated that although the high frequencies are not present in the simulation images, these contain still enough information. Later on this will be demonstrated with with feature detectors.

6.2.4 Aligning Images

Another challenge during the image acquisition process was found out to be the alignment of images. This had arisen from the cameras spatial co-registration in the simulation software. Though considerable efforts were made to find a solution to this issue (refer to Section 5.5.2), further work will be necessary in the future to find a completely satisfying method for alignment.

As a result, at some locations images were very slightly misaligned (see Figure 6.12). This affected the comparison only to a minor degree, but in order to have a better understanding and to deliver more precise results, alignment of the images was considered.



Figure 6.12: Re-aligned image: reality (lower layer), simulation (upper layer)

The aligning algorithm⁴ detects SIFT points in both images, and trying to match them. The non-correct matches are eliminated by the RANSAC⁵ algorithm [48]. Then the transformation matrix is computed and finally the image is repainted.

Figure 6.13 shows the correct SIFT points in both images which were matched. The non-correct matches are already eliminated. The green lines indicate the correct matches after RANSAC algorithm is applied. The red lines are the strongest matches based on statistics. Again, the lower part of the images contains significantly more matched points. As discussed in section 5.5.3 the point density will decrease with increasing distance of the object surface from the sensor, containing less

⁴The original code written by S. Kimmer and P. Wagner

⁵RANdom SAmple Consensus

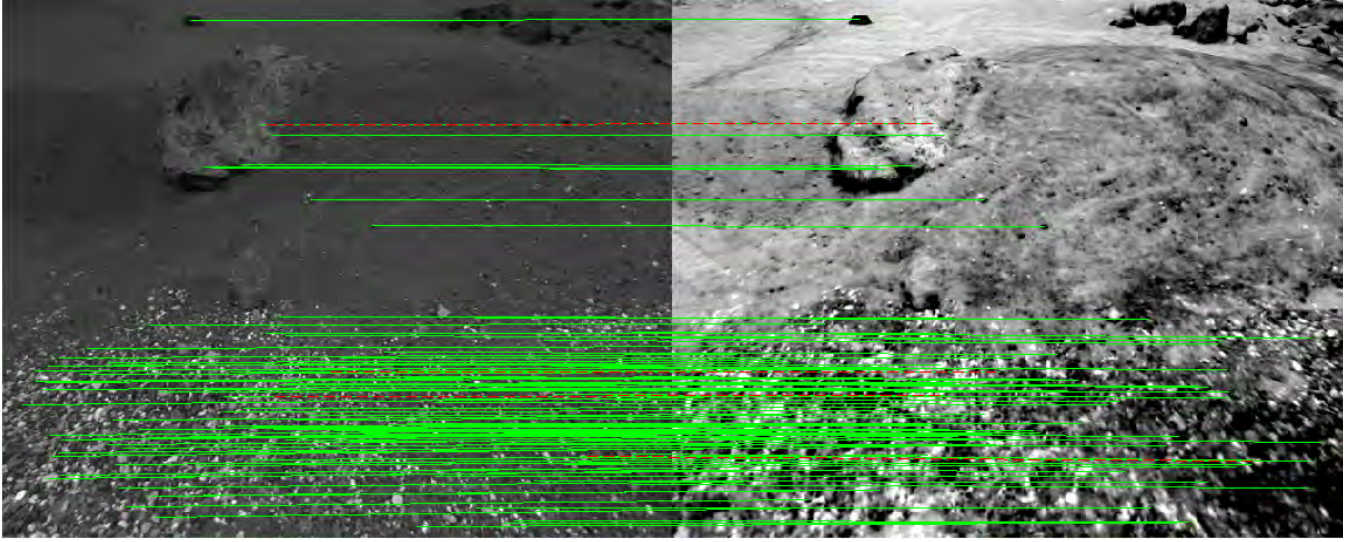


Figure 6.13: Correctly matched SIFT points for the purpose of image alignment

information. Secondly, the lower part contains a significant amount of pebbles, which provide a large number of features for the detectors.

A 50 % resolution degradation was necessary in order to cope with MATLAB's memory limitation. Also the upper part of the images at location three were eliminated due the shadows encountered during the projection of the second simulation. Finally all issues were solved and the analysis can be carried on.

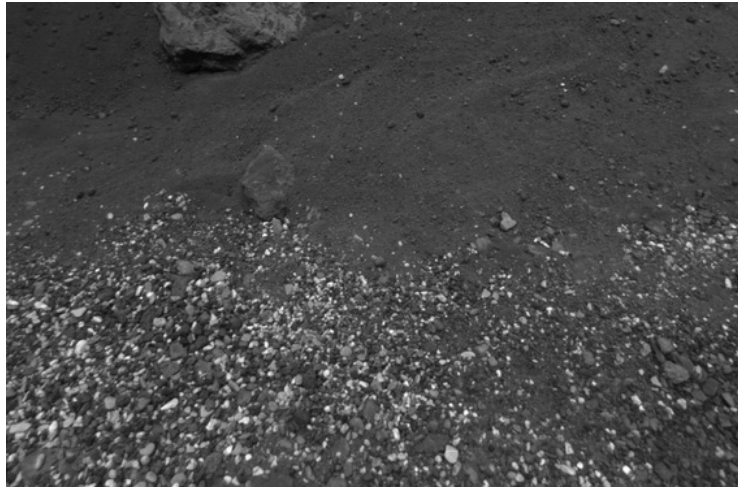
Qualitative comparison of image alignment can be found in Figure 6.14 on page 73. These are auto-contrasted, aligned and cropped at the region of interest (the rover near surrounding). Figure 6.14(a) shows the original image of the testbed, Figure 6.14(b) illustrates the aligned and auto-contrasted image from the multi-view stereo reconstructed simulated DEM, and finally Figure 6.14(c) contains the texture obtained from image projection on a laser scanned DEM.

Tests with different parameter values were carried during the image alignment procedure. The distortion ratio (γ), the threshold in pixel for RANSAC (ι) and the minimum number of inliers (τ) were varied. It was found out that the best results are given by using the following parameters:

- $\gamma = 0.95$;
- $\iota = 2.5$;
- $\tau = 8$.

The time of computation was not of concern in these tests. Simulation 1 in Table 6.1 refers to the scene constructed from images with multi-view stereo reconstruction technique. Simulation 2 in Table 6.1 is based one the laser scanned DEM and of high resolution image projection.

As shown in Table 6.1 the number of SIFT key-points matched in the first case is 35.44 % (Reality and Simulation1), in the second case 38.11 % (Reality and Simulation2). From this percentage the *valid matched points* are only 12.74 %, respectively 22.83 %. As already suspected this is in perfect accordance with the initial assumptions. Namely that the multi



(a) Original cropped image (reality)



(b) Aligned image from 3DROV (multi-view stereo reconstruction)



(c) Aligned image from 3DROV (laser scan and projection)

Figure 6.14: Comparison of real and simulated environments at location 3 in ARL respectively in 3DROV

Image	Nr. of key-points	Matches	Inliers	Iterations	Elapsed time
Real	3364 (limit. 2700)	N.A.	N.A.	N.A.	N.A.
Simulation1	6800 (limit. 2700)	957	122	39559	119.18 seconds
Simulation2	4234 (limit. 2700)	1029	235	5621	21.54 seconds

Table 6.1: Feature points matching with SIFT on 640 x 512 pixel images at location 3 (see Figure 6.13)

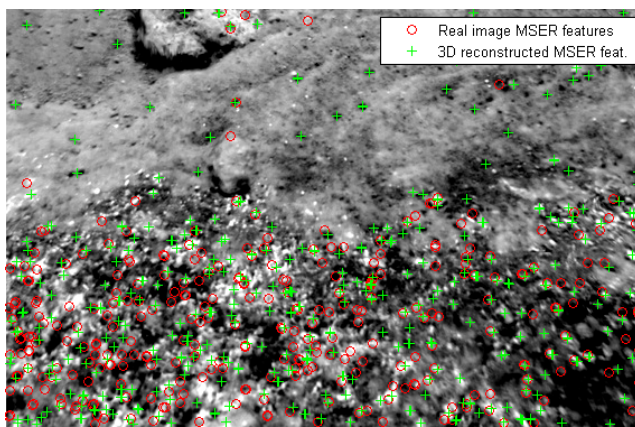
view reconstructed DEM has a lower accuracy by means of features as the latter one. Although the precision of both DEMs are comparable, the texture varies.

6.2.5 Features Location

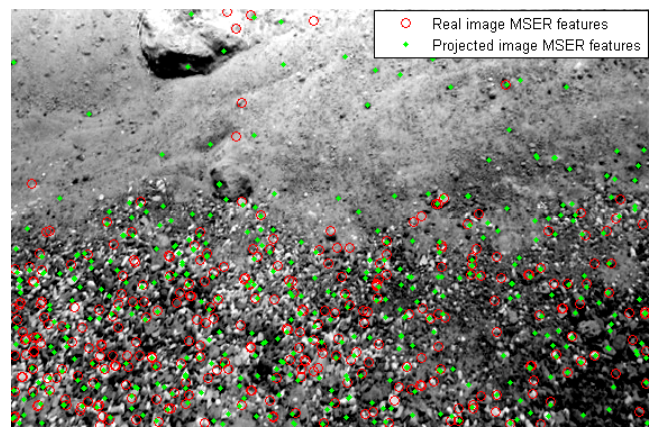
On the aligned images at location three MSER and SURF feature detection were tried out. The parameter settings of the algorithm were tried to be varied with experimental conditions and objectives (see Table 6.2). The objective was to find at the approximate same location the same features in both simulated and real environment.

Image	Detector	Region Area Range	Delta Threshold	Metric Threshold	Num Scale Levels
Figure 6.15(a) - real	MSER	[25 100]	2	N.A.	N.A.
Figure 6.15(a) - simulation	MSER	[25 100]	4	N.A.	N.A.
Figure 6.15(b) - real	MSER	[25 100]	2	N.A.	N.A.
Figure 6.15(b) - simulation	MSER	[25 100]	2	N.A.	N.A.
Figure 6.16(a) - real	SURF	N.A.	N.A.	2500	3
Figure 6.16(a) - simulation	SURF	N.A.	N.A.	3000	3
Figure 6.16(b) - real	SURF	N.A.	N.A.	2500	3
Figure 6.16(b) - simulation	SURF	N.A.	N.A.	5000	3

Table 6.2: Parameters of the applied feature detectors



(a) MSER features location on a multi-view reconstructed terrain image (green) compared to the real image (red)



(b) MSER features location on the texture projected laser scan terrain image (green) compared to the real image (red)

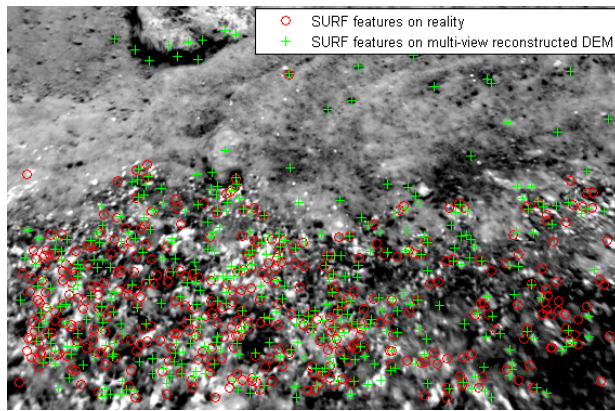
Figure 6.15: MSER region detection comparison

In Figure 6.15 aligned and contrasted images from simulations are used from location three. Different Δ_{th} were used to approximately match the number of MSER features. The region area range parameters were kept constant: [25 100].

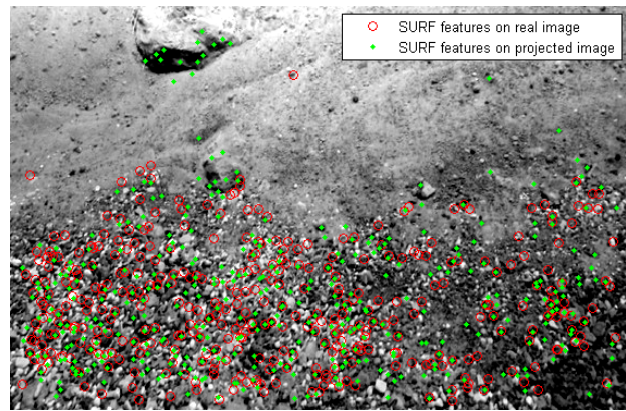
Surprisingly in Figure 6.15(a) a good number of features are located in the very same location. On both figures the red circles represents the MSER features detected in the image from the Exomader's camera. The two simulated environments created by the two different methods are shown in Figure 6.15. On those the MSER points are detected with green crosses respectively green dots. It was found out that approximately one third of the features are found to be on both simulations.

Since the lower part of the images contain only gravel there are certainly more features found there than in the upper half. The distribution of the detected points looks similar in both simulated cases. Apparently if the threshold is lower on the real image, stone edges are not detected (upper part of the image).

In the following, in Figure 6.16 the SURF feature points are detected. Similarly to the previous case only images from the simulations are used with the same conditions. Again the red dots are the SURF features detected on the Exomader's stereo camera image. These are compared with the SURF features location from 3DROV images. One can observe that the green crosses and dots are fairly close, or in many cases, they are at the exact same location as in the real case.



(a) SURF features location on a multi-view reconstructed terrain image



(b) SURF features location on the texture projected laser scan terrain image

Figure 6.16: SURF features comparison

Both simulated digital elevation maps, created with different methods, contain enough feature points for stereo navigation purposes. This was not yet fully demonstrated by actual navigation algorithms tests, but the throughout analysis of the images certainly imply that the simulation tool could be used for such an early phase mission validation.

Chapter 7

Conclusions and Future Work

Planetary rover missions are relatively new to both the European Space Agency, and the space industry in Europe. Testing Martian and lunar robots is an important concern, regardless of the phase in which the mission is. Until present, testing of rovers took place in various places around the globe, either in a natural environment (e.g. deserts in Spain or Chile) or in a pre-built test facility (e.g. Mars Yard). Simulations are a cost-effective alternative to the early phase mission validation and tests. Simulations have the advantage of flexibility and repeatability. Moreover, these systems could be tested thoroughly in realistic simulated environments. The present work paves the way to a highly representative virtual terrain. This will facilitate widespread adoption of simulations during the design phase of the mission, which will reduce the need of costly hardware tests. It is conceivable that in the future real life testing will only be necessary for the validation phase of a mission.

The current research project achieved the creation of digital maps of the planetary testbed found at the ESA's ESTEC site, involving university research group partners (Center for Machine Perception) and also industry (Trasys). Active and passive techniques were used to recreate the Martian-like site in a simulator. It was found that the novel method developed by the CMP group using multi-view stereo reconstruction (passive technique) is comparable to terrestrial laser scanning in terms of accuracy, and it is more cost effective. Although the testbed is about 80 m², precision close to 3 cm is achieved with the passive technique. This has comparable accuracy to terrestrial laser scanning systems which has 1.5 cm precision. This result underlines the high precision of the CMP algorithm.

The second, and most important objective of this Master's thesis, was the validation of virtual environments in order to support visual odometry of rovers. Different feature detectors and descriptors (e.g. SIFT, SURF, and MSER) were used to test the accuracy of rover stereo camera images. The main concern was in obtaining the same image from the same location in both simulated and real cases. This could be not fully achieved, so low level post-processing was used to align the images.

It was concluded that each method excelled in different areas. The passive method contains implicit texture information and gives a consistent texture accuracy. The active method with texture overlay (the laser scanned DEM together with the projection of high-resolution images) gives higher accuracy for small features (e.g. pebble, rocks). The drawback of this method is that the

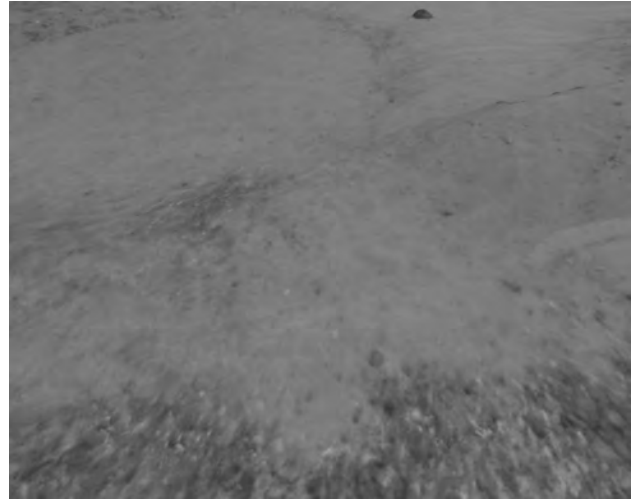
map is lacking the initial texture, so when testing in such an environment one needs to cope with projected shadows. Although the terrain created using the passive method is not precise to the millimeter, for rover navigation path planning purposes it has sufficient texture information.

With TLS equipment containing a high resolution camera, a laser scanned digital elevation map can be improved in terms of texture. The author would strongly recommend this for future work at the ARL. The near future work will examine in more detail the possibility of improving the present digital terrains. Moreover, the analysis will be extended with more feature descriptors and edge detectors. Also a navigation scenario experiment will be set-up to further validate the created terrains. The ultimate goal is to create a database of various types of terrains and large terrains for testing long navigation range. Using such a Martian terrain database in early or middle phase developments of rover testing could lead to substantial improvements in terms of rover navigation path-planning optimization, and mission validation.

Appendix



(a) Real terrain seen through the 'eyes' of Exomader



(b) Simulated terrain from 3D reconstruction by means of images

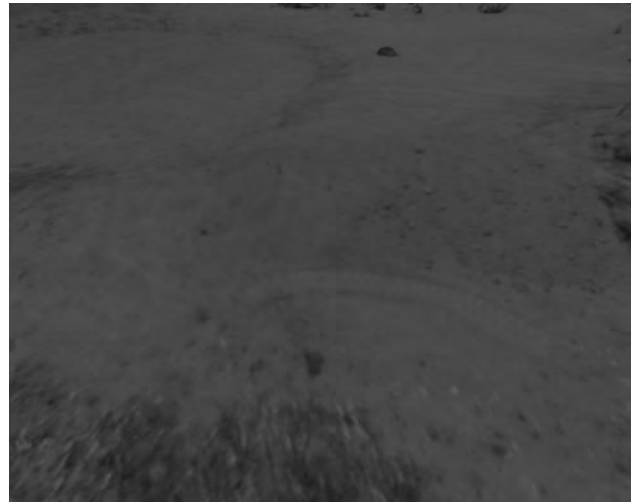


(c) Simulated terrain by means of laser scan and high resolution image projection

Data set of the first location



(d) Real terrain seen through the 'eyes' of Exomader



(e) Simulated terrain from 3D reconstruction by means of images

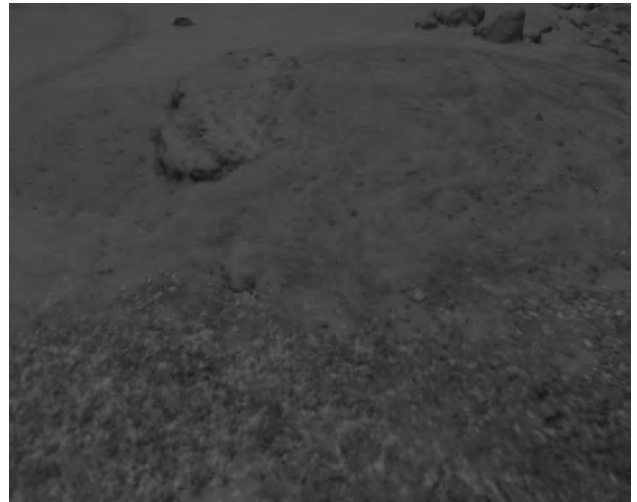


(f) Simulated terrain by means of laser scan and high resolution image projection

Data set of the second location



(g) Real terrain seen through the 'eyes' of Exomader



(h) Simulated terrain from 3D reconstruction by means of images



(i) Simulated terrain by means of laser scan and high resolution image projection

Data set of the third location

Acronyms list

3D	Three-Dimension
ARL	Automation and Robotics Laboratory
CAD	Computer Aided Design
CMP	Center for Machine Perception
CNES	Centre National d'Etudes Spatiales (National Centre for Space Studies)
CSV	Comma Separated Value
DEM	Digital Elevation Model (Map)
DLR	Deutsches Zentrum für Luft- und Raumfahrt (German Aerospace Center)
DoF	Degree of Freedom
DoG	Distance of Gaussian
DTM	Digital Terrain Model
ESA	European Space Agency
ESTEC	European Space research and TEchnology Centre
EXOMADER	EXOMArS DEMonstration Rover
FAST	Features from Accelerated Segment Test
FFT	Fast Fourier Transform
FOV	Field of View
FPS	Frames Per Second
GCP	Ground Control Points
GCS	Global Coordinate System
GIS	Geographical Information System
GNC	Guidance Navigation and Control
GOS	Ground Operation System
GRS	Ground Reference System
ICP	Iterated Closest Points
JPL	Jet Propulsion Laboratory
LASER	Light Amplification by Stimulated Emission of Radiation
LEO	Low Earth Orbit
LIDAR	Light Detection And Ranging
MER	Mars Exploration Rover
MORCA	MOBILE Robot Control Architecture
MSER	Maximally Stable Extremal Regions
NASA	National Aeronautics and Space Administration
NEA	Near Earth Asteroid
PCS	Project Coordinate System
PNG	Portable Network Graphics
RANSAC	RANdom SAmples Consensus
ROI	Region of Interest
RTAI	Real-Time Application Interface
SIFT	Scale-invariant feature transform
SNR	Signal-to-noise ratio
SOCS	Scanner's Own Coordinate System
SURF	Speeded-Up Robust Features
TLS	Terrestrial Laser Scanning

Nomenclature

(In order of appearance)

c	speed of light in vacuum
f	frequency
λ	wavelength
h	Planck's constant
\mathcal{P}	set of points
p	a point
$V(p)$	Voronoi cell associated to a point
$V(\mathcal{P})$	Voronoi diagram
$N_c(x)$	number of cameras associated with a point
T	tetrahedralization
A	matrix of intrinsic parameters
(u, v)	coordinates of projection point
(c_x, c_y)	principal point coordinates
(f_x, f_y)	focal length coordinates
$R t$	joint rotation/translation matrix
(X, Y, Z)	coordinates of a 3D point in the world coordinate space
k_i	radial distortion coefficients
p_i	tangential distortion coefficients
$L(r)$	lens mapping at a point in the object plane at distance r
M	transverse magnification
U	distortion function
E^N	Euclidean space
R	average intensity change
I	image
P	pixel
$I_\Sigma(x)$	entry of an integral image at location x
H	Hessian matrix
g	Gaussian function
σ	standard deviation of the Gaussian function
M_{ij}, R_{ij}	image gradient magnitude, respectively orientation
O_i	origin of SOCS
θ	vertical angle FOV
φ	horizontal angle FOV
α	field of view
d_x	dimension of one pixel in the x direction on chip
d_x	dimension of one pixel in the x direction
S	area
ϱ	point density
τ	minimum number of inliers
ι	threshold in pixel for RANSAC
γ	distortion ratio

Bibliography

- [1] “Mars exploration family portret,” May 2012. [Online]. Available: <http://www.astrosaur.us/>
- [2] S. Chhaniyara, C. Brunskill, B. Yeomans, C. Saaj, S. Ransom, and L. Richter, “Terrain trafficability analysis and soil mechanical property identification for planetary rovers: A survey,” *Journal of Terramechanics*, vol. 49, pp. 115–128, 2011.
- [3] “Electro magnetic spectrum and light,” May 2012. [Online]. Available: <http://9-4fordham.wikispaces.com/Electro+Magnetic+Spectrum+and+light>
- [4] Datasheet, *3D terrestrial laser scanner system: LMS-Z210i*, RIEGL.
- [5] P. Labatut, R. Keriven, and J.-P. Pons, “Efficient multi-view reconstruction of large-scale scenes using interest points, delaunay triangulation and graph cuts,” *Computer Vision Conference*, 2007.
- [6] M. Jancosek and T. Pajdla, “Multi-view reconstruction preserving weakly-supported surfaces,” *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3121–3128, 2011.
- [7] I. Biederman, “Recognition-by-components: A theory of human image understanding,” *Psychological Review*, vol. 2, no. 94, pp. 115–147, 1987.
- [8] T. Tuytelaars and K. Mikolajczyk, “Local invariant feature detectors: A survey,” *Foundations and Trends in Computer Graphics and Vision*, vol. 3, no. 3, pp. 177–280, 2007.
- [9] E. Rosten and T. Drummond, “Fusing points and lines for high performance tracking.” in *IEEE International Conference on Computer Vision*, vol. 2, October 2005, pp. 1508–1511. [Online]. Available: http://edwardrosten.com/work/rosten.2005_tracking.pdf
- [10] D. Lowe, “Object recognition from local scale-invariant features,” *Proceedings of the International Conference on Computer Vision*, vol. 2, pp. 1150–1157, 1999.
- [11] “Ueye - datasheet,” April 2005. [Online]. Available: http://www.klenc.cz/soubory/catalogue_USB_camera_uEye06_engl-Imasys.pdf

- [12] L. David, "Lunar lost and found: The search for old spacecraft," 2006. [Online]. Available: <http://www.space.com/2199-lunar-lost-search-spacecraft.html>
- [13] R. Christy, "Luna - exploring the moon." [Online]. Available: <http://www.zarya.info/Diaries/Luna/Luna21.php>
- [14] "Mars pathfinder," 2012. [Online]. Available: http://www.nasa.gov/mission_pages/mars-pathfinder/
- [15] "Mars," April 2011. [Online]. Available: <http://en.wikipedia.org/wiki/Mars>
- [16] "Water ice in crater at martian north pole," May 2011. [Online]. Available: http://www.esa.int/SPECIALS/Mars_Express/SEMGKA808BE_0.html
- [17] "Nasa spacecraft confirms martian water, mission extended," May 2011. [Online]. Available: http://www.nasa.gov/mission_pages/phoenix/news/phoenix-20080731.html
- [18] "Mars exploration rovers," 2012. [Online]. Available: <http://marsrover.nasa.gov/home/index.html>
- [19] "Chang'e 3," April 2011. [Online]. Available: http://en.wikipedia.org/wiki/Chang'e_3
- [20] "Chandrayaan-2," April 2011. [Online]. Available: <http://en.wikipedia.org/wiki/Chandrayaan-II>
- [21] ESA, "News from moscow," 2011. [Online]. Available: http://ice.sso.esa.int/intranet/public/docs/nfm/nfm2011_10.pdf
- [22] T. P. Gouache, N. Patel, C. Brunskill, G. P. Scott, C. M. Saaj, M. Matthews, and L. Cui, "Soil simulant sourcing for the exomars rover testbed," *Planetary and Space Science*, vol. 59, pp. 779–787, 2011.
- [23] M. V. Winnendael, P. Baglioni, A. Elfving¹, F. Ravera, J. Clemmet, and E. Re, "The exomars rover overview of phase b1 results," *International Symposium on Artificial Intelligence, Robotics and Automation in Space*, 2008.
- [24] M. D. West, J. D. Clarke, M. Thomas, C. F. Pain, and M. R. Walter, "The geology of australian mars analogue sites," *Planetary and Space Science*, vol. 58, no. 4, pp. 447–458, 2010.
- [25] M. Golombek, A. Haldemann, R. Simpson, R. Fergason, N. Putzig, R. Arvidson, J. B. III, and M. Mellon, "Martian surface properties from joint analysis of orbital, earth-based, and surface observations," *Cambridge University Press*, pp. 468–497, 2008.
- [26] A. Cumani and A. Guiducci, "Comparison of feature detectors for rover navigation," *Proceedings of the Applied Computing Conference*, 2009.
- [27] J. Marlow, Z. Martins, and M. Sephton, "Mars on earth: soil analogues for future mars missions," *Astronomy and Geophysics*, no. 49, pp. 2–20, 2008.
- [28] N. Patel, R. Slade, and J. Clemmet, "The exomars rover locomotion subsystem," *Journal of Terramechanics*, no. 47, pp. 227–242, 2010.

- [29] “Pozzolana,” April 2011. [Online]. Available: <http://en.wikipedia.org/wiki/Pozzolana>
- [30] P. Poulakis, L. Joudrier, S. Wailliez, and K. Kapellos, Eds., *3DROV: A Planetary Rover System Design, Simulation and Verification Tool*. i-SAIRAS, 2008.
- [31] “Laser,” May 2012. [Online]. Available: <http://en.wikipedia.org/wiki/Laser>
- [32] Datasheet, *3D terrestrial laser scanner system: General Description and Data Interfaces*, RIEGL, 2005.
- [33] Vicon, *Vicon MX Hardware System Reference*, revision 1.4 ed., Vicon Motion Systems Limited, 2006.
- [34] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski., “A comparison and evaluation of multi-view stereo reconstruction algorithms,” *CVPR*, 2006.
- [35] J.-D. Boissonnat and M. Yvinec, *Algorithmic Geometry*. Cambridge University Press, 1998.
- [36] “Opencv 2.0 reference,” May 2012. [Online]. Available: <http://opencv.willowgarage.com/documentation>
- [37] H. Ojanen, “Automatic correction of lens distortion by using digital image processing,” 1999.
- [38] G. Bradski and A. Kaehler, *Learning OpenCV*. O’Reilly Media, Inc., 2008.
- [39] R. Laganière, *OpenCV 2 Computer Vision Application Programming Cookbook*. Packt Publishing, 2011.
- [40] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, “Speeded-up robust features (surf),” *Computer Vision and Image Understanding*, pp. 346–359, 2008.
- [41] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” *Proceedings of the Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 511–518, 2001.
- [42] “Sift,” May 2012. [Online]. Available: http://en.wikipedia.org/wiki/Scale-invariant_feature_transform
- [43] J. Matas, O. Chum, M. Urban, and T. Pajdla, “Robust wide-baseline stereo from maximally stable extremal regions,” in *Proceedings of the British Machine Vision Conference*, pp. 384–393, 2002.
- [44] “Maximally stable extremal regions,” June 2012. [Online]. Available: <http://en.wikipedia.org/wiki/MSER>
- [45] M. Alba and M. Scaioni, “Comparison of techniques for terrestrial laser scanning data georeferencing applied to 3-d modelling of cultural heritage,” *Int. Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 36, p. 8, 2007.
- [46] “Cmp sfm web service,” April 2011. [Online]. Available: <http://ptak.felk.cvut.cz/sfmservice/>
- [47] P. Deléglise, “Calibration tool,” European Space Research and Technology Centre, Internal Document, 2011.
- [48] M. A. Fischler and R. C. Bolles, “Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography,” *Comm. of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.