Universitat de les Illes Balears

DOCTORAL THESIS
2019

NEW INSIGHTS ON LASER-BASED STRUCTURED LIGHT FOR UNDERWATER 3D RECONSTRUCTION

Miguel Francisco Massot Campos

Universitat
de les Illes Balears

DOCTORAL THESIS
2019

Doctoral Programme of Information
and Communications Technology

NEW INSIGHTS ON LASER-BASED STRUCTURED LIGHT FOR UNDERWATER 3D RECONSTRUCTION

Miguel Francisco Massot Campos

Thesis Supervisor: Dr. Gabriel Oliver Codina
Thesis Tutor: Dr. Alberto Ortiz Rodríguez

Doctor by the Universitat de les Illes Balears

Statement of Authorship

This thesis has been submitted to the *Escola de Doctorat*, *Universitat de les Illes Balears*, in fulfilment of the requirements for the degree of *Doctor en Tecnologies de la Informació i les Comunicacions*. I hereby declare that, except where specific reference is made to the work of others, the content of this dissertation is entirely my own work, describes my own research and has not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university.

Miguel Francisco Massot Campos

Palma de Mallorca, September, 2019

Funding

The work reported in this thesis was supported by the Ministry of Education of Spain (FPI grant ES-2012-054352), Spanish Ministry of Research and Innovation DPI2011-27977-C03, TIN2014-58662-R, DPI2014-57746-C3-2-R, by the FP7 project TRIDENT (FP7-ICT-2009-4, ID: 248497) and FEDER Funding.

Supervisor's Agreement

I, Gabriel Oliver Codina, Ph.D. in Physics and Professor at the *Department of Mathematics and Computer Science*, *Universitat de les Illes Balears*

ATTEST THAT:

this dissertation, titled *New Insights on Laser-Based Structured Light for Underwater 3D Reconstruction*, submitted by Miguel Francisco Massot Campos for obtaining the degree of *Doctor en Tecnologies de la Informació i les Comunicacions*, was carried out under my supervision and contains enough contributions to be considered as a doctoral thesis.

Dr. Gabriel Oliver Codina

Palma de Mallorca, September, 2019

Abstract

Three-dimensional visual maps of the seafloor provide objective information to reliably characterise these scenes. These maps are of great study value for archaeological sites or geologically interesting areas. Often, spatial scales over 8 orders of magnitude are needed, from the millimetre resolution to the $> 100\ km^2$, to recognise key features of interest to areas over which they are distributed. Delivering precise resolution at scale has been an issue for the last decade. Nowadays, new vision based sensors such as lasers and structured light provide a higher resolution than known acoustic sensors. Using these innovative sensors poses new challenges in underwater localization as most of the mapping error comes now from the self-localization rather than from the measurement sensor.

This thesis presents two different solutions to the visual mapping problem. The first part of the PhD presents a novel laser-based structured light system to increase 3D perception resolution, accuracy and frame rate when compared to acoustic counterparts and cameras. The system consists of a laser Diffractive Optical Element (DOE) that diffracts the beam in 25 parallel planes and a camera to recover the intersected lines to the seafloor. We propose a calibration procedure and solve the correspondence problem using a Maximum Spanning Tree algorithm. The experimental results show that the system draws a better representation of the objects in front and outperforms plain stereoscopy in featureless scenarios.

The second part of the thesis uses a standard laser stripe instead, to cast a single line on the seafloor, in a bathymetric SLAM solution to correct both the navigation and overall shape of the sensed environment. Two different algorithms are presented, one is sub-map bathymetric SLAM, which saves small map portions to be registered at a later stage using ICP and BPSLAM, a 2.5D grid that treats every scan as a particle in a Particle Filter to find the position that better suits the known map. The developed bathymetric SLAM algorithms are tested in a close-to-shore small rocky area in Valldemossa using the AUV SparusII and in a large survey in the Hydrate Ridge using the AUV AE2000f.

Resumen

Los mapas visuales tridimensionales del fondo marino proporcionan información objetiva para caracterizar de manera confiable dichas áreas.Estos mapas son también de gran valor de estudio para sitios arqueológicos o áreas geológicamente interesantes. A menudo, se necesitan escalas espaciales de más de 8 órdenes de magnitud, desde la resolución milimétrica hasta los $> 100\ km^2$, para reconocer las características clave de interés para las áreas sobre las cuales se distribuyen. La resolución precisa a escala ha sido un problema en la última década. Hoy en día, los nuevos sensores basados en visión, como los láseres y la luz estructurada, ofrecen una resolución más alta que los sensores acústicos conocidos. El uso de estos novedosos sensores plantea nuevos desafíos en la localización subacuática, ya que la mayoría de los errores del mapa provienen ahora de la autolocalización y no del sensor de medida.
Esta tesis presenta dos soluciones diferentes al problema del mapeo visual. La primera parte de PhD presenta un novedoso sistema de luz estructurada basado en láser para aumentar la resolución de percepción 3D, la precisión y la velocidad de fotogramas en comparación con sus homólogos acústicos y cámaras. El sistema consta de un elemento óptico difractivo (DOE) con láser que difracta el haz en 25 planos paralelos y una cámara para recuperar las líneas intersectadas al fondo marino. Proponemos un procedimiento de calibración y resolvemos el problema de correspondencia utilizando un algoritmo de árbol de expansión máxima. Los resultados experimentales muestran que el sistema consigue una mejor representación de los objetos y supera la estereoscopía simple en escenarios sin características visuales.
La segunda parte de la tesis usa un láser estándar de una línea en una solución batimétrica SLAM para corregir tanto la navegación como la forma general del entorno detectado. Se presentan dos algoritmos diferentes, uno es SLAM batimétrico a partir de sub-mapas, que guarda porciones de mapa peque nas para ser registradas en una etapa posterior utilizando ICP y BPSLAM, una cuadrícula 2.5D que trata cada medida láser como una partícula en un filtro de partículas para encontrar la posición que mejor se adapte al mapa conocido. Los algoritmos batimétricos desarrollados de SLAM se prueban en una peque na área rocosa cercana a la costa en Valldemossa utilizando SparusII y en Hidrate Ridge usando AE2000f.

Resum

Els mapes visuals tridimensionals del fons marí proporcionen informació objectiva per caracteritzar de manera fiable aquestes àrees. Aquests mapes són també de gran valor d'estudi per a llocs arqueològics o àrees geològicament interessants. Sovint, es necessiten escales espacials de més de 8 ordres de magnitud, des de la resolució mil·limètrica fins als $> 100\ km^2$, per reconèixer les característiques clau d'interès per a les àrees sobre les quals es distribueixen. La resolució precisa a escala ha estat un problema en l'última dècada. Avui dia, els nous sensors basats en visió, com els làsers i la llum estructurada, ofereixen una resolució més alta que els sensors acústics coneguts. L'ús d'aquests nous sensors planteja nous reptes en la localització subaquàtica, ja que la majoria dels errors del mapa provenen ara de l'autolocalització i no del sensor de mesura.

Aquesta tesi presenta dues solucions diferents al problema del mapatge visual. La primera part presenta un nou sistema de llum estructurada basat en làser per augmentar la resolució de percepció 3D, la precisió i la velocitat de fotogrames en comparació dels seus homòlegs acústics i càmeres. El sistema consta d'un element òptic difractiu (DOE) amb làser que difracta el feix en 25 plans paral·lels i una càmera per recuperar les línies intersectades al fons marí. Proposem un procediment de calibratge i resolem el problema de correspondència utilitzant un algoritme d'arbre d'expansió màxima. Els resultats experimentals mostren que el sistema aconsegueix una millor representació dels objectes i supera la estereoscòpia simple en escenaris sense característiques visuals.

La segona part de la tesi fa servir un làser estàndard d'una línia en una solució batimètrica SLAM per corregir tant la navegació com la forma general de l'entorn detectat. Es presenten dos algoritmes diferents, un és SLAM batimètric a partir de sub-mapes, que guarda porcions de mapa petites per ser registrades en una etapa posterior utilitzant ICP i BPSLAM, una quadrícula 2.5D que tracta cada mesura làser com una partícula en un filtre de partícules per trobar la posició que millor s'adapti al mapa conegut. Els algoritmes batimètrics desenvolupats de SLAM es proven en una petita àrea rocosa propera a la costa a Valldemossa utilitzant SparusII i en Hidrate Ridge usant AE2000f.

Acknowledgements

I would like to begin by thanking my supervisor Dr. Gabriel Oliver for his guidance and support throughout the past five (and a bit) years. Systems, Robotics and Vision research group has provided me with an environment in which to think, learn, experience and grow in the underwater field of research.

To Toni, Eric, Xesc and Pepllu for the summer nights by the sea waiting for the laser surveys to complete. And to the rest of the SRV crew, especially Emilio, Xisco, JP, Inés, Alberto - and everyone else - have all helped in some way to make my experience rewarding, I also owe a special thank you. These countless coffee breaks to breathe whithin a day and share our PhD student lives. My PhD life would have probably been more productive, but way less fun without you. Thanks for making this journey more enjoyable.

I would also like to thank Dr. Blair Thornton for inviting me, first to the University of Tokyo and gifting me a 445nm laser to follow-up research. The discussions we had have improved this thesis. I am also grateful for Dr. Yvan Petillot valuable ideas at Ocean Systems Laboratory. Thank you for inviting me to Heriot-Watt University.

A big thanks to ODEMAR (by Dr. Javier Escartín) crew for our days at sea on board *Pourquoi Pas? R/V* and later on Adaptive Robotics (by Dr. Blair Thornton) on board *Falkor R/V*. The experience and know-how I gained there were invaluable.

Now let me write some words in Spanish below.

Me reservo mi último y más grande agracedimiento para mi familia y amigos. Para mis padres por creer en mí y hacer que crea en mi mismo. Sin su apoyo y comprensión no habría sido posible. Finalmente, me gustaría agradecer a mi esposa, Josmy, por su apoyo y amor a lo largo de esta etapa. No puedo describir con palabras todas las formas en las que ella me ha ayudado. Gracias por traer las cosas más importantes a mi vida.

Contents

List of Figures

List of Tables

List of Algorithms

List of Acronyms

1D One-Dimensional

2D Two-Dimensional

3D Three-Dimensional

AHRS Attitude and Heading Reference System

APD Avalanche Photodiodes

AUV Autonomous Underwater Vehicle

CAD Computer-Aided Design

CBEE Consistency Based Error Estimation

CCD Charge-Coupled Device

CMOS Complementary Metal-Oxide-Semiconductor

CTX Context Camera

CW-LLS Continuous Wave Laser Line Scanning

DIDSON Dual-Frequency Identification Sonar

DOE Diffractive Optical Element

DOI Digital Object Identifier

DOF Degrees Of Freedom

DPM Distributed Particle Mapping

DR Dead Reckoning

DVL Doppler Velocity Log

EKF Extended Kalman Filter

FW-SL Fixed Wavelength Structured Light

FOV Field of View

GICP Global Iterative Closest Point

GNSS Global Navigation Satellite System

GPS Global Positioning System

GPU Graphic Processing Unit

GT Ground Truth

HD High Definition

HG Heyne Greenstein

ICP Iterative Closest Point

IMU Inertial Measurement Unit

INS Inertial Navigation System

IS Imaging Sonar

KF Kalman Filter

KLT Kanade-Lucas-Tomasi

LBL Long-BaseLine

LC Loop Closure

LED Light-Emitting Diode

LIDAR Light Detection And Ranging

LMS Least Mean Square

LS Laser Stripe

LLS Laser Line Scanning

MB Multibeam

MRO Mars Reconnaissance Orbiter

MST Maximum Spanning Tree

NASA National Aeronautics and Space Administration

NOAA National Oceanic and Atmospheric Administration

PC Personal Computer

PG-LLS Pulse Gated Laser Line Scanning

PhS Photometric Stereo

PID Proportional-Integral-Derivative

PMT Photo Multiplier Tube

PN-MP Pseudorandom coded Modulated Pulse

PSNR Peak Signal-to-Noise Ratio

RANSAC Random Sample Consensus

RGB Red-Green-Blue

RBPF Rao-Backwellized Particle Filter

ROI Region Of Interest

ROS Robot Operating System

ROV Remotely Operated Vehicle

SfM Structure from Motion

SAM Smoothing and Mapping

SAS Synthetic Aperture Sonar

SB Single Beam

SBL Sort Base Line

SI International System of Units (French: Système International d'unités)

SL Structured Light

SLAM Simultaneous Localization and Mapping

SNR Signal to Noise Ratio

SONAR Sound Navigation and Ranging

SRV Systems, Robotics and Vision group

SSS Sidescan Sonar

ST-MP Single Tone Modulated Pulse

SURF Speeded Up Robust Features

SIFT Scale-Invariant Feature Transform

SV Stereo Vision

ToF Time of Flight

UIB University of the Balearic Islands

USBL Ultra Short Base Line

USS United States Ship

UUV Unmanned Underwater Vehicle

UWTD Underwater Target Detection

VSF Volume Scattering Function

VW-SL Variable Wavelength Structured Light

Symbols and Notation

XYZ	Generic coordinate frame axes. For the case of a robot body frame, respectively, longitudinal, lateral and vertical axes.
$x/y/z$	Position along the $X/Y/Z$ axis in a specific coordinate frame
$\varphi/\theta/\psi$	Roll/pitch/yaw. Rotation around the $X/Y/Z$ axis of a body coordinate frame
$\dot{x}/\dot{y}/\dot{z}$	Linear velocity along the $X/Y/Z$ axis of a body coordinate frame
$\dot{\varphi}/\dot{\theta}/\dot{\psi}$	Angular velocity around the $X/Y/Z$ axis of a body coordinate frame
$\ddot{x}/\ddot{y}/\ddot{z}$	Linear acceleration along the $X/Y/Z$ axis of a body coordinate frame
$\mathbf{x}_k$	State at time k
$\mathbf{F}_k$	State transition model at time k within a Kalman Filter
$\mathbf{w}_k$	Process noise at time k within a Kalman Filter
$\mathbf{Q}_k$	Process noise covariance at time k within a Kalman Filter
$\mathbf{z}_k$	Sensor measurement at time k
$\mathbf{H}_k$	Observation model at time k within a Kalman Filter
$\mathbf{v}_k$	Observation noise at time k within a Kalman Filter
$\mathbf{R}_k$	Observation noise covariance at time k within a Kalman Filter
Σ	Covariance
$\sigma_{dist}/\sigma_{vel}$	Covariance of a distance/velocity measure
$\mathcal{P}$	Point cloud
$\mathcal{T}$	Rigid transform, including a translation and a rotation

Chapter 1

Introduction

In this chapter, the objectives of the thesis are presented, followed by an introduction to the scope and some basic concepts of this dissertation. Finally, the structure of the document and the publications derived from this work are enumerated.

1.1 Objectives of the Thesis

The aim of this thesis is to progress towards the definition of methods to reconstruct the underwater environment in 3D, even in featureless regions. Diverse methodologies are presented suitable for applications requiring a different resolution: a one-shot reconstruction aimed to close-range, manipulation applications and two self-consistent, high-resolution surveys using bathymetric SLAM. A common characteristic for the methods developed is the use of laser-based structured light sources, which allows operating in featureless regions. In particular, the following objectives are pursued:

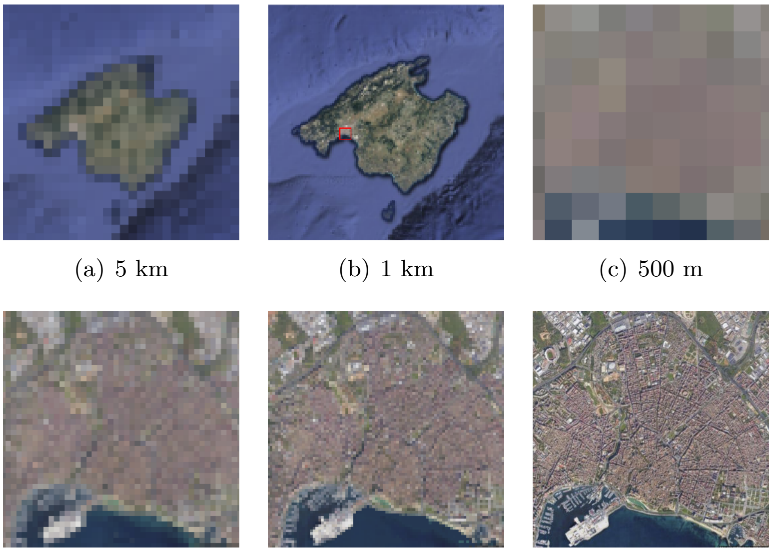
- To better understand underwater 3D reconstruction methods and sensors.
- To be able to map featureless objects and/or terrains.
- To achieve self-consistent underwater maps.

Two methodologies are proposed: (1) for a close-range, high frame rate and uniform spatial resolution, a multiline laser structured light device is presented; (2) for a medium-range survey-like environment, two existing SLAM frameworks are proposed to a novel domain, i.e. underwater laser bathymetry.

1.2 Motivation

1.2.1 Underwater exploration

The world's oceans cover a 71% of our planet and are of great interest to mankind, although parts of space are better known and researched than the seafloor. According to the National Aeronautics and Space Administration (NASA) [10], the Context Camera (CTX) on NASA's



(a) 5 km (b) 1 km (c) 500 m

(d) 100 m (e) 50 m (f) 10 m

Figure 1.1: How different maps of Palma de Mallorca would look like if the island was mapped from 5 km resolution to 10 m.

Mars Reconnaissance Orbiter (MRO) has been taking images of Mars for more than 10 years at an average resolution of 20 m covering more than the 99% of the planet. The images are sharp enough to show the shape of features as small as a tennis court ($24 \times 10\,m$) while being 54.6 million kilometres away from Earth.
The ocean seafloor is completely explored, although at 5 km resolution from space. Only about a 10%-15% coverage has been mapped *in situ* at 100 m resolution, given the high degree of difficulty and cost in exploring our ocean using technologies such as sonar to generate maps of the seafloor. According to the National Oceanic and Atmospheric Administration in the United States (NOAA) [11], more than 80 percent of our ocean is unmapped, unobserved and unexplored. The reader is encouraged to see the resolution differences between 5 km and 10 m in figure 1.1, where the island of Mallorca has been pixelated to match the scale. The figure clearly shows that at resolutions higher than 100 m it is impossible to see streets and avenues. Even at 10 m cars or pedestrians remain unnoticeable.

In order to provide a closer look into the ocean, underwater seafloor exploration can be accomplished using underwater submersibles and robots to capture images of the seafloor and the environment. In the field of this research, Unmanned Underwater Vehicles (UUV) can be mainly classified in two groups: Remotely Operated Vehicles (ROVs) and Autonomous Underwater Vehicles (AUVs). An ROV is an unmanned underwater robot that is connected to a ship by cables. These transmit command and control signals between the operator and the ROV, allowing remote navigation of the vehicle. An AUV operates independently from the ship as it has no connecting cables and carries a battery that normally is the most limiting factor in mission endurance. The platforms are normally conceived for different kinds of

operation. ROVs are targeted at delicate work such as rock sampling, assembly and repair of underwater structures, vessel hull inspection among others, whilst AUVs can be useful for long and repetitive missions such as detecting and mapping submerged wrecks, rocks, and obstructions that can be a hazard to navigation for commercial and recreational vessels.
In the last decade, UUVs have been established as a tool for sea exploration, inspection and intervention. The use of high-resolution seafloor bathymetry is becoming increasingly routine in marine research. ROVs have reduced the need for manned submersibles whilst increasing safety, mission time and repeatability. AUVs have increased their autonomy and lowered the costs of gathering ocean data and imagery. These vehicles and the sensors they carry present some challenges at multiple levels that still need to be addressed. In particular, underwater imaging absorption and scattering, and the generation of high resolution bathymetry.

1.2.2 Underwater optical imaging and laser light

The performance of traditional optical imaging systems such as cameras are limited by absorption and scattering when used underwater. Absorption is a loss of direct illumination due to the interaction of photons with water molecules, and scattering describes small changes within the light path caused by reflections and refractions caused by temperature transitions, suspended particles, dissolved solids or biological *snow*. Scattering in particular is different for different size particles, ranging from Mie to Rayleigh, and the scattering patterns are very different based on the ratio of wavelength to particle size [12]. To overcome absorption one may think that the solution is to increase the light power. But that would also increase backscattering and it might blind the receiver, losing contrast. Even when a system has been optimized to reduce backscatter it may become limited by absorption. In this situation, the propagating signal (light) is too weak to be detected by the corresponding sensor, and the system is said to be power limited. If the power is increased the scattering increases. It can increase so much that the sensor cannot differentiate the true signal from the noise. In this case the system is said to be contrast limited and can be measured with the Signal to Noise Ratio (SNR). Other authors have explored polarization filters to enhance underwater images [13]. In his paper, Schechner and Kapel show how taking multiple images of the same scene at different polarisation angles can increase contrast and colour correction.

The performance in any case can be enhanced by choosing the light source wavelength to match the optimal underwater wavelength that minimizes both absorption and scattering coefficients. As seen in figure 1.2, absorption and scattering coefficients vary depending on the wavelength of the light source. In order to transmit the maximum light these coefficients have to remain low. Blue-green colour spectra present a good compromise between absorption and scattering. Laser light is a type of light source whose wavelength can be tailored to fit in this regard. A blue-green laser light source will not only achieve larger distances, but also filters can be used to discard scattered light, such as bandpass filters.
A fixed-wavelength laser light source can be chosen to minimize these terms and since it is

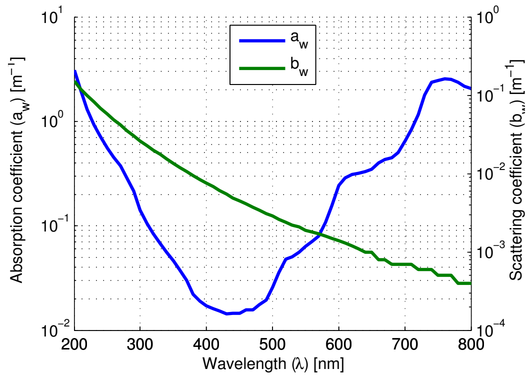


Figure 1.2: Absorption and scattering coefficients in pure seawater. Reproduced from Smith and Baker (1981).

inherently collimated, a high irradiance is achieved with low backscatter interaction, increasing the achievable imaging range underwater. This is accomplished by illuminating a small volume of water instead of a flooding light to every direction. Laser lines, laser patterns or just pointers would therefore illuminate a small volume of water. If the illuminated seafloor is grabbed using an imaging sensor, sparse three-dimensional measurements will be gathered.
Particularly, in underwater manipulation where an UUV has landed or is close to the seafloor, it is common that suspended particles increase scattering. In such conditions, being able to recover a sparse three-dimensional model may be enough to grasp an object [14].

1.2.3 High resolution sensors

On an UUV, Inertial Navigation Systems (INS), Attitude Heading Reference Systems (AHRS), Doppler Velocity Log (DVL), Global Positioning System (GPS), Ultra Short Baseline (USBL) and/or Long Baseline (LBL) acoustic positioning systems all provide options for improving navigation accuracy, each with varying levels of attainable precision [15]. A stand alone dead-reckoning solution would naturally degrade over time with an error position of $< 0.1\%$ of the distance travelled. This error will depend on the navigations sensor used. For example, Phins Subsea states that their dead-reckoning will degrade up to 0.05% of the distance travelled [16].
While the instruments used to generate seafloor maps have significantly increased the spatial resolution of bathymetric maps, this improvement is only meaningful if it can be matched by an accurate vehicle localization.
The use of an Extended Kalman Filter (EKF) can bound estimates of position uncertainty. As a shipboard USBL typically has an uncertainty of 1% of the slant range, the resulting position accuracy does not produce consistent visual maps. LBL acoustic positioning also provides navigation estimates with bounded error but this requires additional infrastructure to be in place, as well as the LBL transponder net to be accurately surveyed in. Furthermore LBL transponder nets are subject to a trade-off between accuracy and coverage.

Applied directly for map making, all of these standard positioning methods will contribute to errors and inconsistencies. Thus, it would be beneficial to develop additional positioning constraints beyond the direct measurements to obtain visually consistent maps.

1.3 Document Overview

The dissertation is divided into six chapters as follows:

- **Chapter 2** extensively reviews the background for this thesis, the published research on imaging based underwater reconstruction and its sensors during the last fifteen years.
- **Chapter 3** presents a complete three dimensional reconstruction pipeline and two calibration methods for laser-stripe-based structured light sensors.
- **Chapter 4** presents a novel one-shot laser structured light sensor
- **Chapter 5** describes two SLAM frameworks: a submap SLAM and a particle filter SLAM framework. These two approaches are compared using two different AUV missions.
- **Chapter 6** concludes this dissertation by summarizing the main contributions of the thesis and by highlighting the differences of the introduced approaches with other similar solutions. Some future work to extend the research described here is also suggested.

1.4 Related Publications

Parts of this thesis have been published in international journals and conference proceedings. The following list gives an overview about the individual publications.

Journal Articles

- **Miquel Massot Campos** and Gabriel Oliver-Codina, **Optical Sensors and Methods for Underwater 3D Reconstruction**, *MDPI Sensors*, Dec. 2015, vol. 15, no. 12, pp. 31525-31557. DOI: 10.3390/s151229864 [17].
- Francisco Bonin-Font, Gabriel Oliver, Stephan Wirth, **Miquel Massot Campos**, Pep Luís Negre, and Joan Pau Beltran, **Visual Sensing for Autonomous Underwater Exploration and Intervention Tasks**, *Ocean Engineering*, 2014, vol. 93, pp. 25-44. DOI: 10.1016/j.oceaneng.2014.11.005 [18].

Conference Proceedings and Workshops

- **Miquel Massot Campos**, Blair Thornton and Gabriel Oliver. **Laser stripe bathymetry using particle filter SLAM**, in *IEEE/MTS Oceans*, 2019, in press.

- Michael Leat, Adrian Bodenmann, **Miquel Massot Campos** and Blair Thornton, **Analysis of Uncertainty in Laser-Scanned Bathymetric Maps** in *IEEE/OES Autonomous Underwater Vehicles (AUV)*, 2018 [19].

- **Miquel Massot Campos**, Gabriel Oliver, Adrian Bodenmann and Blair Thornton, **Submap bathymetric SLAM using structured light in underwater environments**, in *IEEE/OES Autonomous Underwater Vehicles (AUV)*, 2016, pp. 181-188. DOI: 10.1109/AUV.2016.7778669 [20].

- **Miquel Massot Campos**, Francisco Bonin Font, Pep Luís Negre Carrasco, Eric Guerrero, Antoni Martorell and Gabriel Oliver Codina, **A 3D mapping, obstacle avoidance and acoustic communication payload for the AUV SPARUS II**, in *Instrumentation Viewpoint*, 2016, no. 19, pp. 31-33. ISSN: 1886-4864 [21].

- **Miquel Massot Campos**, Gabriel Oliver, Hashim Kemal, Yvan Petillot and Francisco Bonin-Font, **Structured light and stereo vision for underwater 3D reconstruction**, in *IEEE/MTS Oceans*, 2015, DOI: 10.1109/OCEANS-Genova.2015.7271433 [22] .

- **Miquel Massot Campos** and Gabriel Oliver Codina, **One-Shot Underwater 3D Reconstruction**, in *Proc. 19th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA 2014)*, 2014, DOI: 10.1109/ETFA.2014.7005282 [23].

- **Miquel Massot Campos** and Gabriel Oliver Codina, **Underwater laser-based structured light system for one-shot 3D reconstruction**, in *Proc. IEEE Sensors*, 2014, DOI: 10.1109/ICSENS.2014.6985208 [24].

Technical Reports

- **Miquel Massot Campos** and Gabriel Oliver Codina, **Survey on Underwater 3D Reconstruction**, in *Technical Reports*, 2014, no. A-04-2014, Universitat de les Illes Balears, pp. 1-10 [25].

1.5 Unrelated Publications

In parallel to this work, European and National projects such as TRIDENT, TRITON or MERBOTS were developed. Publication in other fields not directly related to this thesis, but relevant in research.

Journal Articles

- F. Bonin-Font, J. Lalucat, G. Oliver-Codina, **M. Massot**, E. Guerrero Font, and P. L. Negre Carrasco. **Evaluating the impact of sewage discharges on the marine**

environment with a lightweight AUV Marine Pollution Bulletin, vol. 135, no. July, pp. 714-722, 2018.

- J. Escartín, C. Mevel, S. Petersen, D. Bonnemains, M. Cannat, M. Andreani, N. Augustin, A. Bezos, V. Chavagnac, Y. Choi, M. Godard, K. Haaga, C. Hamelin, B. Ildefonse, J. Jamieson, B. John, T. Leleu, C. MacLeod, **M. Massot**, P. Nomikou, J. Olive, M. Paquet, C. Rommevaux, M. Rothenbeck, A. Steinführer, M. Tominaga, L. Triebe, R. Campos, N. Gracias, R. Garcia. **Tectonic structure, evolution, and the nature of oceanic core complexes and their detachment fault zones (13 20' N and 13 30' N, Mid Atlantic Ridge)**. In Geochemistry, Geophysics, Geosystems 18 (4), 1451-1482, February, 2017.

- N. Palomeras, A. Peñalver, **M. Massot**, P. L. Negre, J. Javier, P. Ridao, P. J. Sanz, G. Oliver. **I-AUV Docking and Panel Intervention at Sea**. In Sensors, Basel (Switzerland), MDPI, vol. 16, no. 1673, October, 2016.

- J. Escartín, F. Leclerc, J. Olive, C. Mevel, M. Cannat, S. Petersen, N. Augustin, N. Feuillet, C. Deplus, A. Bezos, D. Bonnemains, V. Chavagnac, Y. Choi, M. Godard, K. Haaga, C. Hamelin, B. Ildefonse, J. Jamieson, B. John, T. Leleu, C. MacLeod, **M. Massot**, P. Nomikou, M. Paquet, C. Rommevaux, M. Rothenbeck, A. Steinführer, M. Tominaga, L. Triebe, R. Campos, N. Gracias, R. Garcia, M. Andreani. **First direct observation of coseismic slip and seafloor rupture along a submarine normal fault and implications for fault slip history**. In Earth and Planetary Science Letters, Elsevier, vol. 450, pp. 96-107, September, 2016.

- F. Bonin-Font, **M. Massot**, P. L. Negre, G. Oliver, J. P. Beltran. **Inertial Sensor Self-Calibration in a Visually-Aided Navigation Approach for a Micro-AUV**. In Sensors, MDPI, vol. 15, no. 1, pp. 1825-1860, 2015.

Conference Proceedings and Workshops

- E. Guerrero, F. Bonin-Font, P. L. Negre, **M. Massot**, G. Oliver. **USBL Integration and Assessment in a Multisensor Navigation Approach for AUVs**. In The 20th World Congress of the International Federation of Automatic Control (IFAC World Congress), Toulouse, 2017.

- F. Bonin-Font, **M. Massot**, P. L. Negre, G. Oliver, E. Guerrero, E. García. **Towards a new Methodology to Evaluate the Environmental Impact of a Marine Outfall Using a Lightweight AUV**. In MTS/IEEE Oceans , Aberdeen, 2017.

- E. García, A. Ortiz, **M. Massot**. **Visual Control of an AUV for Multi-Robot Intervention Tasks**. In Jornadas Automar (Marine Automation Workshop), Castelló, 2017.

- E. García, J. P. Company, A. Ortiz, **M. Massot**, G. Oliver. **Multifunctional Cooperative Marine Robots for Intervention Domains: Target Detection, Tracking and Recognition Issues**. In Jornadas Nacionales de Robótica (Spanish Robotics Workshop), Valencia, 2017.

- F. Bonin-Font, **M. Massot**, G. Oliver. **Towards Visual Detection, Mapping and Quantification of Posidonia Oceanica using a Lightweight AUV**. In IFAC International Conference on Control Applications in Marine Systems, Trondheim, pp. 500-505, 2016.

- E. Guerrero, **M. Massot**, P. L. Negre, F. Bonin-Font, G. Oliver. **An USBL-Aided Multisensor Navigation System for Field AUVs**. In IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI) , Baden-Baden, pp. 430-435, 2016.

- P. L. Negre, F. Bonin-Font, **M. Massot**, G. Oliver. **Stereo-Vision Graph-SLAM for Robust Navigation of the AUV SPARUS II**. In IFAC Workshop on Navigation, Guidance and Control of Underwater Vehicles (NGCUV), Girona, 2015.

- Jorge Sales, **Miquel Massot Campos**, Pep Lluís Negre, Gabriel Oliver, and Pedro J. Sanz, **Integración de técnicas ópticas de reconstrucción 3D para mejorar la planificación de agarres en tareas de manipulación arqueológica subacuática**, in *XXXVI Jornadas Nacionales de Automática*, 2015 [26].

- F. Bonin-Font, A. Ćosić, **M. Massot**, G. Oliver. **Visual Characterization and Automatic Detection of Posidonia Oceanica for Meadows Mapping using an AUV**. In International workshop on Marine Technology (Martech), Cartagena, 2015.

- N. Palomeras, A. Peñalver, **M. Massot**, G. Vallicrosa, P. L. Negre, J. Javier, P. Ridao, P. J. Sanz, G. Oliver, A. Palomer. **I-AUV Docking and intervention in a subsea panel**. In IEEE/RSJ International Conference on Intelligent Robots and Systems , Chicago, Illinois, USA, 2014.

- J. Escartín, F. Leclerc, M. Cannat, S. Petersen, N. Augustin, A. Bezos, D. Bonnemains, V. Chavagnac, Y. Choi, M. Godard, K. Haaga, C. Hamelin, B. Ildefonse, J. Jamieson, B. John, T. Leleu, **M. Massot**, C. Mevel, P. Nomikou, J. Olive, M. Paquet, C. Rommevaux, M. Rothenbeck, A. Steinführer, M. Tominaga, L. Triebe, R. Garcia, N. Gracias, N. Feuillet, C. Deplus. **Quantifying Coseismic Normal Fault Rupture at the Seafloor: The 2004 Les Saintes Earthquake (Mw 6.3) Along the Roseau Fault (French Antilles)**. In AGU, San Francisco, 2014.

- J. Escartín, D. Bonnemains, C. Mevel, M. Cannat, S. Petersen, N. Augustin, A. Bezos, V. Chavagnac, Y. Choi, M. Godard, K. Haaga, C. Hamelin, B. Ildefonse, J. Jamieson,

B. John, T. Leleu, C. MacLeod, **M. Massot**, P. Nomikou, J. Olive, M. Paquet, C. Rommevaux, M. Rothenbeck, A. Steinführer, M. Tominaga, L. Triebe, M. Andreani, R. Garcia, R. Campos. **Insights into the internal structure and formation of striated fault surfaces of oceanic detachments from in situ observations (13°20'N and 13°30'N, Mid-Atlantic Ridge)**. In AGU, San Francisco, 2014.

- **M. Massot**, G. Oliver, L. Ruano, M. Miró. **Texture Analysis of Seabed Images: Quantifying the Presence of Posidonia Oceanica at Palma Bay**. In Proceedings of the IEEE/MTS Oceans Conference, Bergen, Norway, 2013.

Chapter 2

State of the art on underwater 3D sensing methods

In this chapter we review the state of the art of three dimensional reconstruction sensors and techniques, especially visual sensors, applied to underwater mapping and object reconstruction. The text is partially based on [17], a review article from the thesis' author. Most surveys existing in the literature are centred in underwater sensors and imaging techniques, but only few examples can be found focusing on underwater 3D reconstruction and seafloor mapping, which is the aim of this chapter.

The chapter is structured as follows: section 2.2 presents sensing methods to gather 3D data, section 2.3 reviews the different types of hardware sensors and techniques, section 2.4 shows some commercial solutions and finally, in section 2.5 a discussion is held.

The most used sensors and techniques are studied: Lidar, Stereo Vision (SV), Structure from Motion (SfM), Structured Light (SL), Laser Stripe (LS) and Laser Line Scanning (LLS) are described in detail, while sonar is only presented as a reference to be compared with. Features such as range, resolution and ease of assembly are given for underwater conditions.

2.1 Introduction

Jaffe *et al* [27] surveyed in 2001 the different prospects in underwater imaging, foreseeing the introduction of blue-green lasers and multidimensional Photo Multiplier Tubes (PMT) arrays. An application of these prospects is shown in Foley and Mildell [28], who covered in 2002 the technologies for precise archaeological surveys in deep water such as image mosaicking and acoustic three-dimensional bathymetry.

In [29], Kocak *et al* outlined the advances in the field of underwater imaging from 2005 to 2008, basing their work on a previous survey [30]. Caimi *et al* [31] centred their survey in 2008 on underwater imaging as well, and summarized different extended range imaging techniques as well as spatial coherency and multi-dimensional image acquisition. Years later, Bonin *et al* [32] surveyed in 2011 different techniques and methods to build underwater imaging and illuminating systems.

Finally, Bianco *et al* [33] focused in underwater 3D reconstruction on close-range underwater objects in 2013, but only comparing structured light and passive stereo. The same year, Erič *et al* [34] explored 3D reconstruction from the point of view of the documentation of underwater heritage sites. The methods presented there are structure from motion, gathered from divers carrying a camera, and structured light for the modelling of underwater statues and busts. Structure from motion is also compared by Jordt [35], who included in 2014 different surveys on 3D reconstruction, image correction calibration and mosaicking on her PhD memorandum, where she studied structure from motion and stereoscopy.

2.2 Sensing Methods

Three dimensional sensors can be classified in three major classes depending on the measuring method: Triangulation, Time of Flight and Modulation. A sensor can belong to more than one class, which means that it uses different methods or a combination of them to obtain three dimensional data, as depicted in figure 2.1. There is also another traditional classification depending on whether the sensing device is active or passive. All methods in the figure are active except for passive imaging.

These methods and devices will be compared in terms of range, resolution, precision and accuracy when available. The relationship between precision and accuracy is explained in figure 2.2.

2.2.1 Active or Passive

Sensors can also be classified as passive or active depending on whether they interact or not with the medium.

Active sensors are those that either illuminate, project or cast a signal to the environment in order to measure the data to gather. An example of an active system is sonar, where a sonic pulse is sent onto the scene to reconstruct.

Passive methods only get data from the measurable signals in the underwater environment, with no alteration or change on the scene. An example of that is Structure from Motion, where a monocular camera travels looking for features for a posterior 3D triangulation. Camera-based sensors are the only ones that can be passive for 3D reconstruction, as the other are based on sound or on light projection.

2.2.2 Time of Flight

Time discrimination methods are based on controlling the time the signal travels. By knowing the speed of the signal in the medium where it travels, the distance can be inferred. These methods achieve longer distances, especially sonar, but can be affected by changes in water temperature, salinity and pressure, as the speed of sound changes with them.

At shorter distances, a small time delay in the timing can cause a big error in the measurement. Furthermore, some sensors require a minimum distance at which they can measure depending on their geometry.

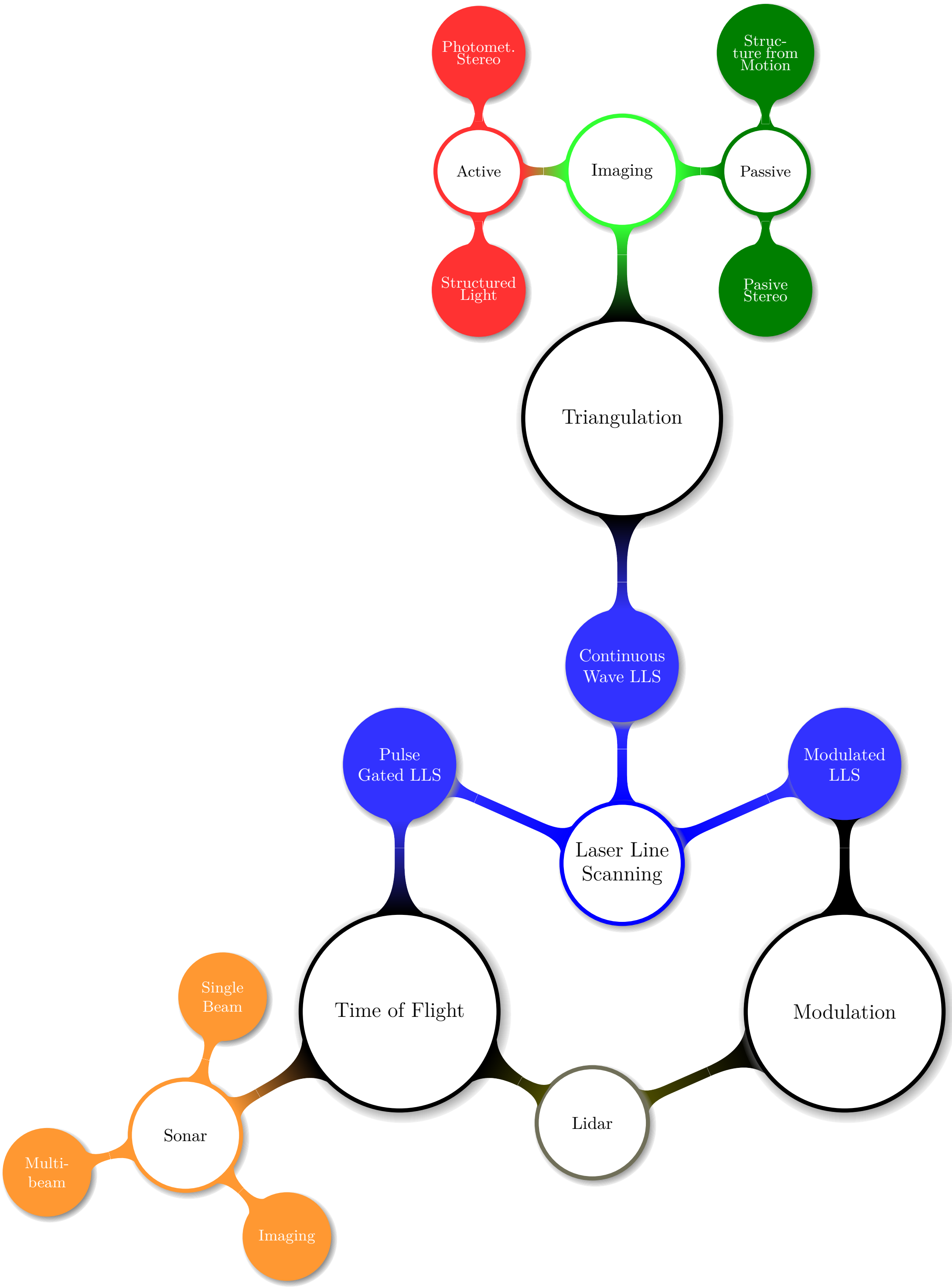


Figure 2.1: 3D reconstruction sensors classification

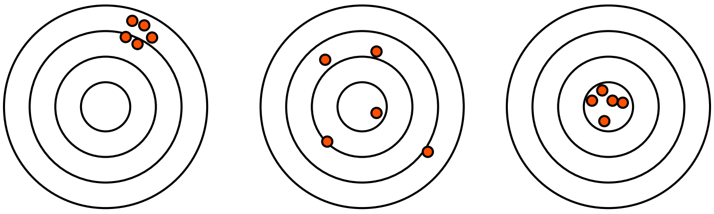


Figure 2.2: Accuracy vs precision: the target on the left has been shot with a precise, non accurate weapon, whereas the target in the centre has been shot with an accurate, non precise weapon. What is desired is an accurate and precise target as shown on the right. Accuracy accounts for the average error to the target and precision for its dispersion.

Sonar, Lidar and Pulse Gated Laser Line Scanning (PG-LLS) are some examples of sensor hardware using this principle to acquire 3D data.

2.2.3 Triangulation

Triangulation methods are based on measuring the distance from two or more devices to a common feature or target with some known parameters.

For example, two cameras can obtain depth (e.g. a stereo rig) by searching on the right camera features found on the left. Once these features have been matched and filtered, the remaining features can be projected on the world as light rays coming from these two cameras. The triangle formed between the feature in the space, and the two cameras is the basis of triangulation. In these methods, artificial lights (lamps and spotlights, for example) are used just to illuminate the scene if needed, and are not employed in the triangulation of the 3D points, which, in turn, is based on the knowledge of similar points in the image sequence, found through stereo matching algorithms.

The limitation of triangulation sensors is their field of view. Triangulation-based devices tend to be better at close distances and worse at far. Also, the bigger is the separation of the cameras (baseline), the better is the z resolution, provided there exists a common view region [36].

Different sensors exist that compute 3D information by triangulation: Structured Light, Laser Stripe and Photometric stereo (PhS) from active imaging, Structure from Motion and Stereo Vision from passive imaging and Continuous Wave Laser Line Scanning (CW-LLS) from Laser Line Scanning.

2.2.4 Modulation

While the time domain approach uses amplitude and time to discriminate multiple scattered, diffused photons, the frequency domain uses the differences in amplitude and phase of a modulated signal to perform this task. The diffused photons that undergo many scattering events produce temporal spreading of the transmitted pulse. Only low frequency components are efficiently transmitted whilst high frequency components are lost.

It is known that coherent modulation/demodulation techniques at optical frequencies in underwater environments fall apart due to the high dispersion in the sea water path [29], as well as for the different absorption and scattering coefficients depending on the optical wavelength. There is a minimum for these coefficients in the blue-green colour spectra. Thus, intensity modulation of a laser carrier is the most used modulation technique in underwater reconstruction. This method has been reported in the literature both from airborne platforms and from underwater vehicles. They usually modulate the amplitude in frequencies in the order of GigaHertz, thus requiring very sensitive sensors and accurate time scales. The receivers are usually Photomultiplier Tubes (PMT) or modern photon counters made of avalanche photodiodes (APD). These sensors are triggered during a time window and incoming light is integrated. After the demodulation step, 3D information can be obtained from phase difference.

2.3 Types of sensors

2.3.1 Sonar

The term sonar is an acronym for sound, navigation and ranging. There are two major kinds of sonars, active and passive.

Passive sonar systems usually have large sonic signature databases. A computer system frequently uses these databases to identify classes of ships, actions (i.e. the speed of a ship, or the type of weapon released), and even particular ships. These sensors are evidently not used for 3D reconstructions, thus they are not considered in this study.

Active sonars create a pulse of sound, often called a “ping”, and then listen for reflections of the pulse. The pulse may be at constant frequency or a chirp of changing frequency. If a chirp, the receiver correlates the frequency of the reflections to the known signal. In general, long-distance active sonars use lower frequencies (1 - 600 kHz) whilst short-distance high-resolution sonars use high frequencies (1 - 3 MHz).

In the active sonar category, we can find three major representative types of sonars: Side Scan Sonar (SSS), Multi-Beam Sonar (MB) and Single Beam sonar (SB). If the across track angle is wide, they are usually called imaging sonars. Otherwise, they are commonly named profiling sonars as they are mainly used to gather bathymetric data. Moreover, these sonars can be mechanically operated to perform a scan, towed or mounted on a vessel or underwater vehicle.

Sound propagates in water faster than in air, although its speed is also influenced by water temperature and salinity. One of the main advantages of sonar soundings is its long range compared to electromagnetic waves, making them a feasible sensor to gather bathymetry data from a surface vessel, even for thousands of meters’ depth. At this distance, a resolution of tenths of meters per sounding is a good result, whilst if an AUV is sent to *fly* at 40 m to perform a survey, a resolution of less than a meter can be achieved.

One of the clearest examples of bathymetric data gathering is performed using MB sonar, as in [37]. This sensor can also be correlated to a colour camera to obtain not only 3D but also colour information, as in [38], where its authors scan a pool using this method. However, in this case, its range lowered to the visual available range.

SB scanning sonars can carry out a 3D swath by slowly rotating their head [39], as if the SB was a 1D range sensor mounted on a pan and tilt head. The data retrieval is not as fast as with a MB.

MB sonars can also be mounted on pan and tilt systems to perform a complete 3D scan. They are usually deployed using a tripod or mounted on an ROV, requiring the ROV to remain static while the sweep is done, like in [40].

Profiling can also be done with SSS, which is normally towed or mounted in an AUV to perform a gridded survey. The SSS is mainly used on-board of a constant speed vehicle describing straight transects. Even though SSS can be considered as a 2D imaging sonar, 3D information can be inferred from it, as depicted in [41].

Imaging sonars differ from MB or SB sonar by a broadened beam angle (e.g. they capture an *image* of the sea bottom instead of a thin profile). For instance, in [42] Brahim *et al* use an imaging sonar with a field of view of $29°$(azimut) $\times$ $10.8°$(elevation) producing 96×512 px azimut by range sonar images.

Other exotic systems have been researched, combining imaging sonar with conventional cameras to enhance the 3D output and to better correlate the sonar correspondences. In [43], Negahdaripour uses a stereo system formed by a camera and a imaging sonar. Correspondences between the two images are described in terms of conic sections. In [44] a forward looking sonar and camera use manually provided correspondences between the sonar image and the camera image to perform reconstructions.

Other solutions prove that imaging sonars can be used to recover depth information. For example, in [45] 3D data is recovered from a set of images using SfM; and in [46] the same information is inferred from the shadow casted by an object sounded by a forward looking sonar.

The basics in sonar or in time of flight methods is depicted by equation (2.1).

$$d = \frac{v\Delta t}{2} \tag{2.1}$$

where d is the distance between the target and the receiver, v is the speed of sound underwater and Δt is the time passed between the pulse and its echo being received. Bearing can be estimated via two means: using an array sensor or by a moving platform. If there is more than one receiver with calibrated relative positioning, the relative bearing of the target can be computed using the time difference of arrival. Alternatively, if there is only one receiver, target bearing can be estimated from two different receiving positions (e.g. if a ship carrying the sonar moves along whilst sending sonar pulses).

Table 2.1 shows a comparison of the 3D reconstruction techniques using sonar. In the references where the resolution has been obtained from graphic plots provided by the authors or similar information, a *less than* (<) symbol has been used.

References	Sonar Type	Scope	Accuracy
Pathak [37]	MB	Rough map for path planning	1 m at 100 m
Rosenblum [47]	MB	Small object reconstruction	8 cm at 1 m
Hurtos [38]	MB + Camera	Projects images on 3D surfaces	2.3 cm at 1.5 m
Guo [39]	SB	Small target 3D reconstruction	2.6 cm at 1 m
Coiras [41]	SSS	Seabed elevation with UW pipe	13 ± 19 cm at 124m
Brahim [42]	IS	Sparse scene geometry	0.5 m at 10 m
Aykin [46]	IS	Smooth surfaces 3D reconstruction	-
Negahdaripour [43–45]	IS + Camera	Alternative to optical stereo systems	< 5 cm

Table 2.1: Sumary of Sonar 3D reconstruction solutions. Note that not all authors provide measures of accuracy. The annotated values have been taken from their publication.

2.3.2 Lidar

Lidar, an acronym for Airborne scanning Light Detection And Ranging, is widely used as a mapping tool for coastal and near shore ocean surveys. Similar to LLS, but surveyed from an aircraft, a laser line is scanned throughout the landscape and the ocean. Depending on the laser wavelength, Lidar is capable of recovering both the ocean surface and the sea bottom. In this particular case, a green 532 nm laser that penetrates the ocean water over 30 m [48] is used in combination with a red or infrared laser. Both lasers return the echo from the sea surface, but only one reaches the underwater domain.

Lidar has been used for Underwater Target Detection (UWTD), usually mines, as well as for coastal bathymetry [49,50]. It is normally surveyed at heights of hundreds of meters (Pellen *et al* survey mostly uniformly at 300 m [50]) with a swath of 100 to 250 m. Its accuracy is in the order of decimeters. In [48] an accuracy of 0.7 m is achieved. Moreover, Lidar signal can be modulated, enhancing its range capabilities and rejecting underwater backscatter [51, 52].

Although this chapter focuses on underwater sensors, Lidars have been shortly mentioned as they are capable of reconstructing certain coastal regions from the air.

Lidar, as sonar, uses a TOF equation similar to (2.1), and it derives into (2.2),

$$d = \frac{c\Delta t}{2} \tag{2.2}$$

where d is the distance between the target and the receiver, c is the speed of sound in air and Δt is the time passed between the pulse and its echo being received. Bearing is estimated from the scanning angle.

In table 2.2 two 3D reconstruction references using Lidar are compared.

References	Class	Wavelength	Lidar Model	Combination	Accuracy
Reineman [48]	ToF	905 nm	Riegl LMS-Q240i	Camera, GPS	8.7 cm at 120m
Cadalli [49]	ToF	532 nm	U.S. Navy prototype	PMT + 64x64 CCD	< 10 m
Pellen [50]	UWTD	532 nm	ND:YAG laser	PMT	-
Mullen [51, 52]	UWTD	532 nm	Nd:YAG laser	PMT + Microwave	-

Table 2.2: Summary of Lidar 3D reconstruction solutions. Note that not all authors provide measures of accuracy. The annotated values have been taken from their publication.

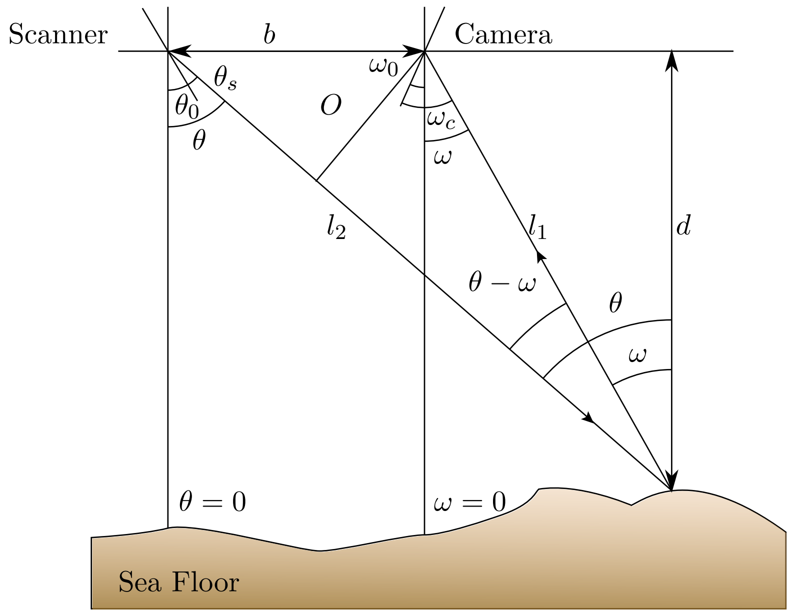


Figure 2.3: Triangulation geometry principle for a Laser Scanning System

2.3.3 Laser Line Scanning (LLS)

For finer details in 3D reconstructions, laser and imaging devices are mostly used. However, imaging is limited by the absorption and scattering of the water. Customers of the oil and gas companies often do not want the risk of being 1-2 meters from an expensive underwater asset in order to perform inspections. LLS achieves larger ranges exploiting range gated receivers and narrow laser pulses, as a matter of example, to get rid of backscatter.

This kind of sensors tend to use Photomultiplier Tubes (PMT) as photon counters. Some modern approaches use photodiodes or even cameras with a fine-tuned mechanical shutter. As laser source, green lasers working at 532 *nm* are a common solution, as this wavelength presents a trade-off between good price, availability and low absorption and scattering coefficients for pure seawater.

There are three main categories of LLS: Continuous Wave LLS (CW-LLS), Pulse Gated LLS (PG-LLS) and Modulated LLS (Mod-LLS). In table 2.3 there is a summary of the different LLS 3D reconstruction solutions.

References	Aim	Type	Wavelength	Receiver	Resolution	Accuracy	Precision
Moore [53]	3D	CW-LLS	532 nm	Linescan CCD	90 mm at 10 m	-	-
Moore [54]	3D	CW-LLS	532 nm	Linescan CCD	-	2.5 mm at 2.87m	0.3 mm std
McLeod [55]	3D	PG-LLS	-	-	1 mm	-	5 mm at 8 m
Imaki [56]	3D	PG-LLS	532 nm	PMT	-	-	2 cm std at 20 m
Cochenour [57]	3D	Mod-LLS	532 nm	PMT	-	-	-
Rumbaugh [58]	3D	Mod-LLS	532 nm	APD	-	1 cm at 50 cm	-
Dominicis [59]	3D	Mod-LLS	405 nm	PMT	-	5 mm at 8.5 m	-
Dalgleish [1]	Img.	CW-LLS	532 nm	PMT	50 mm at 10 m	-	-
Mullen [2]	Img.	Mod-LLS	532 nm	PMT	-	-	0.1 m at 3.7 m
Gordon [60]	Img.	PG-LLS	488-514.5 nm	PMT	-	-	-

Table 2.3: Sumary of Laser Line Scanning 3D reconstruction solutions. Note that not all authors provide measures of resolution, accuracy or precision. The annotated values have been taken from their publication.

2.3.3.1 Continuous Wave LLS

This subcategory uses a triangulation method to recover the unknown depth. A camera-based triangulation device using a laser scan concept can be built using a moving laser pointer made of a mirror galvanometer and a line-scan camera, as shown in [53, 54].

The geometric relationship between the camera, the laser scanner and the illuminated target spot is shown in figure 2.3. The depth D of a target can be calculated from (2.3).

$$d = l_1 \cos(\omega) \tag{2.3}$$

as

$$l_1 = \frac{b\cos(\theta)}{\sin(\theta-\omega)} \tag{2.4}$$

since

$$\sin(\theta-\omega) = \frac{O}{l_1}, \text{ and } O = b\cos(\theta) \tag{2.5}$$

therefore

$$d = \frac{b}{\tan(\theta)-\tan(\omega)} \tag{2.6}$$

where b is the baseline or separation between the centre of the scanning mirror and the centre of the primary receiving lens of the camera. Here, θ and ω are the scanning and camera pixel viewing angles, respectively.

The angles ω_0 and θ_0 are the offset mounting angles of the scanner and camera, and θ_s and ω_c are the laser beam angle known from a galvanometer or an encoder and the pixel viewing angle (with respect to the camera housing). Thus,

$$\theta = \theta_0 + \theta_s \tag{2.7}$$

$$\omega = \omega_0 + \omega_c \tag{2.8}$$

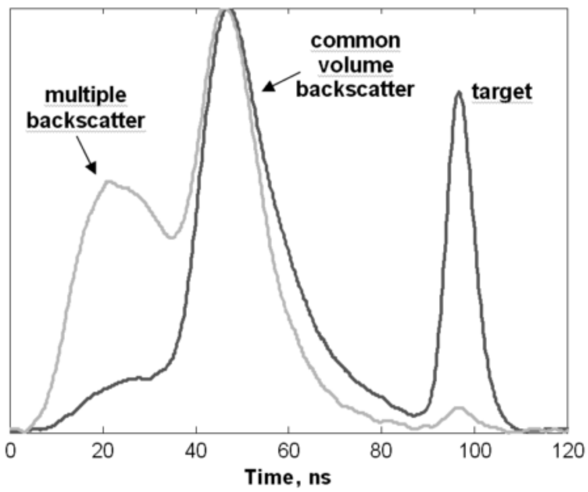


Figure 2.4: Example of a returning signal from a Laser Scanning System. At higher turbidity (gray signal) the backscatter peak is stronger and the target return is weaker. The common volume backscatter is light that has been deflected once, whilst the multiple backscatter has been deflected twice or more times. Reproduced from [1].

$$d = \frac{s}{\tan(\theta_0 + \theta_s) - \tan(\omega_0 + \omega_c)} \tag{2.9}$$

Both θ_0 and ω_0 have to be computed by calibration so that afterwards, the distance to the target can be computed.

2.3.3.2 Pulse Gated LLS

This ToF sensor has a simple principle: it illuminates a narrow area with a laser light pulse while keeping the receivers shutter closed. Then, it waits for the return of the light from the object by estimating its distance from the sensor and then opens the shutter so that only the light returning from the target is captured. For instance, in figure 2.4 the shutter should have been opened from 80 to 120 *ns* to get rid of the unwanted backscatter.

This setup has been highly used in extended range imagery. In the early 90s the LLS system in [60] was used in USS Dolphin research submarine and as a towed body to perform high resolution imagery at an extended range. This prototype used Argon Ion gas laser, with a high power budget not available for ROVs nor AUVs.

Dalgleish *et al* [1] compared PG-LLS with CW-LLS as imaging sensors. The experimental results demonstrate that the PG imager improved contrast and SNR (Signal to Noise Ratio). Their sensor becomes limited by forward backscatter at 7 attenuation lengths, whilst CW at 6.

In true ToF 3D reconstruction, McLeod *et al* [55] published a paper about a commercial sensor [61] mounted on Marlin AUV. Their setup achieves a precision of 5 *mm* in a good

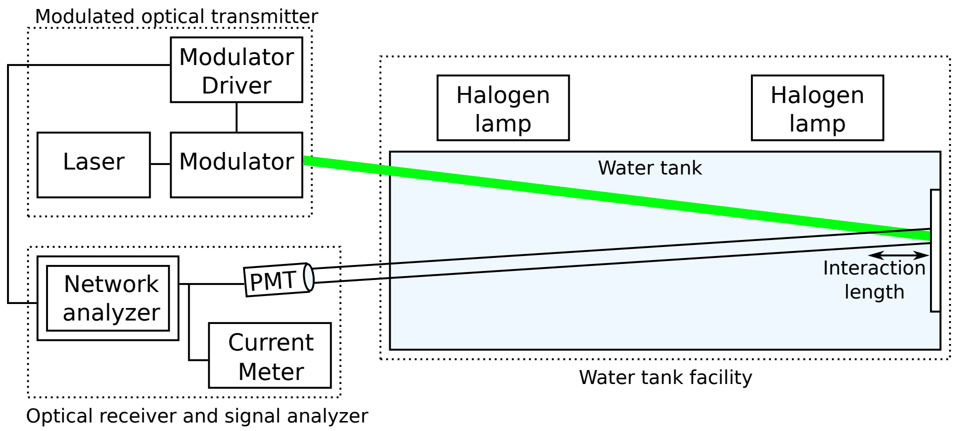


Figure 2.5: Laser line scanning setup including a modulated optical transmitter, an optical receiver and signal analyzer, and a water tank facility. The interaction length is the distance over which the transmitted beam and the receiver field of view overlap. Reproduced from [2].

visibility scenario, when measuring a point at 8 m. Similarly but in a longer range, Imaki [56] *et al* achieve a precision of 2 cm at a range of 20 m with their own developed system.

2.3.3.3 Modulated LLS

A Modulated LLS characterizes for the use of the frequency domain instead of the spatial or time domain to discern a change in the sent signal. In Sonar chirps (Radar as well) the modulation and posterior de-modulation of the signal gives insight on the distance from the sensor to the target.

As stated before, intensity modulation is the only realizable modulation in underwater scenarios. The original and the returned signal are subtracted and the distance is obtained by demodulation of the remainder.

The same approach can be used for extended range imaging as well, as seen in [2], where Mullen *et al* have developed a Modulated LLS that uses frequency modulation and demodulation in the laser source in order to identify the distance at which the target has been illuminated. The optical modulation is used to discriminate scattered light. There, they compared different frequencies and found that a higher frequency, 90 MHz reached further than 50 MHz or 10 MHz. The setup used by the authors can be seen in figure 2.5.

In [57] different modulation techniques based on ST-MP (Single Tone Modulated Pulse) and PN-MP (Pseudorandom coded Modulated Pulse) are compared for one dimensional ranging. The results show that in clear water, the PN-MP stands as an improvement over the ST-MP due to their excellent correlation properties of pseudorandom codes.

In [58] a one axis ranging solution is proposed. Although the authors characterize the solution as Lidar, their setup is more similar to LLS, and the measurements are not taken from a plane. In the paper, a resolution of 1 cm from a distance of 60 cm is reported, thus

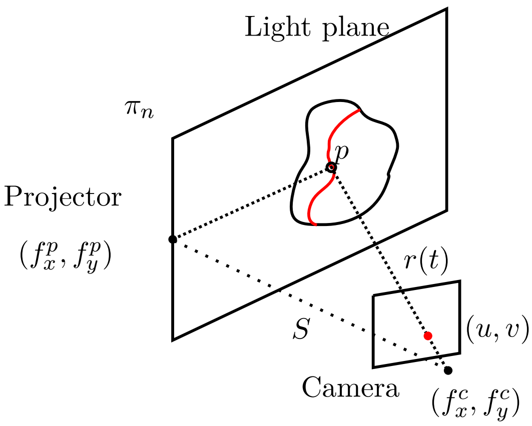


Figure 2.6: Triangulation geometry principle for a Structured Light System

1.7% of error. This system could then be swept for a 3D reconstruction and work as a true LLS.

In [59], a simpler approach using an amplitude modulated blue laser (405 nm) at 80 MHz was used, called MODEM-based 3D laser scanning system. It is capable of reconstructing objects at 8.5 meters away within a 5% accuracy. The system is similar to the ones presented before, but in this case 3D reconstructions are presented to the reader, showing the potential of long range underwater reconstruction using this technique.

2.3.4 Variable Wavelength Structured Light (VW-SL)

These systems consist of a camera and a colour, or white light projector. The triangulation principle is used between these two elements and the projected object.

The projector casts a known pattern on the scene, normally a set of light planes, as shown in figure 2.6, where both the planes and the camera ray are known. The intersection between both is unknown and can be calculated as follows:

Mathematically, a line can be represented in parametric form as:

$$r(t) = \begin{cases} x & = & \frac{u-c_x}{f_x}\, t \\ y & = & \frac{v-c_y}{f_y}\, t \\ z & = & t \end{cases} \tag{2.10}$$

where (f_x, f_y) is the camera focal length in x and y axes, (c_x, c_y) is the central pixel in the image and (u, v) is one of the detected pixels in the image. Supposing a calibrated camera and the origin in the camera frame the light plane can be represented as in (2.11).

$$\pi_n: \ Ax + By + Cz + D = 0 \tag{2.11}$$

To find the intersection point, (2.10) is substituted in (2.11), giving (2.12).

$$t = \frac{-D}{A\frac{u-c_x}{f_x} + B\frac{v-c_y}{f_y} + C} \tag{2.12}$$

Different projecting patterns have been used in the literature [62]. Binary patterns are the most used, although multilevel ones have also been proposed. The former option is simpler in terms of processing and easier to achieve with a projector. Binary patterns use only two states of light, usually white light, illuminating stripes in the scene. At the beginning, there is only one division (black-to-white) in the image. In the following scans, subdivisions of the previous areas of the scene are covered until the camera cannot differentiate the two states. The correspondence of consecutive light planes is solved using time multiplexing. The number of light planes achievable with this method is fixed, normally by the resolution of the projector. On the other hand, phase shifting patterns use sinusoidal projections in the same operating mode to cover wider values in gray scale. By unraveling the phase value, different light planes can be obtained for just one state in the equivalent binary pattern. Phase shifting patterns are also time multiplexing patterns.

These methods use more than one projection pattern to obtain range information. De Bruijn sequences can achieve one-shot reconstructions by using pseudo-random sequences formed by alphabets of symbols in a circular string. If this theory is brought to matrices instead of vectors (e.g. strings), then those patterns are called M-arrays. These can be constructed by folding a pseudo-random sequence [63]. Usually these patters use color to better distinguish the symbols in the alphabet. However, not all kind of surface finishes and colors reflect correctly the incoming color spectra back to the camera. Direct coding patterns, either in gray levels or in color, have also been used in air. However, to the best knowledge of the authors, there are no reports in underwater scenarios of some of the previously explained codification strategies.

In the literature, Zhang *et al* project a gray scale four-step sinusoidal fringe [64]. Therefore, it is a time multiplexing method using four different patterns. In their article, SL is compared to SV showing better behavior in SL on textureless objects. Same reports were obtained projecting 20 different gray coded patterns in a pool [3]. An accuracy in z direction of 2% was achieved with this system.

Bruno *et al* [65] also project gray coded patterns with a final code shift of 4 pixel wide bands. With this last shifts, better accuracy can be obtained compared to narrowing the pattern to only one pixel wide patterns, where finding all the thin black and white lines is more difficult. In this setup, a total of 48 patterns were used. However, this particular setup calculates the 3D points using the positions of two cameras determined during the calibration phase. The projector is only used to establish the correspondences and it is not involved in the triangulation. This system would be an hybrid between SL and SV.

Another way to triangulate information using structured light is to sweep a light plane.

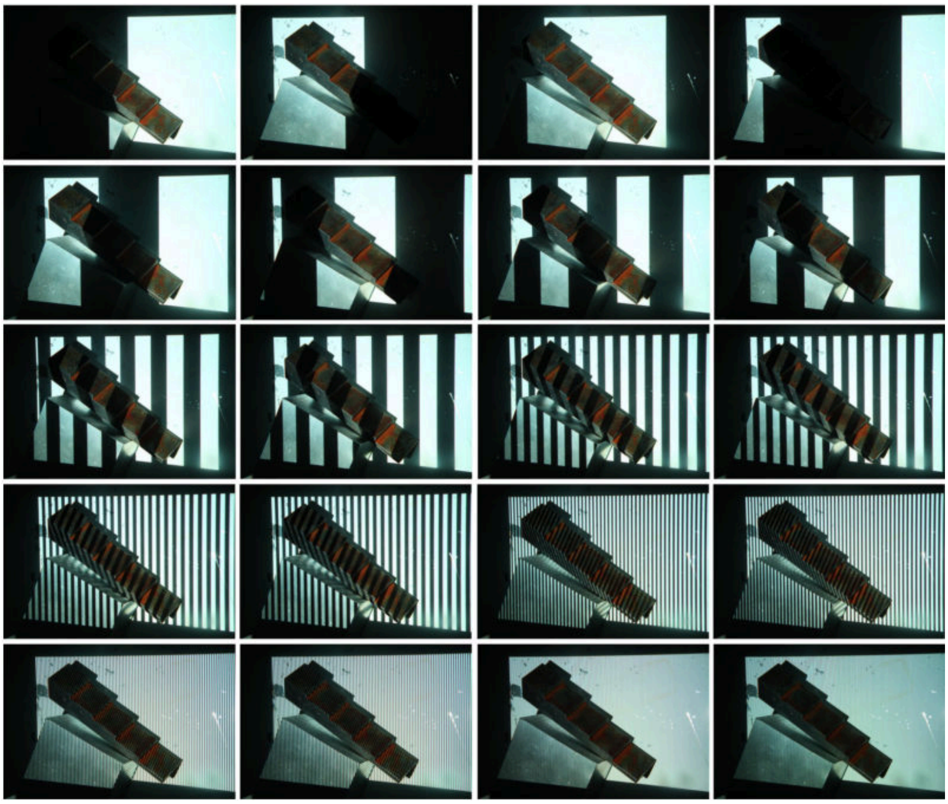


Figure 2.7: A set of captured pictures captured using the projected Gray code bit planes. Extracted from [3].

This light plane can be swept either using the available pixels in the projector of by moving the projector. Narasimhan and Nayar [66] sweep a light plane into a tank with dilute milk and recover 3D information even in high turbidity scenarios. The authors show that in high turbidity medium and with conventional floodlight, it is impossible to see anything but backscattering. By narrowing the illuminated area to a light plane, the shapes of the objects in the distance can be picked out and therefore triangulated.

2.3.5 Fixed Wavelength Structured Light Systems (FW-SL)

The systems presented in this section project fixed wavelength light into the environment. This light is normally obtained from lasers.

2.3.5.1 Laser Stripe (LS)

LS systems are a subgroup of SL systems, where although the pattern is fixed to be a line (a laser plane), the projector is swept across the field of view of the camera. Thus, for this setting, a motorized element is needed in addition to the laser if the system holding the camera and laser is not moving. The relative position and orientation of the laser and camera system must be known in order to perform the triangulation process. The resolution of these systems is usually higher than stereoscopy, but they are still limited by absorption and scattering. The range of LS does not normally go over 3 m in clear waters, as will be seen later in the commercial solutions.

Using an underwater stripe scanning system was initially proposed by Jaffe and Dunn in [67] to reduce backscattering. Tetlow and Spours [68] show in their article a laser stripe system with an automatic threshold setup for the camera, making this sensor robust to pixel saturation if the laser reflection is too strong. To do that, they programmed a table with the calibrated thickness of the laser stripe depending on the distance to the target. In their results, they achieved resolutions up to five millimetres at a distance of three meters.

Kondo *et al* [69] tested a LS system in *Tri-Dog I* AUV. Apart from using it for 3D reconstructions, they also track the image online to govern the robot to keep a safe distance to the seabed centering the laser line in the camera image by changing the depth of the vehicle. They report a resolution of 40 mm at three meters.

Hildebrandt *et al* [70] mount a laser line onto a servomotor that can be rotated 45° with an accuracy of 0.15°. The camera is a 640×480 CMOS 200 fps with a 90° FOV (Field of View). Returns $300k$ points in 2.4 seconds. Calibration is shown in his article with a novel calibration rig consisting of a standard calibration pattern (checkerboard) next to a grey surface on one side. The laser is better detected on a grey surface because on white, light is strongly reflected and the camera capturing the scene has to compensate it opening the shutter for a shorter period of time. The detection of the laser in the same plane of the calibration pattern is used to calculate the position of the laser sheet projector with respect to the camera.

In [71] a system consisting on a camera, a laser line and a LED light are mounted on the AUV *Tuna Sand* to gather 3D information as well as imagery. The laser is pointed at the upper part of the image whilst the LEDs light the lower part, so that there is enough contrast to detect the laser line. In [72–74] a similar system called SeaXerocks 3D mapping device is mounted on the ROV *Hyper-Dolphin*. With this system, the authors perform 3D reconstructions in real intervention scenarios such as in hydrothermal sites and shipwrecks.

In [75] *Tuna Sand* AUV is also used with a different sensor. In this case, a laser stripe is mounted on a rotary motor and, in a different enclosure, a camera is mounted. By keeping the robot as static as possible, the laser is projected onto the scene whilst rotating it. The camera then captures the line deformation. In this paper, multiple laser scans from sea experiments at Kagoshima Bay are combined using Iterative Closest Point (ICP) algorithm. The authors reconstruct a hydrothermal vent whose chimney is 3 meters tall at a depth of 200 meters.

In [59,76] a dual laser scanner is used to increase the field of view of the laser stripe, so that almost 180° are illuminated by the laser. The system is very similar to the commercial sensor in [77]. They approximate the detected laser lines to be Gaussian and explain an optimization method to calibrate the camera-to-laser transformation. The authors claim that the achieved measuring error is below 4%.

Prats *et al* [78–80] use a laser stripe on a underwater manipulator (which is mounted onto the AUV *Girona 500* in some experiments). They take advantage of the arm motion to sweep the stripe. The resulting point cloud is used to autonomously grasp known objects. Although the robot slightly changes its position when the arm moves to scan the scene, they track the sea bottom with a template tracking algorithm to estimate the robot motion, thus correcting small misalignments between the data and the real environment.

Different approaches to the common laser stripe scanning have been also reported. In [81] two almost-parallel laser stripes are projected to compute the distance between these lines captured from a camera, to know the distance to the target. These values are used as an underwater rangefinder. However 3D reconstruction was not the aim of the paper.

In [82], Caccia mounts four laser pointers lined with a camera in a ROV. The four imaged pointers are used to calculate the altitude and the heading of the vehicle, assuming the seabed is flat.

Yang *et al* mount a camera and a vertical laser stripe in a translation stage [83]. They recover 3D data interpolating from a data table previously acquired from calibration. Whenever a laser pixel is detected in the image, its depth value is calculated from the four closest points in the calibration data.

In table 2.4, the different SL references are compared. For the solutions with no clear accuracy results, the resolution has been deduced from the graphics in their respective articles.

References	Type	Color / Wavelength	Pattern	Accuracy
Zhang [64]	SL	Grayscale	Sinusoidal Fringe	-
Tornblom [3]	SL	White	Binary pattern	2 mm at 250 mm
Bruno [65]	SL	White	Binary pattern	0.4 mm at 1.2 m
Narasimhan [66]	SL	White	Light plane sweep	5 mm at 100 mm
Bodenmann [73, 84]	LS	532 nm	Laser line	200 mm at 10 m
Yang [83]	LS	532 nm	Laser line	-
Kondo [69]	LS	532 nm	Laser line	-
Tetlow [68]	Mot. LS	532 nm	Laser line	30 mm at 3 m
Hildebrandt [70]	Mot. LS	532 nm	Laser line	-
Prats [78]	Mot. LS	532 nm	Laser line	-
Nakatani [75]	Mot. LS	532 nm	Laser line	1 cm at 2 m
Jakas [59, 76]	Dual LS	405 nm	Laser line	See [77]
Massot [22]	LbSLS	532 nm	25 laser lines	3.5 mm at 70 cm

Table 2.4: Summary of Structured Light 3D reconstruction solutions. *Massot* method will be presented in chapter 3. Note that not all authors provide measures of accuracy. The annotated values have been taken from their publication.

2.3.6 Photometric stereo

In situations where light stripe scanning takes too long to be practical, photometric stereo provides an attractive alternative. This technique for scene reconstruction requires a small number of images captured under different lightning conditions. In figure 2.8 there is a representation of a typical PhS setup with four lights.

3D information can be obtained by changing the location of the light source whilst keeping the camera and the object in a fixed position. Narasimhan and Nayar present a novel method to recover albedo, normals and depth maps from scattering media [66]. Usually, this method requires a minimum of 5 images. In special conditions such as the ones presented in [66], four different light conditions can be enough.

In [85], Tsiotsios *et al* show that three lights are enough to compute tridimensional information. They also compensate the backscatter component by fitting a backscatter model for each pixel.

2.3.7 Structure from Motion (SfM)

SfM consists in travelling in space taking camera shots. From these camera shots, image features are detected and matched between consecutive frames to know the relative camera motion, and thus its 3D trajectory.

Given m images of n fixed 3D points, we need to estimate m projection matrices $\boldsymbol{P}_i$ and n 3D points $\boldsymbol{X}_j$ from the $m \cdot n$ correspondences $\boldsymbol{x}_{ij}$.

$$\boldsymbol{x}_{ij} = \boldsymbol{P}_i \boldsymbol{X}_j,\ i = 1, \dots, m,\ j = 1, \dots, n \tag{2.13}$$

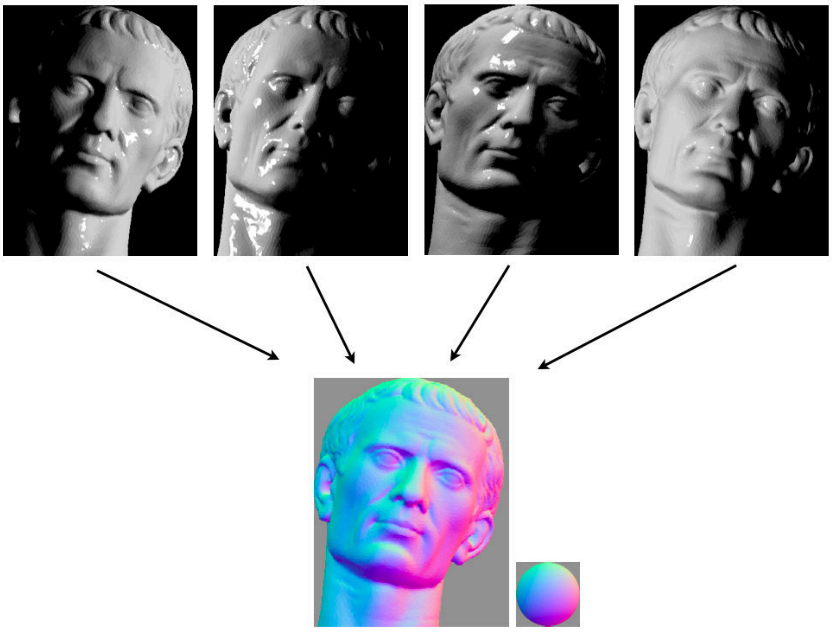


Figure 2.8: Photometric stereo setup: four lights are used to illuminate an underwater scene. The same scene lid from different sources are the images used to recover three-dimensional information. Reproduced from: `http://perception.csl.illinois.edu/matrix-rank/stereo.html`

Therefore, if we scale the entire scene by some factor k and, at the same time, scale the projection matrices by a factor of $1/k$, the projection of the scene points remain the same. Thus, only with SfM, scale is not available, although there are methods that compute the scale from known objects or by knowing the constraints of the robot or vehicle carrying the camera [86].

$$\boldsymbol{x} = \boldsymbol{P}\boldsymbol{X} = \left(\frac{1}{k}\boldsymbol{P}\right)(k\boldsymbol{X}) \tag{2.14}$$

The one-parameter family of solutions parametrized by λ is

$$\boldsymbol{X}(\lambda) = \boldsymbol{P}^{+}\boldsymbol{x} + \lambda\boldsymbol{c} \tag{2.15}$$

where $\boldsymbol{P}^{+}$ is the pseudo-inverse of $\boldsymbol{P}$ (i.e. $\boldsymbol{P}\boldsymbol{P}^{+} = I$) and $\boldsymbol{c}$ its null-vector, namely the camera centre, defined by $\boldsymbol{P}\boldsymbol{c} = 0$.

The approach of SfM is the cheapest in terms of hardware, and the easiest to install in a real robot. Only a still camera or a video recorder is needed, with enough storage to keep a full dive in memory. Later, the images can be processed to obtain the required 3D models.

In the underwater medium, both feature detection and matching suffer from diffusion (absorption and scattering), sun flickering and non uniform light, making the detection of the same feature more difficult from different viewpoints. Depending on the distance from the camera to the 3D point, the absorption and scattering components vary, changing the colours and the sharpness of that particular feature in the image.

RANdom SAmple Consensus (RANSAC) is an iterative method to discard outliers widely used in imaging. Sedlazeck *et al* show in [87] a real 3D scenario reconstructed from ROV Kiel 6000 using an HD color camera. The feature detector is a corner detector based on image gradients. RANSAC procedure is used to filter outliers after the features have been matched

Pizarro *et al* [88] use SeaBED AUV to perform optical surveys, equipped with a 1280×1024 *px* CCD camera. The feature detector used is a modified Harris corner detector and its descriptor is a generalized colour moment.

In [89] Meline *et al* compare Harris and SIFT features using a 1280×720 *px* camera in shallow water. In the article, the authors reconstruct a statue bust. They conclude that SIFT is not robust to speckle noise, contrary to Harris. Furthermore, Harris presented better inlier count on the different scenarios.

McKinnon *et al* [90] use GPU SURF features and a high resolution camera 2272×1704 *px* to reconstruct a piece of coral. This setup presents several challenges in terms of occlusions of the different views. With their SfM approach, they achieve 0.3 *mm* accuracy at $1 - 1.5$ *m*.

Jordt-Sedlazeck and Koch develop a novel refractive structure from motion algorithm that takes into account the refraction of glass ports in water [91]. By considering the refraction coefficient between the air-glass-water interface they improve the results of their SfM, called Refractive SfM.

Cocito *et al* [92] use images captured by divers that always contain a scaling cube to recover scaled 3D data. The processing pipeline requires an operator to outline silhouettes of the area of interest of the images. In the case of the application in that paper, they were measuring bryozoan colonies volume.

In [4], the documentation of an archaeological site where experimental cleaning operations were conducted is shown. A commercial software, Photoscan by Agisoft, was used to perform a multi-view 3D reconstruction.

References	Feature	Matching method	Accuracy	Scale
Sedlazeck [87]	Corner	KTL Tracker	-	Medium
Pizarro [88]	Harris	Affine invariant region	3.6 cm RMS	Large
Meline [89]	Harris	SIFT	-	Small
McKinnon [90]	SURF	SURF	1 mm at 1 m	Small
Jordt-Sedlazeck [91]	-	KLT Tracker	-	Small
Cocito [92]	Silhouettes	Manually	< 1 cm	Small
Bruno [4]	SIFT	SIFT	4.5 mm	Small

Table 2.5: Summary of Structure from Motion 3D reconstruction solutions. Note that not all authors provide measures of accuracy. The annotated values have been taken from their publication.

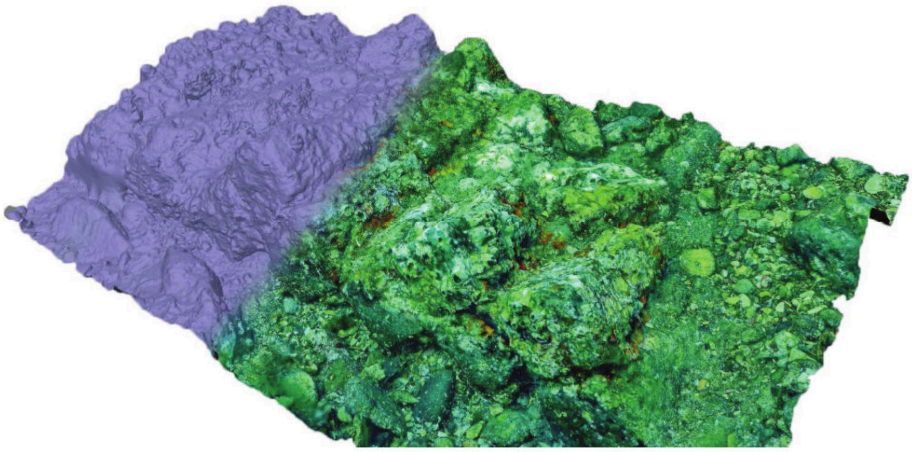


Figure 2.9: 3D reconstruction using SfM. The images were recorded by divers in an archaelogical site. From [4].

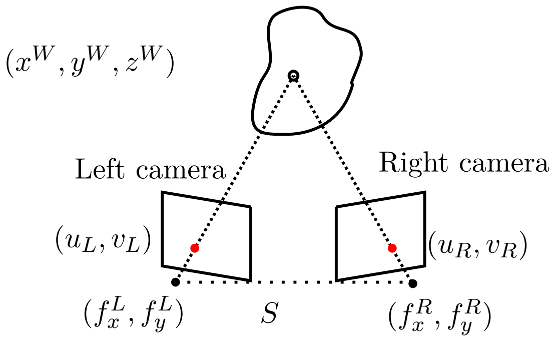


Figure 2.10: Triangulation geometry principle for a Stereo System

The solutions presented are summarized in table 2.5. Most of them do not have results on resolution given that this 3D reconstruction method cannot recover the correct scale. In the solutions where a result is given, the authors have scaled manually the resulting point cloud to match a particular feature or human made object.

2.3.8 Stereo Vision (SV)

Stereoscopy follows the same working principle as SfM, but features are matched between left and right frames of a stereo camera to compute 3D correspondences. From calibration, the relative position of the left camera with respect to the right one is known and, therefore, there is no longer a degree of freedom in the scale of the reconstruction.

By knowing the relative position of the cameras and the location of the same feature in both images, the 3D coordinates of the feature in the world can be computed by triangulation. The closest crossing point (if the lines do not intersect due to calibration) is considered as the corresponding 3D point. In figure 2.10, the corresponding 3D point of the image coordinates

$x = (u_L, v_L)$ and $x' = (u_R, v_R)$ is the point $p = (x^W, y^W, z^W)$, which can also be written as $x'\boldsymbol{F}x = 0$ where $\boldsymbol{F}$ is the fundamental matrix [93].

Once the camera rig is calibrated (known baseline and no distortion in the images) 3D can be obtained calculating the disparity for each pixel e.g. perform a 1D search for each pixel in the left and right images. A block matching is normally used. The disparity is the difference in pixels from left to right, where the same patch has been found. It gives a direct measure of depth as

$$z = \frac{f \cdot b}{d} \tag{2.16}$$

where d is the disparity in pixels, f is the focal distance in pixels, b is the baseline in meters and z is the depth or distance of the pixel perpendicularly to the image plane, in meters.

Once this 3D data has been gathered, the registration between consecutive frames can be done using 2D or 3D features, or even 3D registration methods such as Iterative Closest Point (ICP).

Fairly different feature descriptors and matchers have been used in the literature. SIFT [94–99] is one of the most used. SURF [100], or even direct 3D registration with SIFT 3D [95] or ICP [94]. For instance in [101], Servos *et al* perform refractive projection correction on depth images generated from a Bumblebee2 camera (12 *cm* baseline). The results obtained with this correction have better accuracy and bigger number of pixel correspondences compared to standard methods. The registration is directly done in the generated point cloud using ICP. There is no feature matching.

Schmidt *et al* [97] use commercial GoPro cameras to set a 35 *mm* baseline stereo rig and perform micro bathymetry using SIFT features. They achieve an accuracy of 3 *mm* in their reconstructions.

In [99], the stereo system IRIS is hung from the tip of the arm of *Victor6000* ROV. Their system uses SIFT combined with RANSAC to discard outliers. After that, a sparse bundle adjustment is performed to correct the navigation to survey natural underwater objects.

In [102], Hogue *et al* combine a Bumblebee stereo and a inertial unit housed in a watertight case, called *Aquasensor*. This system is used to reconstruct and register dense stereo 3D points. The reconstruction show a lot of drift if the IMU is not used and presume erroneous camera model to cause part of it. The system is used by the authors to perform a reconstruction of a sunken barge.

Beall *et al* [100] use a wide baseline stereo rig and extract SURF features from left and right image pairs. They track these features to recover the structure of the environment after a SAM (Smoothing and Mapping) step. Then the 3D points are triangulated using Delaunay triangulation and the image texture is mapped to the mesh. This setup is applied to reconstruct coral reefs in Bahamas.

Negre *et al* [5, 103] perform 3D reconstruction of underwater environments using a Graph SLAM approach in a micro AUV equipped with two stereo rigs.

Nurtantio *et al* [96] use three cameras and extract SIFT features. The reconstruction of

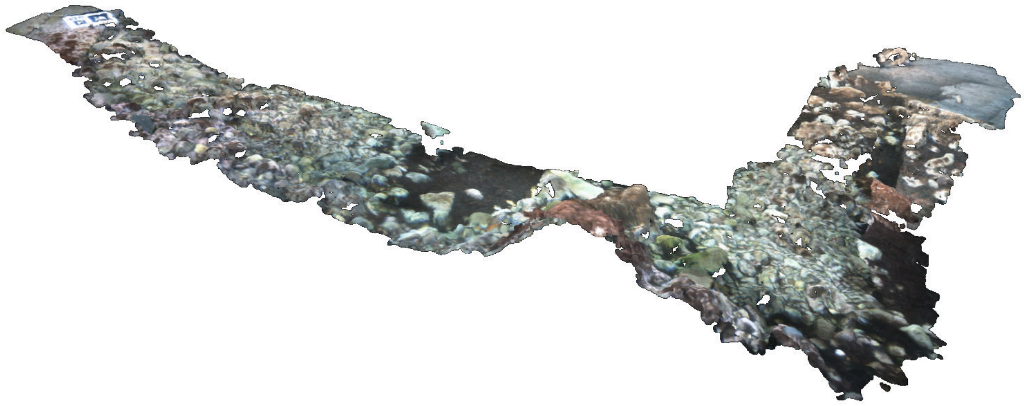


Figure 2.11: 3D reconstruction from SV using Graph SLAM. From [5].

the multi-view system is triangulated using Delaunay triangulation. However they manually preprocess the images to select whether they are suitable for an accurate reconstruction. The outlier removal stage is also manual.

Inglis and Roman constrain stereo correspondences using multibeam sonar [104]. From *Hercules* ROV, navigation data, multibeam and stereo is preprocessed to reduce error and then the sonar and optical data are mapped into a common coordinate system. They backproject the range data coming from the sonar to the camera image and limit the available z range the stereo correspondence algorithm has. To simplify this approach, they tile the sonar backprojections into the image, and generate a tiled minimum and maximum disparity values for an image region (e.g. a tile). The number of inliers obtained with this setup increases significantly compared to an unconstrained system.

References	Feature	Matching method	Baseline	Resolution	Scale
Kumar [94]	SIFT	RANSAC and ICP	-	-	Small
Jasiobedzki [95]	SIFT	SIFT3D and SLAM	-	-	Large
Nurtantio [96]	SIFT	SIFT	-	-	Small
Schmidt [97]	SIFT	SIFT	35 mm	3 mm	Small
Brandou [99]	SIFT	SIFT	-	-	Medium
Beall [100]	SURF	SURF and SAM	60 cm	-	Large
Servos [101]	-	ICP	12 cm	< 27 cm	Small
Hogue [102]	Corners	KLT tracker	12 cm	-	Large
Inglis [104]	SIFT	SIFT	42.5 cm	-	Small

Table 2.6: Sumary of Stereoscopy 3D reconstruction solutions. Note that none of the authors presented results of accuracy nor precision. Resolution has been used instead when available.

In table 2.6, the different solutions are presented and compared.

Comercial Solutions Name	Company	Range (m) Min	Max	Depth (m)	Resolution (mm)	Field of view	Motorized	Method
Echoscope [105]	Coda Octopus	0.5	120	250	20	$50° \times 50°$	yes	TOF
INSCAN [107]	Teledyne CDL	2	25	3000	5	$30 \times 30 \times 360°$	yes	TOF
SL1 [61]	3D at Depth	2	30	3000	4	$30 \times 30 \times 360°$	yes	TOF
3DLS [77]	Smart Light Devices	0.3	2	4000	0.1	t.b.a.	t.b.a.	Triangulation
ULS-100 [109]	2g Robotics	0.1	1	350	1	$50 \times 360°$	yes	Triangulation
ULS-200 [110]		0.25	2.5	350	1	$50 \times 360°$	yes	Triangulation
ULS-500 [111]		1	10	3000	3	$50 \times 360°$	yes	Triangulation
Cerberus	Savante	-	10	6000	-	t.b.a.	t.b.a.	Triangulation
SLV-50	Savante	-	2.5	6000	1	60°	no	Triangulation
Lumeneye	Savante	-	-	6500	-	65°	no	Laser only
SeaStripe [112]	Tritech	-	-	4000	-	64°	no	Laser only

Table 2.7: Commercially available 3D reconstruction hardware.

2.4 Commercial hardware solutions

Different commercial hardware solutions exist to gather 3D data. In table 2.7 a collection of the alternatives is shown.

Echoscope from Coda Octopus [105] commercialises a high resolution imaging sonar for shallow water applications. Conversely, Teledyne sells Blueview, an imaging sonar that offers less resolution, but mounted on a pan and tilt unit [106].

Teledyne sells an underwater LLS called INSCAN [107]. This system must be deployed underwater or fixed to a structure. The device samples 1 m^2 in 5 s at 5 m range.

3D at Depth sells a similar device called SL1 [61]. They have worked with Teledyne in their designs [108]. The specifications are the same as with the previous sensor.

3DLS is a triangulation sensor formed by an underwater dual laser projector and a camera. It is produced by *Smart light devices*. By default, uses a 15 W green laser.

2G Robotics has three different triangulation devices depending on the demanded range [109–111]. They all are motorized, so they must be deployed and static during their scan.

Savante provides three different solutions. Cerberus is a triangulation sensor formed by a laser pointer and a receiver, capable of recovering 3D information. SLV-50 is another triangulation sensor formed by a laser stripe and a high sensitivity camera and finally Lumeneye is a laser stripe that only casts laser light on the scene.

Tritech provides a simple green laser sheet called SeaStripe [112] similar to the previously presented by *Savante*. The 3D reconstruction must be performed by the client camera and software. They only provide the projecting laser. No motorization is involved.

2.5 Discussion

Roman *et al* [113] compared laser stripe scanning to stereoscopy and multibeam sonar in a real underwater scenario using *Hercules* ROV. The stereo data showed less definition than the sonar data, and the stereo 3D points were triangulated exclusively from SIFT descriptors, which are few compared to other sensors resolution. The comparison was made during a survey where laser images where collected at 3 Hz at $2-5$ cm/s at 3 m above the bottom, whilst stereo imagery was captured on a separate survey at 0.15 Hz and at a speed of 15 cm/s and a distance of $1.5-3$ m giving a minimum overlap of 50%. Multibeam was captured during the laser survey at 5 Hz.

As seen in these numbers, the data rate from the different sensors is not equal, and therefore the gathered data is temporally different, giving less or more spatial density depending on the setup. What is clear is that the authors used SIFT features in the stereo camera and abandoned the disparity images that could have been calculated with the left-right pairs, giving a denser point cloud compared to the (rather few) inliers achieved in the survey.

Stereoscopy is the easiest way to obtain the depth of a scene [65], followed by multibeam and laser stripe data, which are the easier sensor types to concatenate given a dataset with information on a set of the navigation sensors used in the ROV or AUV of the experiment.

Structure from Motion is easy to implement from a hardware point of view, as it only needs a camera attached to an already operative robot. The processing needed afterwards is, however, intensive. Also, if there is not a previous setup on a known target for measurement, depth data will be obtained without scale. A scale can be obtained afterwards by fusing the reconstruction with the navigation data or with a known target size. These approaches have already been integrated in commercial software and are accessible to the public [86].

Then, LLS and SL overcome some of the problems of the previously commented sensors. LLS travels further, achieving longer ranges minimizing the common volume scattering and taking advantage of fixed wavelength light sources to minimize absorption. On the other hand, when a precise and closer look to an object or structure is needed, LLS is not always available as it counts time from a signal to travel and return. Often, ToF sensors have a large minimum measuring distance. SL sensors are capable of hitting closer distances with high resolution, no matter what the texture or colour of the scanned object.

All the benefits and limitations of the sensors and methods presented in this chapter lead to choose the best suited sensor for each application. In the context of this thesis, three key topics have been raised: underwater exploration, light absorption and scattering and high resolution sensors. In this concern, we propose two solutions for underwater reconstruction that provide better insight in either size, speed or resolution of the reconstructed targets.

- **Laser-based Structured Light Systems** are capable of overcoming light limitation in the medium and provide sparse and quick three dimensional information of a small area, targeted at autonomous manipulation. The device can be simply attached to an underwater manipulator to provide high-framerate information, as it is just formed by a laser projector and a camera.

- **Laser Stripe** (LS) excels in mapping extensive areas as it requires a simple setup. One camera and one laser stripe mounted on a cruising AUV provide a robust and effective reconstruction solution, with an easy integration in work-class robots. The resulting range measurements can be integrated to a SLAM framework to provide better mapping accuracy and consistency.

The methodology for both of these approaches is further discussed in chapter 3. Once the foundations of triangulation and laser detection are settled, the document presents in chapter 4 a one-shot LbSLS, together with its experimental results. A second approach to laser structured light is delivered in chapter 5, where two SLAM solutions using LS are presented and discussed.

Pipeline for Laser-based Structured Light Systems

In this thesis, two three dimensional reconstruction methods and sensors are presented. Even if different, they use the same methodology and computational pipeline for a laser projector with a known pattern and a camera. The differences between them are the projected pattern and the decoding. This last step is only needed for one of the approaches: a multi-line laser pattern. This chapter first presents an introduction to laser-camera geometry in 3.1, followed by section 3.2, where the laser peak detection method is depicted. The triangulation of the detected points is explained in section 3.3 and finally in section 3.5 a summary of the chapter can be found.

In the following two chapters, a novel Laser based Structured Light System (LbSLS) [23, 24, 114] is proposed (chapter 4) and two bathymetric SLAM approaches applied to laser stripe 3D reconstruction are presented (chapter 5). They both have this pipeline in common.

3.1 Introduction

A laser-based structured light system is a sensor formed by one or more cameras and a laser projector. The projector casts a known laser pattern whose shape, deformed by the target area, is recovered by a camera. A relation can be established between range and resolution depending on where the projector is located with respect to the camera (e.g. its baseline). There is clearly a compromise between spatial resolution and range. At a larger distance to the target surface to measure, the size of a pixel on the target increases, decreasing its resolution.

The geometry of the system has to be designed depending on the desired range of the device. In figure 3.1 the most important geometry variables are indicated. The geometry will fix the maximum and minimum working distance between system and object, as described by equations 3.1 and 3.2.

$$d_{max} = \frac{b}{\tan\left(\psi + \frac{\varphi}{2}\right) - \tan\left(\frac{\theta}{2}\right)} \tag{3.1}$$

$$d_{min} = \frac{b}{\tan\left(\psi - \frac{\varphi}{2}\right) + \tan\left(\frac{\theta}{2}\right)} \tag{3.2}$$

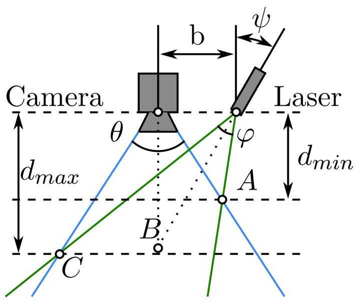


Figure 3.1: Laser to camera geometry. The point B is the crossing point of the camera axis and the laser axis. d_{min} and d_{max} distances are related to the field of view of the camera, being d_{max} the maximum distance at which all the projected pattern from the laser is seen from the camera, and d_{min} the minimum.

where b is the baseline, θ is the camera field of view and φ is the laser beam divergence. The relative rotation between the camera and laser is ψ. In some cases, the maximum working distance is infinite, if and only if the laser beam crosses the camera field of view only once. These relationships also apply for laser triangulation uncertainty. The uncertainty introduced by using a discrete laser line extraction can be calculated from the geometry in figure 3.2 using equations (3.3) to (3.6).

$$z_{max} = \frac{b}{\tan\left(\alpha - \frac{\beta}{2}\right) + \tan\left(\psi - \frac{\varphi}{2}\right)} \tag{3.3}$$

$$z_{min} = \frac{b}{\tan\left(\alpha + \frac{\beta}{2}\right) + \tan\left(\psi + \frac{\varphi}{2}\right)} \tag{3.4}$$

$$\Delta z = z_{max} - z_{min} \tag{3.5}$$

$$\Delta x = z_{max} \tan\left(\alpha + \frac{\beta}{2}\right) - z_{min} \tan\left(\alpha - \frac{\beta}{2}\right) \tag{3.6}$$

where z_{max} and z_{min} are the the maximum and minimum possible distances from the camera to the triangulated point. This distances vary depending on the laser tilt ψ and divergence angle φ, the aperture of one camera pixel β and its angle α and the baseline. Therefore, the assignment of a laser detection to an integer pixel position restricts the depth resolution as shown in figure 3.3.

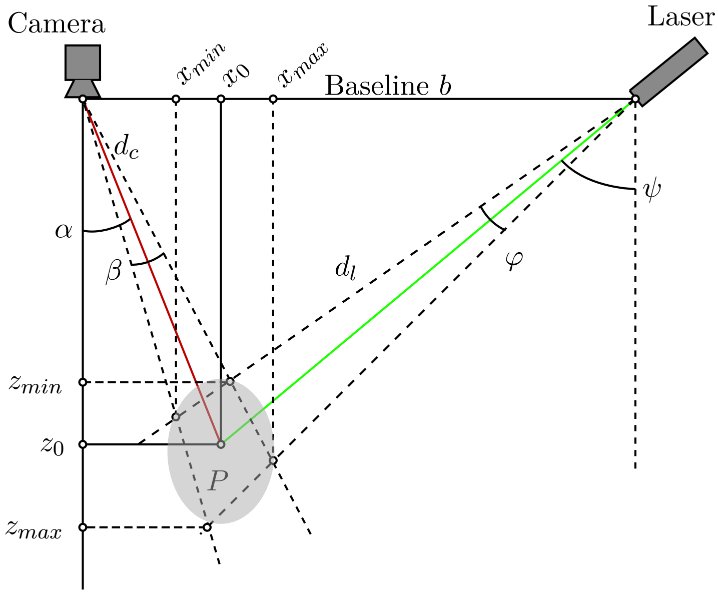


Figure 3.2: A camera ray d_c at an angle α from the principal axis of the camera. The ray crosses the focal point of the camera and projects to a pixel. The pixel size in 3D space translates to a cone with an aperture angle β. At a distance b, a laser with a beam divergence angle φ is tilted at an angle ψ with respect to the vertical. The laser beam d_l crosses the camera ray d_c at a point P in space. The differences in aperture angle and beam divergence at a distance z_0 causes a depth uncertainty and a position uncertainty depicted by $\Delta z = z_{max} - z_{min}$ and $\Delta x = x_{max} - x_{min}$.

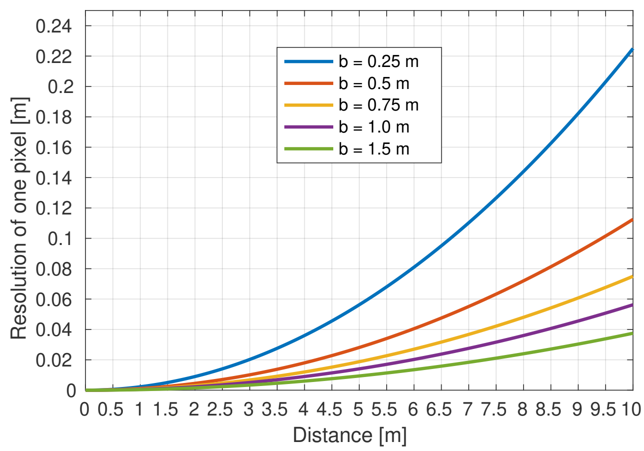


Figure 3.3: Camera pixel resolution vs distance to the target for a camera with a focal distance of $8mm$ and a pixel size of $4.5\mu m$. The resolution depends on the baseline b of the structured light sensor.

.

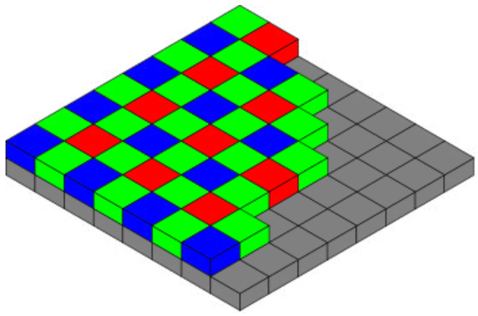


Figure 3.4: Bayer pattern on sensor. The pattern is formed by a series of passband lenses (shown in blue, green and red) that only allow a certain light wavelength band to pass to the sensor, depicted in gray [6].

3.2 Laser peak detection

The detection of laser light is performed by imaging sensors. These colour cameras are usually manufactured with a Bayer filter in front of the sensor. These filter the light spectrum to what is commonly known red, blue and green channels, or *RGB*. Each channel has different spectrum response, given by the manufacturer. Commonly, the spectral response of these channels overlap up to some point. On the other hand, a laser has a narrow wavelength light that might fall in between these overlapping colour bands. If this is the case, we can weight our three channel input images.

A demosaicking or debayering algorithm is applied to the captured image to reconstruct the actual colour of the light that entered the camera lens. Most commonly, the resulting RGB value for a pixel in the three channel image is interpolated among the nine closest pixels in the Bayer array.

Then, given a three channel image $I : I_R, I_G, I_B$, the image sensor sensitivity and the laser wavelength, the linear function that better measures the laser wavelength can be determined. This transformation is shown in equation (3.7).

$$I_L = f(k_R, k_G, k_B) = k_R I_R + k_G I_G + k_B I_b, \quad ||(k_R, k_G, k_B)|| \leq 1 \tag{3.7}$$

where k_R, k_G, k_B have to be retrieved from the sensor sensitivity chart.

3.2.1 Light underwater

As seen in section 1.2.2, absorption and scattering play an important role in the behaviour of light underwater. Absorption by seawater is weak in the blue colour spectra, whilst strong in the red. It varies with temperature and salinity, as well as with algal or phytoplankton particles. This can be modelled as an exponential loss in irradiance (E $[W/m^2]$) as depicted in equation (3.8), where E_0 is the source irradiance, a is the absorption coefficient and z is the distance from the source to the receiver.

$$E = E_0 \exp(-az) \tag{3.8}$$

Scattering is the deflection of a photon by a scattering particle, considered as a sphere with a particular geometrical size. This sphere redirects incident photons into new directions and prevents the forward transmission of photons. The scattering coefficient b $[m^{-1}]$, describes a medium containing a volume density ρ_S $[m^{-3}]$ of scattering particles with an effective cross-section of σ_S $[m^2]$. The scattering coefficient is essentially the cross-sectional area per unit volume of medium, defined in equation (3.9).

$$b = \rho_S \sigma_S \tag{3.9}$$

The angular dependence of scattering is called the volume scattering function (VSF) $p(\theta)$ $[sr^{-1}]$. This function describes the probability of a scattered photon into a unit solid angle oriented at an angle θ relative to the original photon trajectory. This scattering function is assumed azimuthally symmetric. The scattering coefficient (b $[m^{-1}]$) is a measure of the overall magnitude of the scattered light without regard to its angular distribution, which is the integral of the VSF over all angles, as shown in equation (3.10).

$$b = \int_0^{4\pi} p(\Omega)\, d\Omega = 2\pi \int_0^{\pi} p(\theta) \sin\theta \, d\theta \tag{3.10}$$

The backscattering coefficient is defined as the total light scattered back to the hemisphere from which light came from, shown in equation (3.11).

$$b_b = \int_{2\pi}^{4\pi} p(\Omega)\, d\Omega = 2\pi \int_{\pi/2}^{\pi} p(\theta) \sin\theta \, d\theta \tag{3.11}$$

And the backscattering ratio is defined in (3.12).

$$\tilde{b} = \frac{b_b}{b} \tag{3.12}$$

Normally, scattering is defined by the phase function, which is the VSF normalized to the total scattering. It provides information about the shape of the VSF regardless of the intensity of the scattered light, defined in equation (3.13).

$$\int_0^{\pi} \tilde{p}(\theta) 2\pi \sin\theta \, d\theta = 1, \quad \tilde{p}(\theta) = \frac{p(\theta)}{b} \tag{3.13}$$

An isotropic phase function (3.14) would scatter light equally into all possible directions:

$$p(\theta) = \frac{1}{4\pi}, \text{ such that } \int_0^{\pi} p(\theta) 2\pi \sin\theta \, d\theta = 1 \tag{3.14}$$

The anisotropy, g is a dimensionless measure of the amount of forward direction retained after a single scattering effect. In ocean waters, g varies from 0.85 to 0.95. Imagine that a

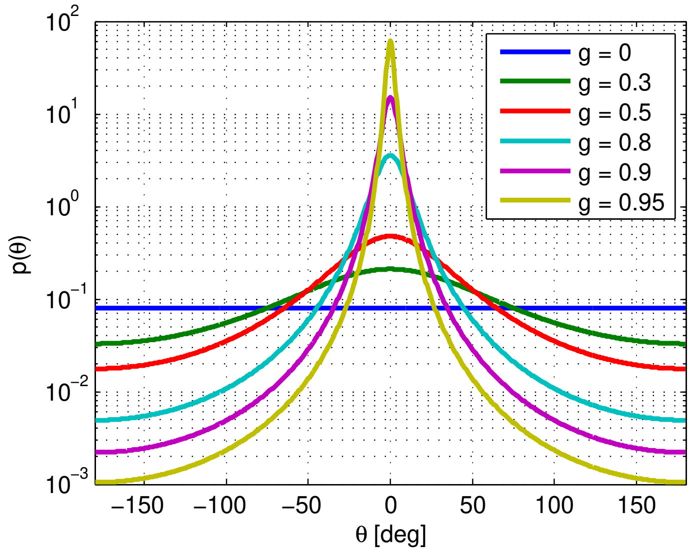


Figure 3.5: Heyney-Greenstein phase function

photon is scattered by a particle so that its trajectory is deflected by a deflection angle θ. Then the component of the new trajectory which is aligned in the forward direction is $\cos(\theta)$. There is an average deflection angle whose value $\cos(\theta)$ is defined as the anisotropy, shown in equation (3.15).

$$g = 2\pi \int_0^\pi p(\theta) \cos\theta \sin\theta \, d\theta = \langle \cos\theta \rangle \tag{3.15}$$

Heyney-Greenstein (HG) phase function (3.16) has been widely used in oceanography as it is an analytic formula that approximates the shape of an actual phase function. Note that it is normalized. In figure 3.5 the phase function is plotted for different g.

$$P_{HG}(\theta) = \frac{1}{4\pi} \left[\frac{1 - g^2}{(1 - 2g\cos\theta + g^2)^{3/2}} \right] \tag{3.16}$$

Its backscatter ratio can be computed from equation (3.17)

$$\begin{aligned}
\frac{b_b}{b} &= 2\pi \int_{\pi/2}^{\pi} P_{HG}(\theta) \sin\theta \, d\theta && (3.17)\\
&= 1 - 2\pi \int_0^{\pi/2} P_{HG}(\theta) \sin\theta \, d\theta && (3.18)\\
&= 1 - \frac{1+g}{2g} \left[1 - \frac{1-g}{\sqrt{1+g^2}} \right] && (3.19)
\end{aligned}$$

3.2.1.1 Application on laser light

Assuming that a Gaussian is a reasonable fit to the phase function (valid in the limit of many collisions), the anisotropy can be written as equation (3.20).

$$g = 1 - \frac{\langle \left(2\sin\frac{\theta}{2}\right)^2 \rangle}{2} \approx 1 - \frac{\langle \theta_W^2 \rangle}{2} \tag{3.20}$$

The irradiance, E $[W/m^2]$ of a Gaussian laser beam profile is described as a function of the radial position r $[m]$ from the central axis of the beam, shown in (3.21).

$$E_{LU}(r) = P\frac{2}{\pi\sigma_L^2}\exp\left(-\frac{2r^2}{\sigma_L^2}\right) \tag{3.21}$$

where P $[W]$ is the power of the laser beam, and σ_L $[m]$ is the laser radius taking divergence into account. This divergence is modelled as if the beam came from such source point at a distance L_L $[m]$ that would have the appropriate size at the output of the system (e.g. water window interface) as modelled in (3.22).

$$\sigma_L^2 = \theta_L^2(L_L + z)^2 \tag{3.22}$$

where θ_L $[rad]$ is the laser divergence and z is the actual working distance from the water window interface. $E_{LU}(r)$ is the laser beam propagation without scattering.

Example 3.2.1. *Consider a collimated laser beam delivering* $1\ W$ *of power to a circular 1-mm-diameter aperture. The laser beam has a Gaussian beam profile with a* $1/e^2$ *radius of* $w_0 = 0.5\ mm$ *at the work point. A detector sits behind the aperture and is greater than* $1\ mm$ *in diameter.*

$$\begin{aligned}
P_{collected} &= \int_S E(S)\,dS = 2\pi\int_0^a E(r)r\,dr && (3.23)\\
&= 2\pi\int_0^a P\frac{2r}{\pi\sigma_L^2}\exp\left(-\frac{2r^2}{\sigma_L^2}\right)dr && (3.24)\\
&= \frac{4P}{\sigma_L^2}\int_0^a r\exp\left(-\frac{2r^2}{\sigma_L^2}\right)dr && (3.25)\\
&= \frac{4P}{\sigma_L^2}\cdot\frac{\sigma_L^2}{4}\exp\left(-\frac{2r^2}{\sigma_L^2}\right)\Bigg|_{r=0}^{r=a} && (3.26)\\
&= P\left[1-\exp\left(-\frac{2a^2}{\sigma_L^2}\right)\right] && (3.27)\\
&= 1\left[1-\exp\left(-\frac{2\cdot 0.5^2}{0.5^2}\right)\right] = 0.865\ W && (3.28)
\end{aligned}$$

The forward scattered beam can be approximated by an expanding Gaussian fan of light. The effect of absorption has been neglected, and the approximation of small angle has been used.

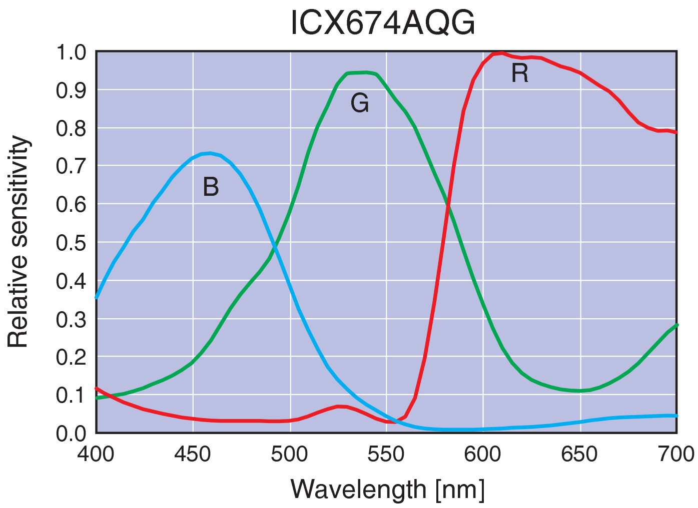


Figure 3.6: Color spectrum response of Sony ICX674. Every wavelength has a different gain in the color channels RGB, red, green and blue respectively.

3.2.2 Gaussian laser detection

As seen in the previous section, a collimated laser source spreads with distance and can be modelled using a Gaussian shape. This light can be easily detected in an image because its wavelength is known. Therefore a reward function can be tailored to be more sensitive to the colour spectrum of the light source. For the particular case of green laser light at 532 nm and a Sony ICX674 sensor, the coefficients k_r, k_g, k_b would have to be $0.08, 0.85$ and 0.2 respectively, as can be seen in figure 3.6.

Common line image detectors such as Canny do not work well for laser stripes, mainly because they are sensible to straight edges and there are no such edges underwater, apart from man-made objects, such as structures, hubs or cables. Therefore the laser line has to be detected by other means, and this thesis proposes to use a sliding window integral.
Other authors [115] have proposed to use the first derivative to find the subpixel centre, but this approach is not suitable for underwater applications. Although the laser stripe yields a Gaussian profile, the discrete-nature of the camera sensor makes interpolation necessary and the larger distances from the camera to the seafloor compared to air-solutions reduces the number of pixels lit by the laser.

Given an image I_L, we compute the integral over a window of length l_w starting at the row v_0 until $v_0 + l_w$, which can be written as (3.29)

$$G_{v_w} = \sum_{v=v_0}^{v_0+l_w} \left(1 - 2\left|v_0 + \frac{l_w - 1}{2} - v\right|\right) \cdot I_L(u,v) \tag{3.29}$$

where $I_L(u,v)$ is the value at pixel (u,v) of the reward image computed using (3.7). This window is rolled over all available rows in the image to find the highest integrated value. If this value is higher than a minimum integral value threshold, the found window is a candidate laser point and a subpixel laser detection algorithm is used for refinement.

The subpixel detector used is a Gaussian approximation (GA) [116], shown in equation (3.30), where $f(x)$ is the intensity value of a particular row at pixel x, usually an integer in the range 0-255, and $\hat{\delta}$ is the subpixel offset with respect to the centre of the integral window used to locate the most likely laser point.

$$\hat{\delta} = \frac{1}{2} \frac{\ln(f(x-1)) - \ln(f(x+1))}{\ln(f(x-1)) - 2\ln(f(x)) + \ln(f(x+1))} \tag{3.30}$$

In the cases were the denominator of equation (3.30) is zero (e.g. when the pixel intensity is the same in three pixels, or when the laser light has burnt the image at the laser point) the Center of Mass (CoM5) detector is used [116]. Once all columns have their laser peak candidate, the peaks are grouped in patches. A patch is formed by laser peaks that are contiguous, defining contiguous as a maximum distance of three pixels. Then, the patches that have less than five contiguous peaks are removed from the list, considered as spurious and therefore outliers.

3.2.3 Application of laser detection

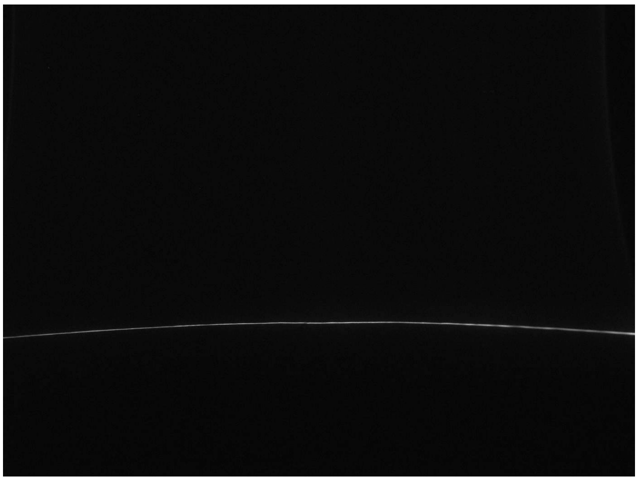
The application of these detection steps is shown in figure 3.7. Figure 3.7(a) shows an example frame gathered during a real survey mission. The laser line can be seen on the bottom part of the image. The column 100 is shown in figure 3.7(b), where the intensity profile clearly shows a peak in the laser portion in the image. Finally, applying this procedure per column, produces the laser line subpixel detection highlighted in figure 3.7(c), where the area around the previous shown column 100 is overlayed with the detected peaks.

3.3 Triangulation

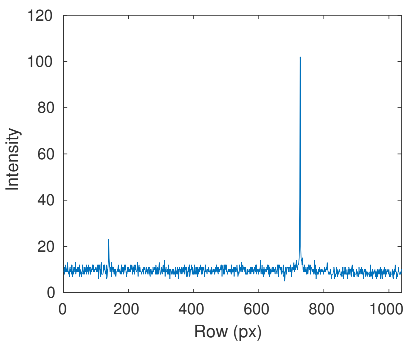
Once the laser peaks are known, three dimensional points can be computed from two dimensional detections. To do that, the camera intrinsic and extrinsic calibration parameters are needed, as well as the laser to camera transformation.

Let C be a camera in 3D space, oriented towards z with x pointing to its right. Using a pinhole camera model, a point X in the space can be back-projected in the camera image, and a point in the camera image can be projected into a line. This setup is shown in figure 3.8.

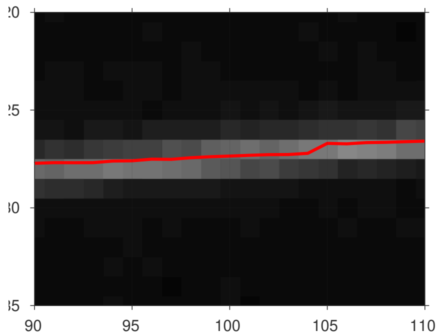
The intrinsic matrix $\boldsymbol{K}$ is parametrized by Hartley and Zisserman [93] as shown in (3.31), and the extrinsic $(\boldsymbol{R}|\boldsymbol{t})$ as a 3×4 transformation matrix comprising a rotation and a trans-



(a) Example laser line image captured at $1392 \times 1040\,px$ resolution.



(b) Intensity values at column 100.



(c) Subpixel laser peak detection around column 100.

Figure 3.7: Laser peak detection.

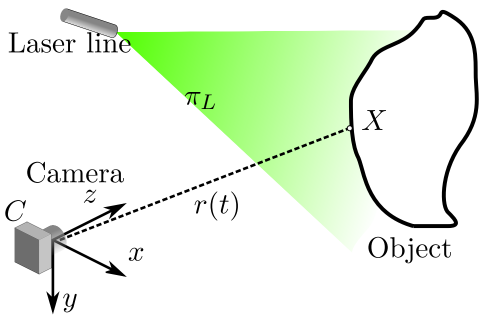


Figure 3.8: Laser-Camera geometry. A point X in space can be computed by projecting a line from the camera at the point where it intersects the laser plane π_L.

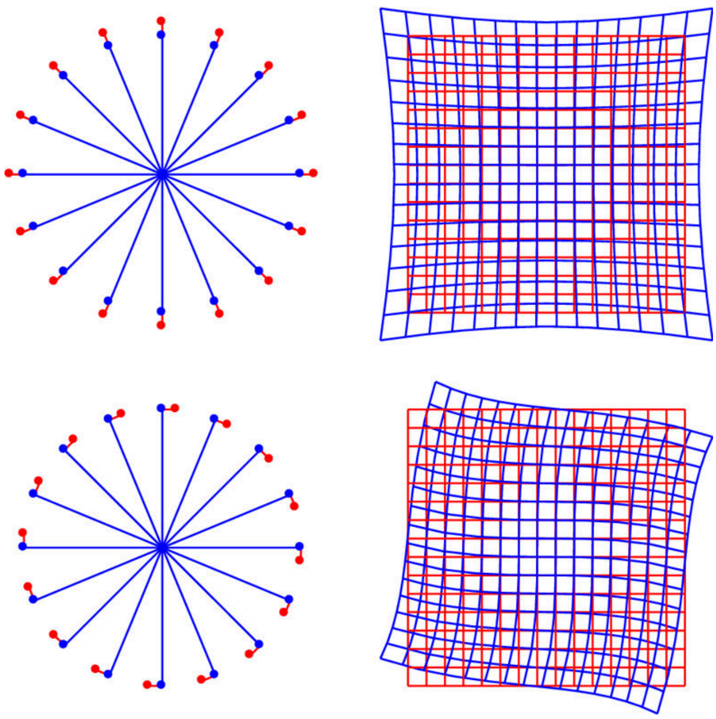


Figure 3.9: Example of how distortion parameters affect the image. On the top, the radial distortion parameter k_1 distorts a circle and a grid, and on the bottom a tangential distortion parameter p_1.

lation.

$$\boldsymbol{K} = \begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix} \tag{3.31}$$

Thus, a point $\boldsymbol{x} = (x, y, z)^\intercal$ can be related to its 2-dimensional image projection $\boldsymbol{p} = (u, v)$ with equation (3.32) when the camera lens produces no distortion at all.

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \boldsymbol{K} \left(\begin{array}{c|c} \boldsymbol{R} & \boldsymbol{t} \end{array} \right) \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \tag{3.32}$$

Real lenses usually have some distortion, as depicted in figure 3.9, mostly radial and slight tangential distortion. Take the position of a point $\boldsymbol{x}_d$ relative to geometric image centre defined by $\boldsymbol{c} = (c_x, c_y)$. Given the pinhole camera model, assume that $\boldsymbol{c}$ is undistorted in the resulting image, but not $\boldsymbol{x}_d$. If the position of $\boldsymbol{x}_d$ is only distorted radially along direction $\overrightarrow{cx_d}$, the distortion is said to be radial. Alternatively, if $\boldsymbol{x}_d$ is also displaced tangentially relative to the circle of radius $r = d_2(\boldsymbol{c}, \boldsymbol{x}_d)$, the distortion is said to be tangential. So, the above model is extended with equations (3.34) to (3.37), where $\boldsymbol{x}_d = (x_d, y_d, z_d)$ is the distorted point expressed in the camera coordinate system.

$$\begin{pmatrix} x_d \\ y_d \\ z_d \end{pmatrix} = \begin{pmatrix} \boldsymbol{R} \mid \boldsymbol{t} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \tag{3.33}$$

$$x' = x_d/z_d \tag{3.34}$$

$$y' = y_d/z_d \tag{3.35}$$

$$x'' = x' \frac{1 + k_1 r^2 + k_2 r^4 + k_3 r^6}{1 + k_3 r^2 + k_4 r^4 + k_5 r^6} + 2p_1 x' y' + p_2 (r^2 + 2x'^2) \tag{3.36}$$

$$y'' = y' \frac{1 + k_1 r^2 + k_2 r^4 + k_3 r^6}{1 + k_3 r^2 + k_4 r^4 + k_5 r^6} + p_1 (r^2 + 2y'^2) + p_2 x' y' \tag{3.37}$$

where $r^2 = x'^2 + y'^2$. The k_i coefficients account for radial distortion and p_j for tangential distortion. These coefficients along with the intrinsic ones, have to be obtained through camera calibration. Finally, using (3.38) the undistorted pixel coordinates can be obtained.

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \boldsymbol{K} \begin{pmatrix} x'' \\ y'' \\ 1 \end{pmatrix} \tag{3.38}$$

In a projective pinhole camera, all the rays of light impinging on it, pass through the focal point. Hence, all the points laying on the same ray of light cast on the same point on the image plane. The direction of these rays is important, the distance to the point that generates them is irrelevant in the image formation. In this sense, one can think that a point on the image plane corresponds to a line in 3-space, which contains both the focal point and the image point. In projective geometry, all the points contained in the same line are equivalent. We can write the line equation in parametric form with (3.39).

$$r(t) = \begin{cases} x & = \frac{u - c_x}{f_x} t \\ y & = \frac{v - c_y}{f_y} t \\ z & = t \end{cases} \tag{3.39}$$

Finally, the laser line is described as a plane π_L in space with known coefficients whose equation can be written as (3.40)

$$\pi_L : Ax + By + Cz + D = 0 \tag{3.40}$$

The intersection between the laser plane and the line originated from the camera, that passed through the detected laser point at (u, v) is the triangulated point $\boldsymbol{x}$ in space, computed by equation (3.41).

$$\boldsymbol{x} = r(t) \cap \pi_L = \begin{cases} x & = \frac{u-c_x}{f_x} t \\ y & = \frac{v-c_y}{f_y} t \\ z & = t \end{cases}, \text{ with } \quad t = -\frac{D}{A\frac{u-c_x}{f_x} + B\frac{v-c_y}{f_y} + C} \tag{3.41}$$

Therefore a point cloud can be defined as (3.42)

$$\boldsymbol{P_i^C} = (\boldsymbol{x_1}, \ldots, \boldsymbol{x_m}) \tag{3.42}$$

where m is the number of points in the i-th frame.

3.3.1 Dead reckoning reconstruction

In navigation, Dead Reckoning (DR) is the process of estimating one's current position by using a previous known position and advancing that position based on known or estimated speeds. For AUVs, DR is usually obtained using sensors such as Global Positioning System (GPS), Inertial Magnetic Unit (IMU), Doppler Velocity Log (DVL), depth sensors and in some cases Long Base Line (LBL) or Short Base Line (SBL) position updates. The estimation is never completely correct, and prone to drift over time. If the system is mounted on a moving platform that gathered N measurements with known position and orientation, its point clouds can be concatenated to form a large set, formulated as (3.43)

$$\boldsymbol{P_{DR}^W} = \bigcup_{i=1}^{N} \left(\boldsymbol{P_i^C} \oplus \boldsymbol{T}_C^R \oplus \boldsymbol{T}_{Ri}{}^W \right), \tag{3.43}$$

where $\boldsymbol{P_i^C}$ is the point set of the i-th frame, $\boldsymbol{T}_C^R$ is the transformation from the camera coordinate system to the vehicle coordinate system, $\boldsymbol{T}_{Ri}{}^W$ is the DR position of the vehicle at the time of the i-th frame with respect to the word, and $\oplus$ is the composition operator.

3.4 Laser Plane Calibration

A large number of computer vision algorithms involve the computation of parameters in the presence of noise or in noisy measurements. These parameters can be part of an equation or the components of a matrix. For example, the intrinsics of a pinhole camera model, the computation of stereo correspondences, homographies and geometric calibrations. In essence, all of the above problems can be defined as the search for an optimal solution to an overdetermined system of linear equations, which can be formalised in equation (3.44), where $\boldsymbol{x}$ are the parameters to minimize and $\boldsymbol{A}$ the linear set of equation that relates them. There is always an avoided trivial solution, when $\boldsymbol{x} = 0$.

$$\boldsymbol{A} \cdot \boldsymbol{x} = 0 \tag{3.44}$$

A camera-laser pair needs to be calibrated to determine the spatial relationship between the laser plane with reference to the camera coordinate system. The plane parameters to compute can either be $\{A, B, C, D\}$ from equation (3.40), or their dual representation π_L : $(\boldsymbol{c}, \vec{n})$ where the plane is defined by a crossing point $\boldsymbol{c}$ and its normal vector $\vec{n}$. Given a set of detected laser points $\boldsymbol{p_i}$, whose centroid is $\boldsymbol{c}$, a plane can be fitted to the points by solving (3.45).

$$\min_{\boldsymbol{c},||\vec{n}||=1} \sum_{i=1}^{n} ((\boldsymbol{p}_i - \boldsymbol{c})^\intercal \vec{n})^2 \tag{3.45}$$

Introducing the $3 \times n$ matrix $\boldsymbol{A} = [\boldsymbol{p}_1 - \boldsymbol{c}, \boldsymbol{p}_2 - \boldsymbol{c}, \ldots, \boldsymbol{p}_n - \boldsymbol{c}]$, the problem can be formulated as shown in equation (3.46), which is the expression of a parameter estimation problem, that can be solved by computing the vector $\vec{n}$ that minimises some cost function of the matrix equation $\boldsymbol{A} \cdot \boldsymbol{x} = 0$.

$$\min_{||\vec{n}||=1} ||\boldsymbol{A}^\intercal \vec{n}||^2 \longrightarrow \boldsymbol{A} \cdot \boldsymbol{x} = 0 \tag{3.46}$$

Using the singular value decomposition $A = USV^\intercal$ the plane normal $\vec{n}$ can be found as the third column of U, equivalently, $\vec{n} = U(:, 3)$. It follows from the orthogonality of U that the plane is spanned by its two first columns.

In this thesis we present two methods for laser plane calibration. The first method uses a chessboard pattern as a known surface where the laser line is projected, and the second method uses a second camera and stereoscopy to compute the 3D laser points and the laser plane.

3.4.1 Calibration with chessboard pattern

The setup for this calibration method requires to place the camera-laser pair at different distances to a calibration plane. The calibration plane π_C can be generated using a common chessboard pattern commonly used for camera calibration. Then the laser points can be triangulated in 3-space by intersection with the calibration plane. The described setup can be seen in figure 3.10.

The calibration data is generated by changing the distance between the sensor and the plane for m camera positions $\boldsymbol{T} = \{t_1, t_2, \ldots, t_m | \boldsymbol{t}_i = [x_i, y_i, zi, \phi_i, \theta_i, \psi_i]\}$. Next, for each calibration plane π_{pi}, a set of 3D plane points $\mathbf{P} = \{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_m | \boldsymbol{p}_i = [p_x, p_y, p_z]\}$ gathered belongs to the unknown laser plane. Finally, the same equations (3.45) and (3.46) are used to obtain the laser plane equation.

Spurious measurements and outliers can appear during calibration and we have to be able

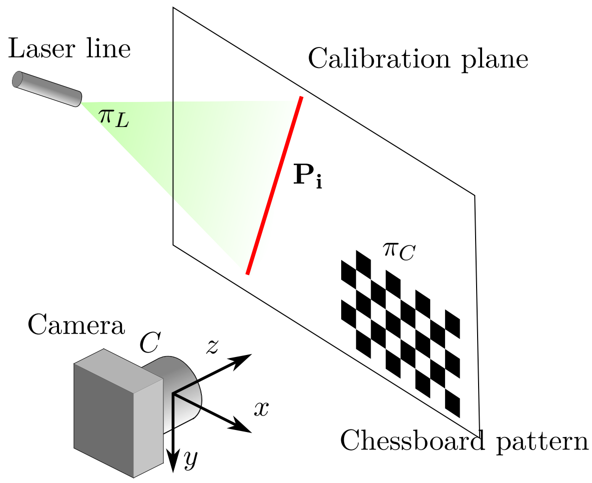


Figure 3.10: Laser calibration with checkerboard.

to either remove them or fit a model without accounting for them. To deal with outliers a Random Sample Consensus (RANSAC) iterative method has been implemented to refine the result. RANSAC is a learning algorithm technique that estimates a model parameters by random sampling the observed data. A random sample of points is used with (3.46) and the obtained parameters are validated against the whole dataset. The validation is the computation of L2 point-plane distance error. Depending on a minimum error threshold, every point can be considered as an inlier, if its error is less than the threshold, or as an outlier. The algorithm is iterated over random samples of data until one of these criteria are met: (1) a maximum number of iterations has been reached, (2) a maximum number of inliers has been reached, and (3) the error has fallen below a predefined threshold, set at the beginning of the algorithm. In any of these cases, the model with the least error is chosen.

Finally, the complete method to calibrate the laser using a calibration plane involves the following steps:

1. Calibrate the camera using Zhang's method [93].

2. Prepare a planar surface with a known marker or chessboard. The surface must be wide enough to accommodate the projected laser line.

3. Place the structured light system with the camera pointing to the calibration pattern and the laser line projected next to it.

4. Use the algorithm 3.1 to obtain the laser calibration.

Algorithm 3.1: Offline Laser to camera calibration algorithm.

Data: N frames with laser line and calibration pattern
Result: Laser plane position π_L with respect to the camera

$\boldsymbol{P} = \varnothing$ ▷ To store 3D points
for *each frame* $i = 1, \dots, N$ **do**
  Detect the calibration plane $\rightarrow \pi_C$
  Detect the laser points $\boldsymbol{P}_i$ and triangulate them to form lines
  Intersect each line with the plane π_C and append the 3D points into $\boldsymbol{P}$
end
RANSAC: Form M subsets of k points from $\boldsymbol{P}$
for *each subset* $j = 1, \dots, M$ **do**
  Compute the centroid of $p_j \longrightarrow \boldsymbol{c}_j$
  Subtract the centroid $\boldsymbol{c}_j$ to all points $\boldsymbol{P}$
  Use SVD to find the plane normal $\vec{n_j}$
  Define $\pi_{j,L} : (\vec{n}_j, \boldsymbol{c}_j)$
  Compute the distance sum d_j of all points $\boldsymbol{P}$ to the plane $\pi_{j,L}$
end
Return the plane that fits most points (e.g. minimizes the distance d_j)

This calibration method has been applied to the structured light system mounted on Turbot AUV, a SparusII model owned by the University of the Balearic Islands (UIB). The AUV mounts a left-right stereo rig with a baseline of 14.5 *cm* at its nose and a laser line at 22 *cm* from the cameras, centred, and tilted 10° degrees forward with respect to the vertical. The geometry has been targeted at low-altitude mapping ranging between 1 to 3 meters. The AUV is shown in figure 3.11 showing the configuration described.

The gathered point cloud together with the resulting laser plane can be seen in figure 3.12(a), and the final plane estimation error histogram is depicted in figure 3.12(b) with a mean of 0.6 *mm* and a standard deviation of 0.65 *mm*.

This method does not require to place the camera-laser system in a water tank. If the camera is calibrated both in air and in water, the relative position of the laser plane with respect to the camera will remain constant as long as the laser pointer is assumed to be perpendicular to its window housing. A solution to calibrate a system with this kind of method is to hang the vehicle with the system mounted on a crane and use the floor as a calibrated plane. If this kind of calibration is not feasible due to size or complexity, the following insitu calibration method can overcome these drawbacks.

3.4.2 Insitu calibration

This calibration method consists of a moving platform with a stereo camera and a laser structured light system, positioned either in fore-aft or left-right arrangement. The technique

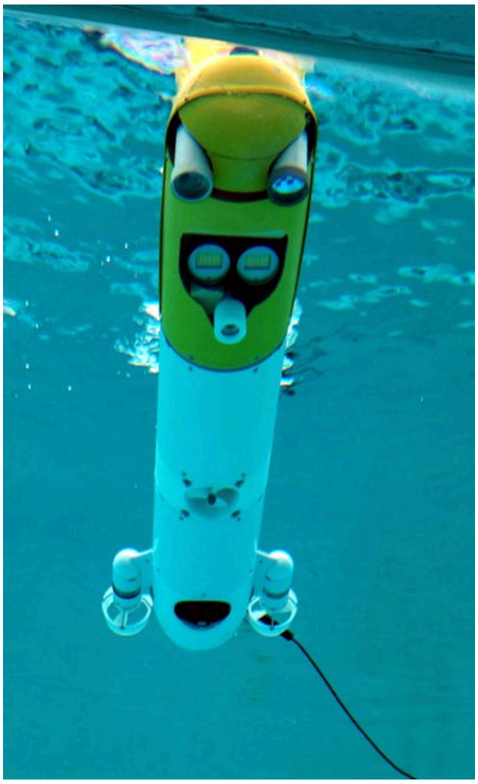
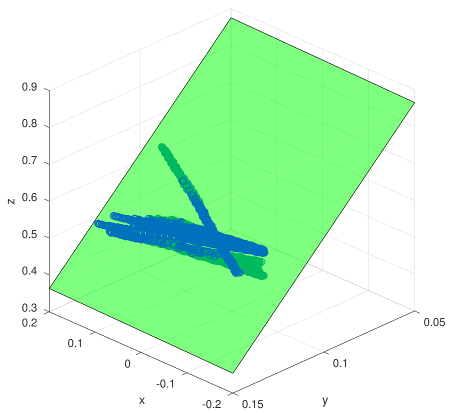
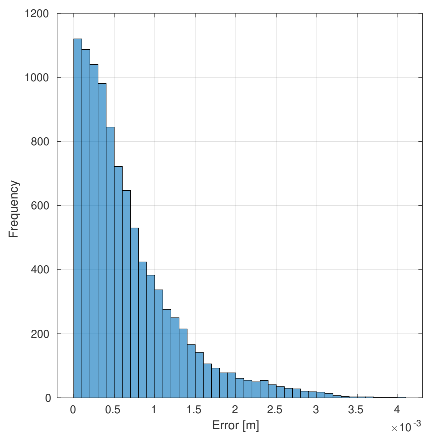


Figure 3.11: Bottom view of Turbot AUV showing (front to back) a stereo rig, two LED lights and a laser (white housing) located at its nose. At the centre of the craft there is a vertical thruster and at the rear, two surge thrusters and a DVL can be spotted.



(a) Three dimensional points used for calibration and the resulting fitting plane.



(b) Error histogram plot.

Figure 3.12: Calibration using a chessboard pattern for the platform Turbot AUV performed in the University of the Balearic Islands (UIB).

uses epipolar geometry of the stereo pair to match the detected laser line and generate a set of 3D points to determine the laser plane. The setup is similar to the one presented by Inglis and Roman [117], where the authors present a left-right stereo rig and a laser line insitu calibration. In their setup, the laser line is parallel to the baseline. This requires a non-flat seafloor terrain to find stereo-laser correspondences for their calibration setup, as well as feature descriptors to remove outliers. In the method presented in this document, the performance of the calibration method is not sensitive to the structure of the seafloor environment and feature descriptors are not required. It follows the work of Leat *et al* [19], extended to compute the position of a camera system with respect to the laser plane.

The method main steps are: (1) collect stereo calibration images, (2) collect laser calibration images and (3) compute the laser plane coefficients. Supposing an already calibrated stereo camera, the gathering of laser calibration images can be performed at the dive phase of a survey. Commonly, non-hover-capable AUVs have to dive in a circular motion whilst increasing their depth, resulting in a helix-like path. During this phase the stereo structured light system can collect a set of images at a decreasing altitude rate, sensing 3D laser points throughout the field of view of the camera. This step is shown in figure 3.13.

Using the method presented in section 3.2, the peaks are detected in both cameras. The distance between two laser peaks on an epipolar line is used to compute the distance to the seabed from the cameras. If the stereo rig is aligned fore-aft, the epipolar lines are vertical in the images, and if the stereo rig is aligned left-right, the epipolar lines are horizontal. Therefore, the laser should not be placed completely horizontal in a left-right setup nor completely vertical in a fore-aft configuration. In the particular case of a fore-aft setup, the coordinates of the points of interest can be computed using (3.47),

$$\boldsymbol{p}_i = \left(\frac{u_{i1} - c_x}{f}, \frac{v_{i1} - c_y}{f}, 1\right) \cdot \frac{f \cdot b}{v_{i2} - v_{i1}} \tag{3.47}$$

where f is the focal length; (c_x, c_y) are the coordinates of the centre of the image; b is the baseline, the distance between the cameras in meters; (u_{ij}, v_{ij}) are the pixel i coordinates of the camera j. Once the set of 3D points $\mathbf{P} = \{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_m = [x, y, z]\}$ have been computed for the length of the helix path, the algorithm 3.2 can be used to calibrate the laser plane.

Algorithm 3.2: Insitu Laser to camera calibration algorithm.

Data: N frames with laser line and known AUV pose

Result: Laser plane position π_L with respect to the camera

$\boldsymbol{P} = \varnothing$ ▷ To store 3D points
for *each frame* $i = 1, \ldots, N$ **do**
  Detect the laser points on the left frame and triangulate them to form lines
  Detect the laser points on the right frame and triangulate them to form lines
  Find the crossing points and append them into $\boldsymbol{P}$
end
RANSAC: Form M subsets of k points from $\boldsymbol{P}$
for *each subset* $j = 1, \ldots, M$ **do**
  Compute the centroid of $p_j \longrightarrow \boldsymbol{c}_j$
  Subtract the centroid $\boldsymbol{c}_j$ to all points $\boldsymbol{P}$
  Use SVD to find the plane normal $\vec{n_j}$
  Define $\pi_{j,L} : (\vec{n}_j, \boldsymbol{c}_j)$
  Compute the distance sum d_j of all points $\boldsymbol{P}$ to the plane $\pi_{j,L}$
end
Return the plane that fits most points (e.g. minimizes the distance d_j)

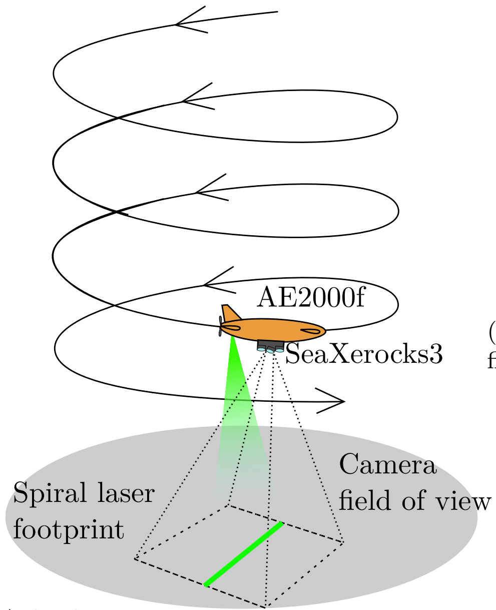
To validate the SVD plane solution, a error histogram is computed in figure 3.14 with a mean of 11.6 mm and a standard deviation of 10 mm. The error is defined as the L2 distance of a point p to its projection in the computed laser plane.

3.5 Summary

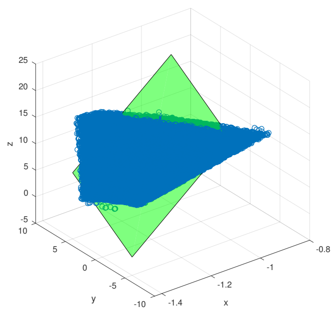
In this chapter, a complete processing pipeline has been presented to reconstruct the environment using a laser line projector and a camera rig, e.g. a laser-based structured light sensor. The proposed Gaussian peak detection with sub-pixel accuracy is backed up with the laser light modelling, where we have seen that a laser beam can be considered Gaussian subject to scattering and absorption.

Two calibration routines have been proposed, the first method based on calibration plane is based on a state of the art calibration for laser lines [70]. The second *insitu* method is based on stereoscopy. Both methods are capable of obtaining a correct laser calibration, the decision of choosing one or the other will depend more on the logistics or on the difficulty and the repeatability.

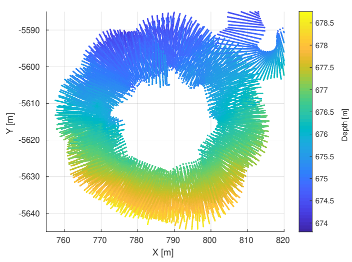
Despite the good results of the first method for the Turbot AUV structured light sensor, the reader has to note two important facts. First of all, the calibration points were gathered at a closer distance (≈ 0.5 m) than the points in AE2000f, which have a centroid at 10 m in



(a) An AUV follows a helix-like path to reach the required altitude. During its descent laser images are acquired, generating a dataset with a large Z disparity.



(b) Laser plane fitting points and final model.



(c) Reconstructed seafloor during the helix descent.

Figure 3.13: Insitu stereo-laser calibration.

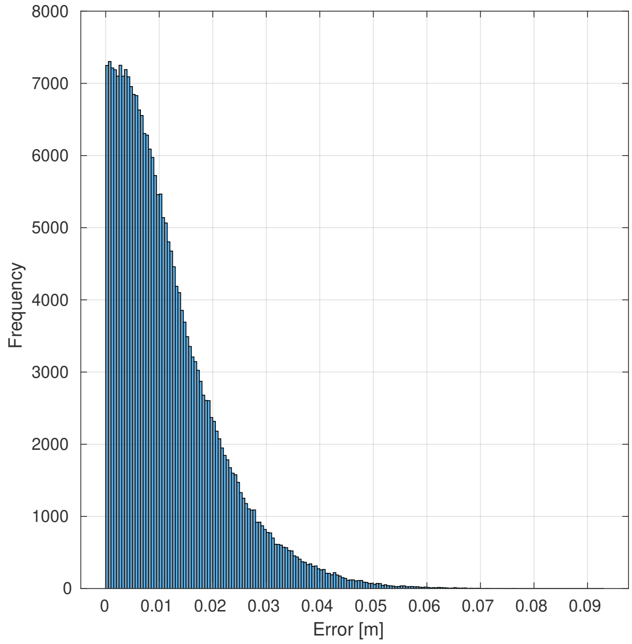


Figure 3.14: *In situ* laser line calibration error histogram with $\mu = 0.0116\,m$ and $\sigma = 0.01\,m$

Z direction. Secondly, the *insitu* calibration can be established as a routine prior to any dive, whereas the first method relies on a calibration performed in a controlled environment and assuming that the laser position or the deformation of its lens has not been altered from the day the calibration was performed to the actual mission day. Therefore, the *in situ* calibration method will account for any small movement of the laser line and for the deformation of its housing at depth. Moreover, the work of [19] has been extended to accommodate for different camera setups and compute the required camera positions and laser plane.

In the following chapter, a one-shot structured laser triangulation sensor is presented, using this same pipeline to reconstruct small objects or areas.

Chapter 4

One-shot reconstruction

In this chapter, a novel one-shot structured light sensor consisting of a camera and a laser projector is presented. The proposed projector casts a 25 parallel line pattern, obtained using a Diffractive Optical Element (DOE) after the laser collimator. The chapter is structured as follows: in section 4.1 the concept of one-shot is explained. In section 4.2 the methodology of this sensing device is shown. Section 4.4 shows three-dimensional reconstruction result, and finally in section 4.5 a discussion is presented.

4.1 Introduction

One-shot reconstruction enables sensing the three dimensional shape of the seafloor in one camera shot. Challenging tasks such as real-time manipulation or landing can be performed autonomously with the aid of a real-time three dimensional point cloud. The proposed sensor is formed by a camera and a 532 nm green laser. In front of the laser source, a DOE shapes the beam to a set of 25 parallel lines. These lines are projected on the underwater scene and recovered by the camera, where their projections are detected in the image, its peaks extracted and matched to their corresponding source laser plane. In figure 4.1 the captured pattern is shown. Using the same principle explained in the previous chapter, a three dimensional cloud is computed from 2D point to plane associations.

4.2 Methodology

The process to obtain 3D information from a frame is split in four steps: acquisition, segmentation, decoding and triangulation.

4.2.1 Acquisition

The acquisition is a process where a camera grabs or acquires a still frame of its field of view. In the context of laser-based structured light, a laser pattern, e.g. a laser line or any other shape obtainable using a DOE, is projected into the seafloor and the deformed pattern recovered in the image. The DOE diffractive behaviour weakens the light power in a very characteristic way. The central part of the DOE is brighter than the rest, due to the inherent

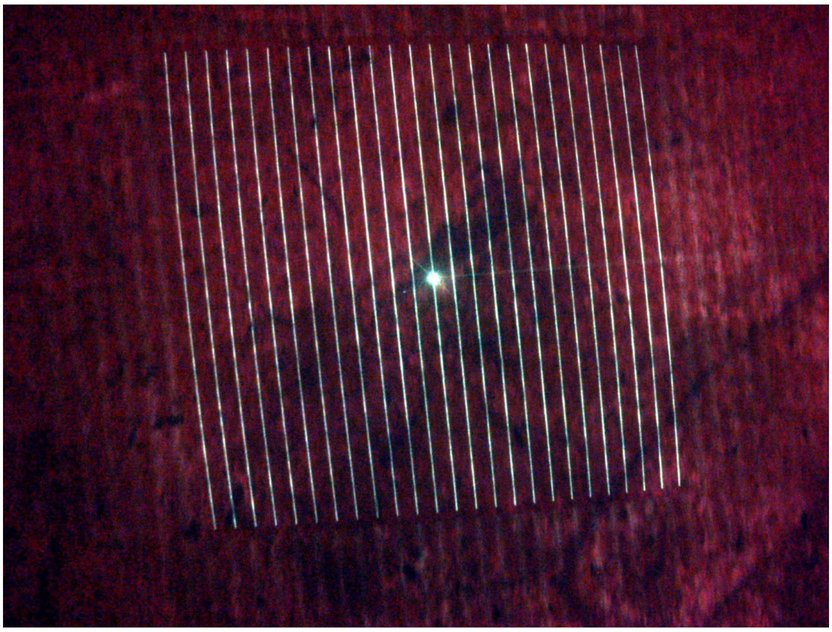


Figure 4.1: 25 parallel line laser pattern projected on an underwater cement wall.

original laser beam shape. To correctly detect the Gaussian shape of the lines, the exposure of the camera is configured to just saturate the pixels of central laser dot.

The DOE units are commonly available for red lasers at a different wavelength of 635 nm. A change in wavelength changes the refractive index, changing not only the field of view of the pattern, but also causes ghosting (e.g. the pattern is repeated in a grid-like shape reducing overall contrast. In figure 4.2 a central-dot row and a common row are shown with their intensity values.

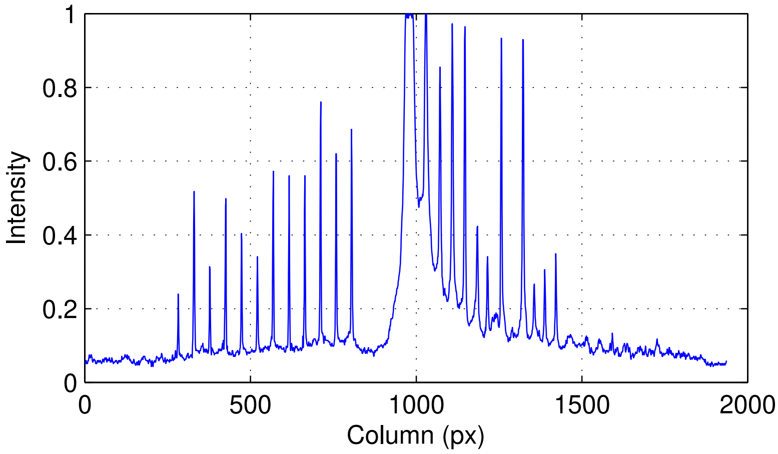
4.2.2 Segmentation

To ease the detection, the background illumination is removed by subtracting the red channel to the green. Then each row is convolved with a median filter, and the result is removed from the original signal. Therefore, the median filter is used as a low pass filter to normalize the intensity in the image, without altering the laser lines. Next, a binary image is computed by simple thresholding of the previous result, and the line centres are found for each row of the image. These steps are shown in figure 4.3.

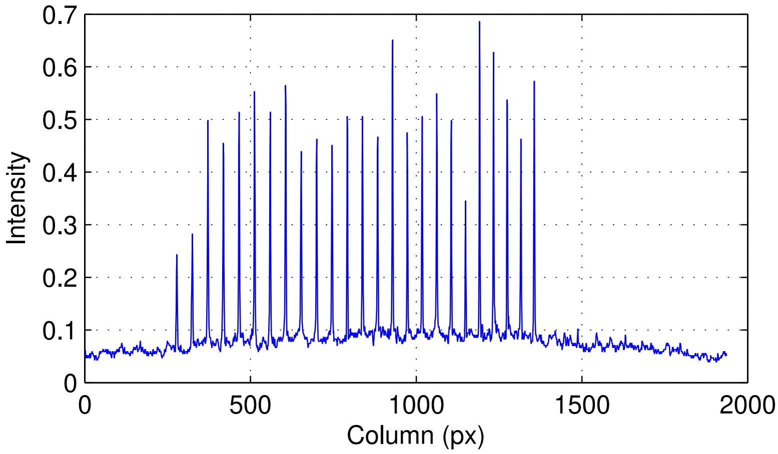
For each centre, the neighbouring values at the original image are checked, and the peaks are found using the peak detection explained in the previous chapter.

4.2.3 Decoding using Maximum Spanning Tree

The pattern used in this structured light system poses a challenge to determine which detected laser points in the image belong to which laser plane of the DOE pattern. To solve this



(a) Central dot row.



(b) Other rows.

Figure 4.2: Intensity values of the rows in the image. Note the saturated value at the central dot.

decoding problem, a relationship between pixels and planes has been proposed. In [118], the authors solve a similar problem where an RGB projector casts parallel lines on a scene using Maximum Spanning Trees (MST). The methodology from the authors has been used in the novel field of laser projection to decode laser planes.

This method is based on the epipolar constraint: given a set laser planes numbered $\pi_1, \pi_2, \ldots, \pi_n$ that project into $l_1, l_2, \ldots, l_m$ laser stripes, if a perpendicular line crossing the stripes is drawn, the crossing points will belong to monolithically increasing indexes of stripes, as long as there are no floating objects between the camera and the background. This supposition is true for underwater environments where the seafloor can be considered smooth and the presence of small objects casting a shadow or partially blocking the view of one of the stripes. These situations are depicted in figure 4.4.

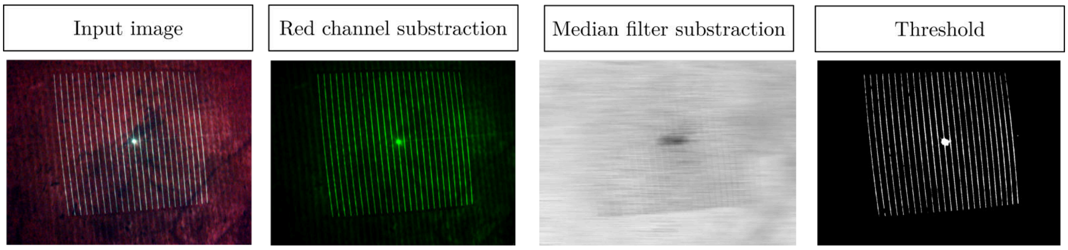


Figure 4.3: Segmentation steps to obtain a clean thresholded image.

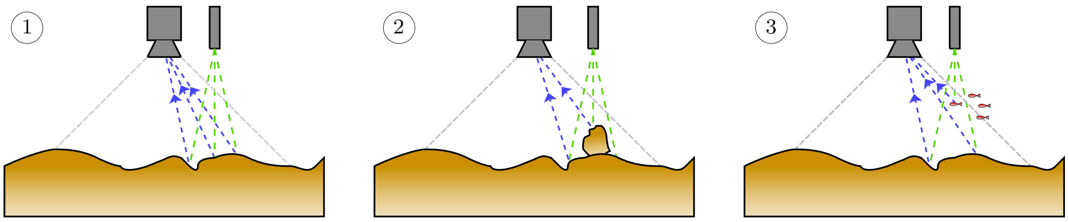
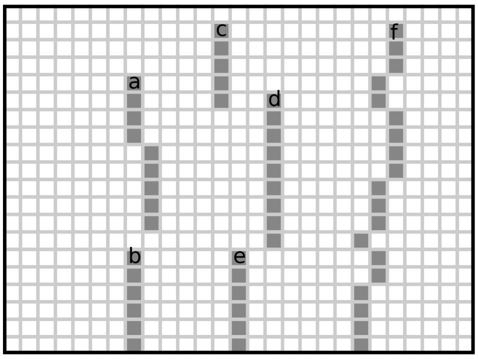


Figure 4.4: Three different situations for decoding: (1) presents a smooth seafloor and all laser lines are seen by the camera; (2) presents a rock blocking the view of one of the stripes; and (3) shows an unlikely situation where the stripe indexing is no longer increasing. The centre plane will be seen by the camera as the last stripe.

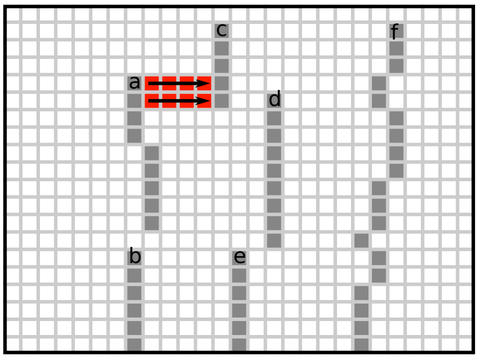
Let 25 parallel lines be defined as 25 groups of indexes or labels $k = \{1, 2, \ldots, 25\}$. Let $\boldsymbol{P}_l = \{\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_m\}$ a group of pixels who share at least one corner. We define $\boldsymbol{p}_a$ a neighbour pixel of $\boldsymbol{p}_b$ if there exists one or more rows of $\boldsymbol{p}_a$ shared with $\boldsymbol{p}_b$ without any other detected laser peak between them. The $\boldsymbol{P}_l$ groups of pixels can then be drawn as nodes in a directed graph $\mathcal{G}$, whose edge weight equals the number of common rows. This directed graph is the input of a MST algorithm. The resulting simplified directed graph is then indexed as follows: the node that does not have any parent is indexed as index 1. Then, the graph is traversed and its indexing increased when an edge is followed from parents to children. This yields an index for every connected vertex. An example of this approach can be seen in figure 4.5. In our application, the pattern has a central dot which belongs to the central line (e.g. index 13). The node belonging to that dot is labelled as $k = 13$ and the indexing occurs traversing the graph forwards and backwards.

4.2.4 Triangulation

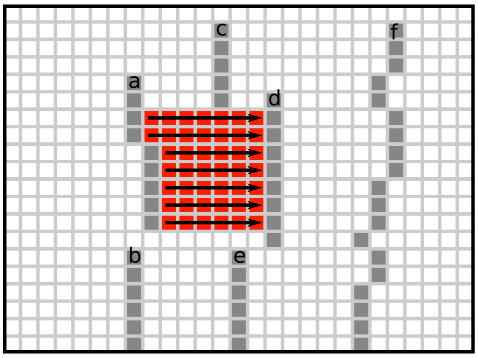
With the labelling and calibration, each 3D point $\boldsymbol{p}(t)$ can be computed by triangulating its corresponding laser plane π_n to the line formed by joining the segmented pixel to the camera focal point, which depends on the scale factor t.



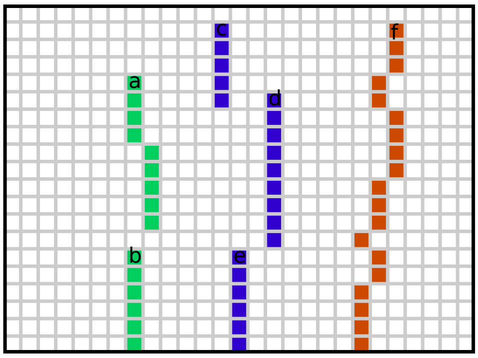
(a) A three-stripe pattern is detected and pixel groups are labeled as a, b, c, d, e, f.



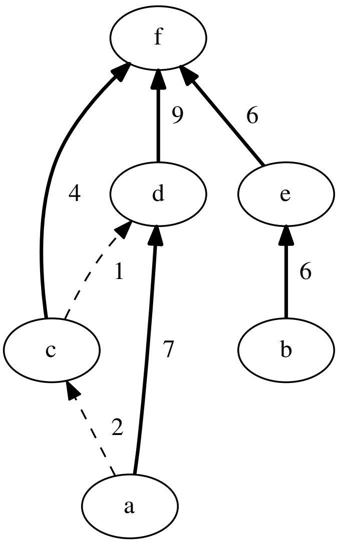
(b) Group a is a neighbour of group c as they share two rows.



(c) Group a is also neighbour of group d as they share 7 rows.



(d) After MST, the pixel groups have been correctly decoded.



(e) Tree view of the example. The weights at every edge are the number of rows shared to their neighbour. Solid edges are the MST solution.

Figure 4.5: Maximum Spanning Tree decoding example.

$$\begin{aligned}
\pi_n &: \quad Ax + By + Cz + D = 0 && (4.1)\\
\boldsymbol{p}(t) &= \left(\frac{u - c_x}{f_x}\, t, \frac{v - c_y}{f_y}\, t, t\right) && (4.2)\\
t &= \frac{-D}{A\frac{u-c_x}{f_x} + B\frac{v-c_y}{f_y} + C} && (4.3)
\end{aligned}$$

where (f_x, fy) is the camera focal length in x and y axes. (c_x, c_y) is the central pixel in the image. (u, v) is the detected laser peak pixel in the image. Replacing 4.3 in 4.2, the 3D coordinates of the point are obtained.

4.2.5 Geometric calibration

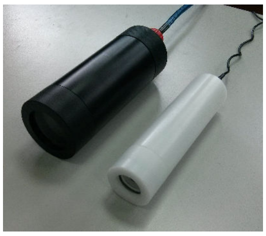
Plane fitting calibration is performed by registering all 3D line points and matching them to their corresponding laser planes. In order to do that, a set of images at different depths are captured, projecting the laser on a flat surface with a calibration pattern on it. Then the laser is detected in the images and ray traced to the corresponding calibration plane. This step is done in every captured frame. After each line has been detected in all images, every line point belonging to the same laser plane is used to fit a 3D plane using a least square approach. The coefficients of those planes are then saved for a posterior triangulation.

Essentially, the calibration of this sensor does not differ much from the previously explained calibration for one laser stripe. For a 25 parallel lines pattern, there are 25 different laser planes to fit. Using the first method, the laser pattern has to be projected into a calibrated plane and the two dimensional projections detected. Using the MST decoding, the points are labelled according to their belonging line. All points are grouped using the line index, triangulated in 3D using the calibrated plane and then a plane equation is fit for each of them.

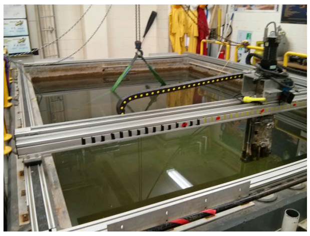
4.3 Experimental setup

A prototype of this LbSLS has been built using a colour camera and a laser pointer with a fitted DOE. The camera is a Manta G-283C from Allied Vision Technologies with a 12 mm optics, a CCD sensor of 1936×1458 pixels running at 20 fps. The laser is a 5 mW ZM-18B green laser from Z-Laser. The projected pattern is formed by 25 parallel lines, inscribed in a perfect square with a field of view of 21° in air, 17° in water. The pattern also has a brighter dot in the centre of the square, due to the direct transmission of the original laser beam through the DOE. These two components have been placed with an approximate baseline of 20 cm, tilting the laser 10° towards the camera.

In figure 4.6(a) the proposed system is shown unassembled, and in figure 4.6(b), it is shown assembled in a pool on a Cartesian robot to perform the experiments. These two



(a) Camera (black) and laser (white) housed.



(b) Cartesian plotter and pool.

Figure 4.6: Experimental setup held at Ocean Systems Laboratory - Heriot Watt University (Scotland, UK).

housings were designed to be mounted on an Autonomous Underwater Vehicle (AUV) for future development.

4.4 Results

Three experimental setups are described below to validate the system. First, an object reconstruction test to validate the ability of the LbSLS to pick 3D objects and their shape. Then, a study on the behaviour of the sensor under different turbidity scenarios is shown. Lastly, a comparison of LbSLS with a stereo rig is proposed during a lawnmower pattern survey.

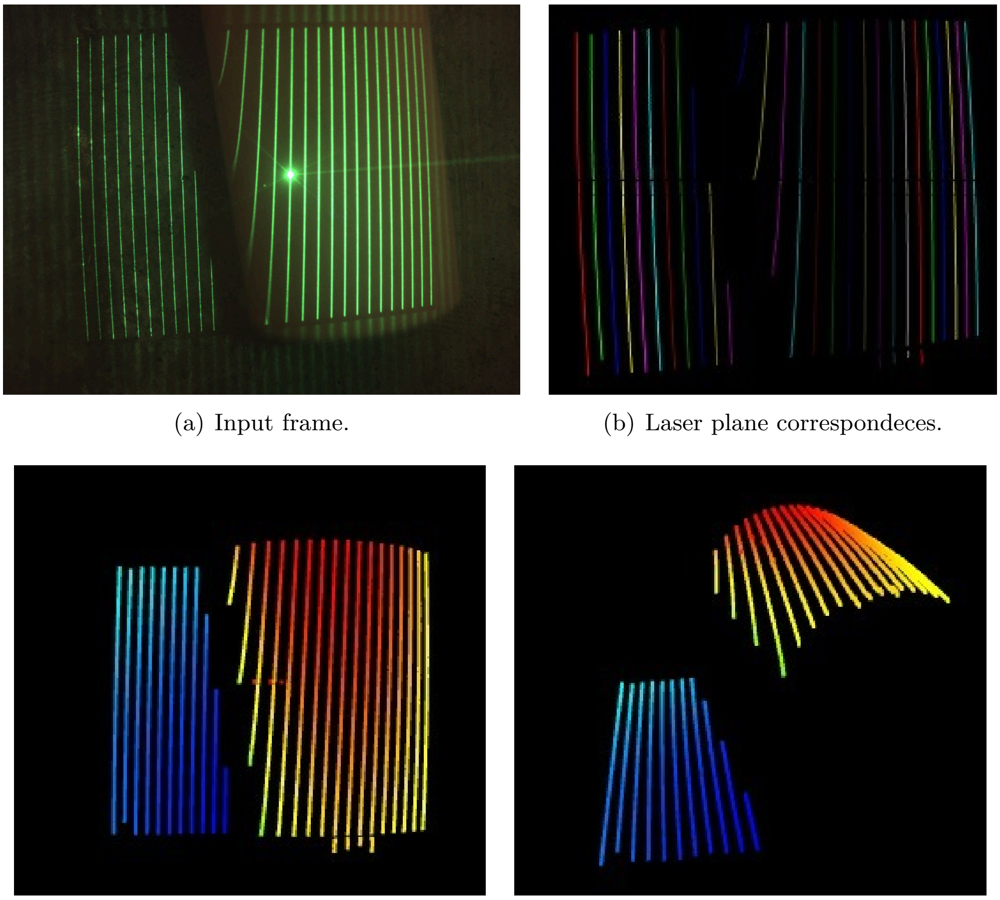
4.4.1 Object Reconstruction

The calibration of the system has been made with eight different frames, taken from different view angles and distances to the checkerboard plane. The output of the calibration has confirmed the angle between laser light planes to be $0.6875° \approx 17°/25$.

Two reconstruction experiments have been carried out. In the first one, a 16 *cm* diameter textureless plastic pipe, Fig. 4.7(a), and in the second, a 15 *cm* plastic weight plate, Fig. 4.16(a).

In both experiments, the correspondence output from the decoding stage can be seen, correspondingly, in Fig. 4.7(b) and 4.16(b), where each line has been drawn in a different colour. In the rows near the area where the central beam hits the target can be seen that the scattering and the light reflectance produces small inconsistencies in the correspondence solving.

Both reconstructions closely reproduce the original geometry. From the point cloud, the pipe roughly measures 13 *cm* width within the visible silhouette and the plate measures 14 *cm* in diameter.



(a) Input frame. (b) Laser plane correspondeces.

(c) Front view. (d) Lateral view.

Figure 4.7: 3D reconstruction of a pipe.

4.4.2 Turbidity

A second experimental setup is used to study how turbidity affects the LbSLS. The system has been deployed in a 125 l, 1.2 $\times$ 0.35 $\times$ 0.35 m water tank and the same scene has been reconstructed in nine different turbidity conditions. In figure 4.8 the setup is depicted. The scene consists of a textureless white bottle at the back placed approximately at 0.7 m from the camera, some stacked tiles and a brown jar on top, approximately at $0,5$ m from the camera.

A conventional camera or stereo system may find enough features in the tiles or the jar to perform any kind of feature matching, but not on a textureless object such as the bottle. Furthermore, with the addition of turbidity, the number of detectable features decreases making even more difficult to extract keypoints.

The camera and the laser have been fixed together at an angle so that the projecting pattern can be seen by the camera from 0.5 m to 1 m due to the water tank dimensions. Once fixed and deployed, a calibration has been performed.

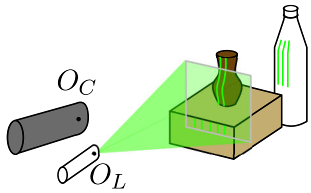


Figure 4.8: Experimental setup. The camera and the laser are deployed in a water tank, pointing to a jar and a bottle. O_C is the coordinate origin of the camera, and O_L is the origin of the laser. The transform between these two coordinates frames needs to be calibrated to obtain a valid 3D reconstruction.

Table 4.1: Number of 3D points detected by the system at different turbidity levels.

Experiment	Turbidity (ml)	Turbidity (%)	3D points
1	0	0	13,202
2	5	1/250	12,661
3	10	1/125	13,516
4	15	3/250	13,315
5	20	2/125	13,062
6	25	1/50	13,013
7	30	3/125	9,882
8	35	7/250	5,600
9	45	9/250	77

Turbidity has been obtained by pouring small quantities of whole milk into the water tank and then the mix has been stirred. Nine different milk volumes, starting from 5 ml up to 45 ml in steps of 5 ml have been measured.

Using the setup depicted in figure 4.8, and nine different turbidity levels, the number of 3D points reported by the system are shown in table 4.1. The number of points do not change until a high turbidity value is reached. Then the number of detected points falls until there is almost no detection at all. That is happening because the sensor is not able to discern the laser from the background, due to the scattered light.

In figure 4.9 experiments 1, 4, 7 and 9 are shown together with the detected points and the triangulated 3D points.

The recovered 3D shape is not affected by turbidity. The geometry of the scene is clear and the laser line segments remain the same throughout low to medium turbidity experiments. Although some points are missing, most of them are missed due to a very steep angle between the scene surface and the projection, causing the line segments to be very thin when projected in the camera image. Besides, the surfaces whose orientation is similar to the heading of the camera do not get affected by pollution until the image loses contrast. As expected, the jar and the bottle have poor contrast and lose colour in the presence of scattering.

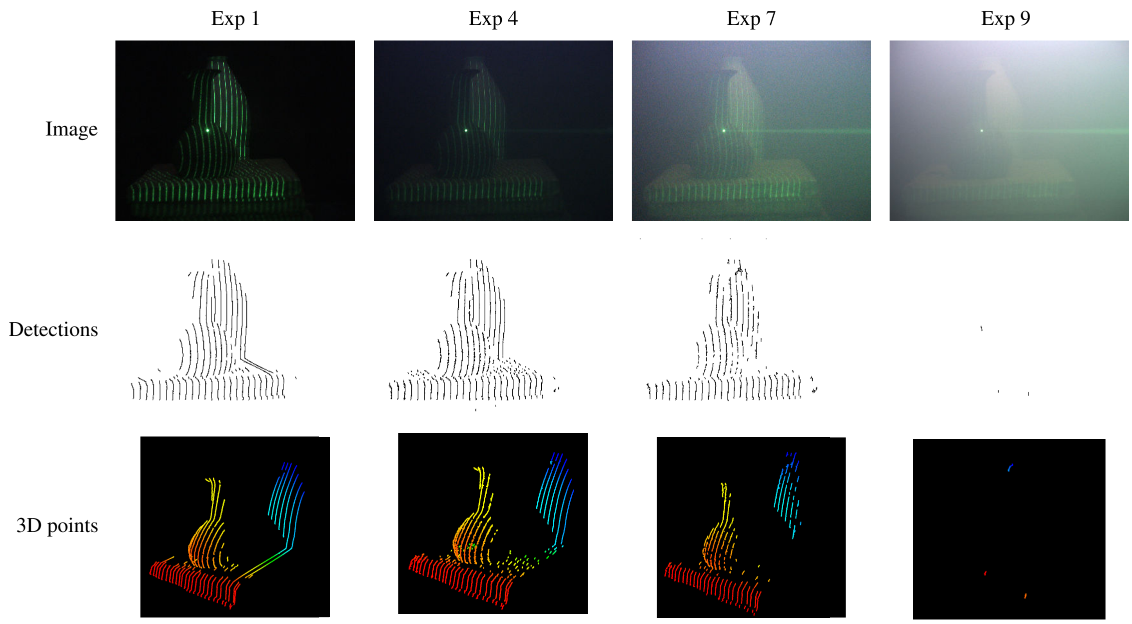


Figure 4.9: Laser images, detections and 3D points for different milk concentrations. The 3D point cloud has been rotated and it is presented in a isometric view, similar to the one in figure 4.8. For better quality, please refer to the digital version.

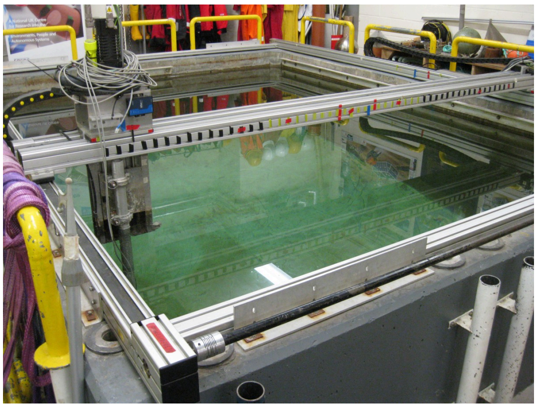
4.4.3 Survey

Finally, the LbSLS is compared to an underwater stereo camera for a small area survey reconstruction. The sensors were mounted, one at a time, on a Cartesian robot and performed a 3 by 2 m lawn-moving survey while recording the imagery. In the case of the LbSLS, the sensor was mounted at 0.7 m of the pool bottom and the spacing between lines was 0.2 m. For the stereo camera, the spacing was 0.5 m and was mounted at 1.3 m.

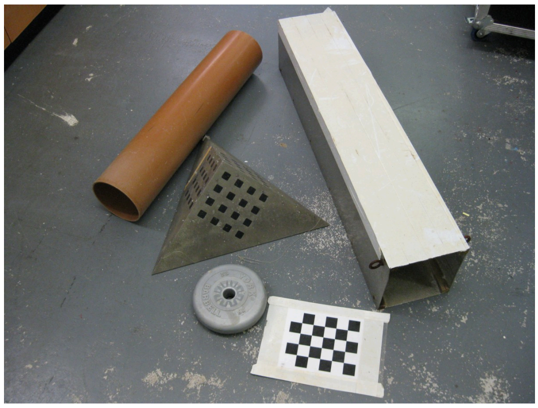
The pool is 4 meters long by 3 meters wide and 2 meters deep, and the Cartesian robot is mounted around the framing of it. It is able to carry an object on its mounting bracket and translate it in space without changing its orientation. This robot also is able to provide a feedback position of the carried element in real time, treated as ground truth.

The stereo camera is an off-the-shelf *FireWire* stereo camera *Bumblebee2* from *Point Grey Research*. This stereo rig includes two 1024×768 px CCD color cameras. The stereo rig is mounted in a bespoke underwater housing with a flat optical port. Its focal length is 3.8 mm. At every frame, 3D points are computed using pixel disparities and colour. For consecutive frames, the 3D points have been filtered using a voxel grid filter, and finally concatenated using the pose of the Cartesian robot. If there is not enough texture in the images, the disparity image will show a blank area, where a stereo correspondence was missed. Thus, the generated point cloud will also miss some points.

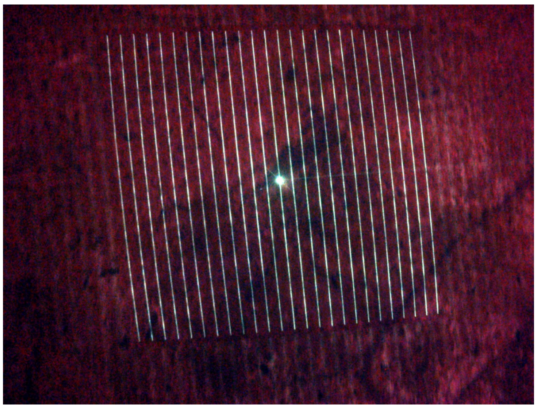
The LbSLS was attached using a pan and tilt unit to the robot. The pan and tilt unit helped to perform a better calibration of the system beforehand and finally the camera was set



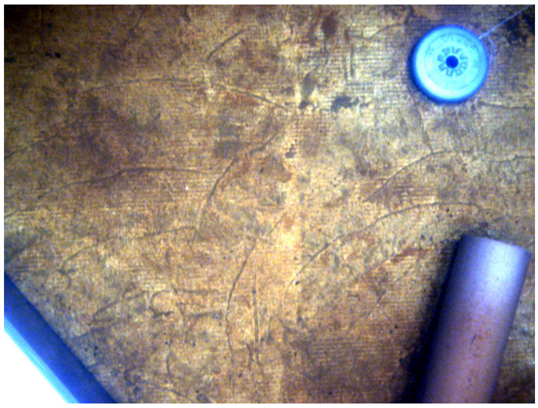
(a) Pool at Heriot Watt University with the cartesian robot on top.



(b) Objects used to verify the 3D reconstruction accuracy.



(c) Image frame from the laser dataset.



(d) Left image frame from the stereo dataset.

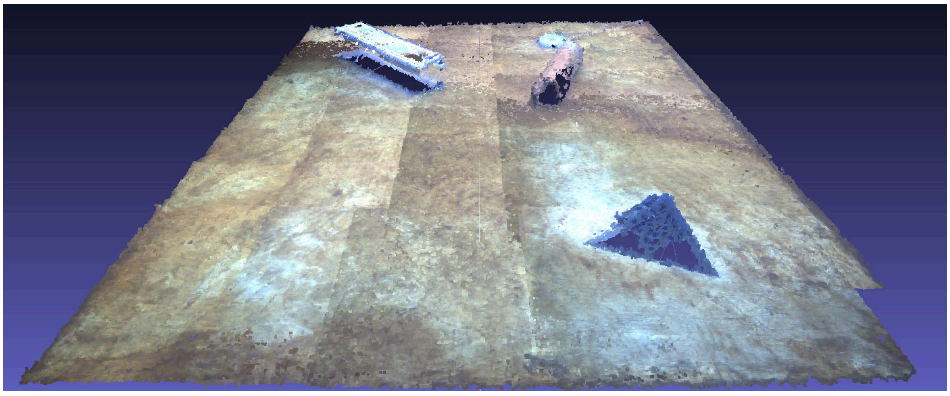
Figure 4.10: Experimental setup. The objects shown were deployed in the pool and the sensors shown were moved in a lanw-moving pattern survey with a cartesian robot.

perpendicular to the floor prior to perform the survey. However, this unit has a considerable weight that increased the inertia of the robot and slightly overshot the robot's position. Whereas the stereo camera was simply attached using a lightweight pole.

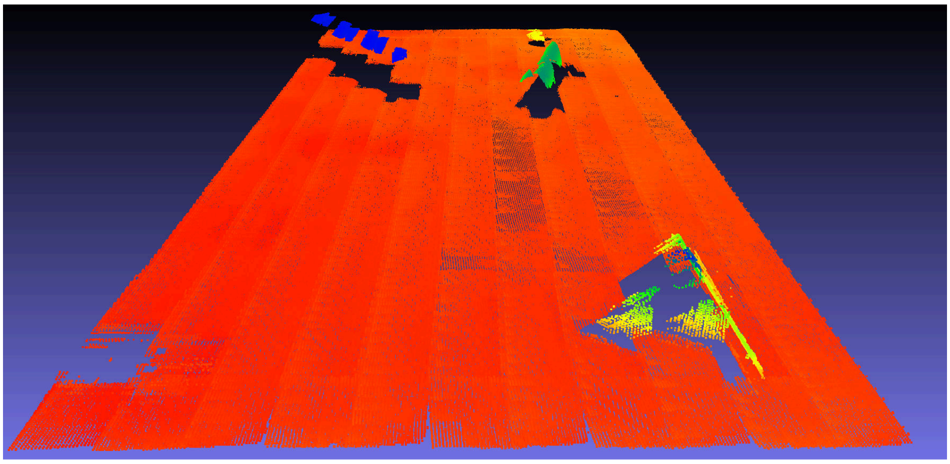
Four different objects were deployed in the pool to test the 3D reconstruction, as explained in section 4.4.4.

The pool, the cartesian robot, and the objects can be seen in figures 4.10(a) and 4.10(b). In figures 4.10(c) and 4.10(d) one frame from each dataset are shown.

The resulting three-dimensional reconstructions can be seen in figures 4.11(a) and 4.11(b). Note that for structured light the pool lights had to be dimmed to sense the laser, thus no RGB data has been captured. Depth is represented in a colour scale, being blue closer and red farther from the camera.



(a) Stereo reconstruction. 685872 points.



(b) Structured light reconstruction. Blue is closer and red is farther from the camera. 370261 points.

Figure 4.11: Sensor 3D reconstructions, available at http://srv.uib.es/pointclouds

Table 4.2: Object measurements for the different sensors. GT means Ground Truth, measurements of the object using a ruler or a caliper.

Object	GT (m)	Laser (m)	Stereo (m)
Square pipe ($L \times H \times W$)	$1 \times 0.2 \times 0.2$	$1.038 \times 0.2102 \times 0.2396$	$1.034 \times 0.1967 \times 0.2113$
Round pipe ($D_{ext} \times L$)	$\varnothing 0.16 \times 0.75$	$\varnothing 0.16 \times 0.8390$	$\varnothing 0.1651 \times 0.7542$
Wheel ($D_{ext} \times D_{int} \times H$)	$\varnothing 0.20 \times \varnothing 0.03 \times 0.04$	$\varnothing 0.1886 \times \varnothing 0.045 \times 0.0511$	$\varnothing 0.1973 \times \varnothing 0.024 \times 0.0630$
Pyramid ($L \times H$)	0.51×0.207	0.5020×0.2064	0.517×0.2186

4.4.4 Object dimensions and accuracy

In the pool, four objects with known dimensions were dropped to check the accuracy of both systems. These objects were a 1 m long 0.2×0.2 m square pipe, a $\varnothing 0.16$ by 0.75 m round pipe, a $\varnothing 0.2$ by 0.04 m lifting wheel with a $\varnothing 0.03$ m central hole in it, and a 0.207 m tall triangular pyramid with an isosceles basis, whose side length is 0.51 m. In figure 4.12, a colour mosaic of the pool floor can be seen, with the objects placed in their reconstructed positions. In table 4.2 the true measurements from these objects and the measurements extracted from both pointclouds are shown.

The obtained measurements from the point cloud have been manually measured using CloudCompare software [119].

4.4.5 Stereoscopy Evaluation

Finally, the two point clouds have been registered using ICP and then gridded to 7 mm resolution to compare the distance from one point cloud to the other. As the stereo point cloud has a higher number of points, it has been taken as the reference model, and the laser point cloud as the test model. Therefore, the distances were computed from the points in the laser dataset to the closest point in the stereo.

In figure 4.13 the distance for each laser point to the closest stereo point is shown. In figures 4.14(a) to 4.14(d) close-up views of the objects can be seen. Note that if there are no stereo points in one area but laser, the error for the laser points will grow even if the points are correctly located. Otherwise, if there are stereo points but no laser points, there is no error. Note this behaviour in figure 4.14(a).

In figure 4.15 the frequency plot of the distance error is shown. A Gaussian distribution has been fitted with a mean of 3.5 mm, and a standard deviation of 1.2 mm. The long right side tail has been split at 0.01 m in figure 4.15(b). The number of bins is the same as in figure 4.15(a), but the vertical scale has been resized to better show the distance distribution.

4.5 Discussion

This system has proved to be capable of performing a one-shot reconstruction. The output 3D data can be used to find objects or to match them to a known object database. The

Figure 4.12: Pool mosaic built from a subset of 40 images (left camera of stereo dataset). From [7].

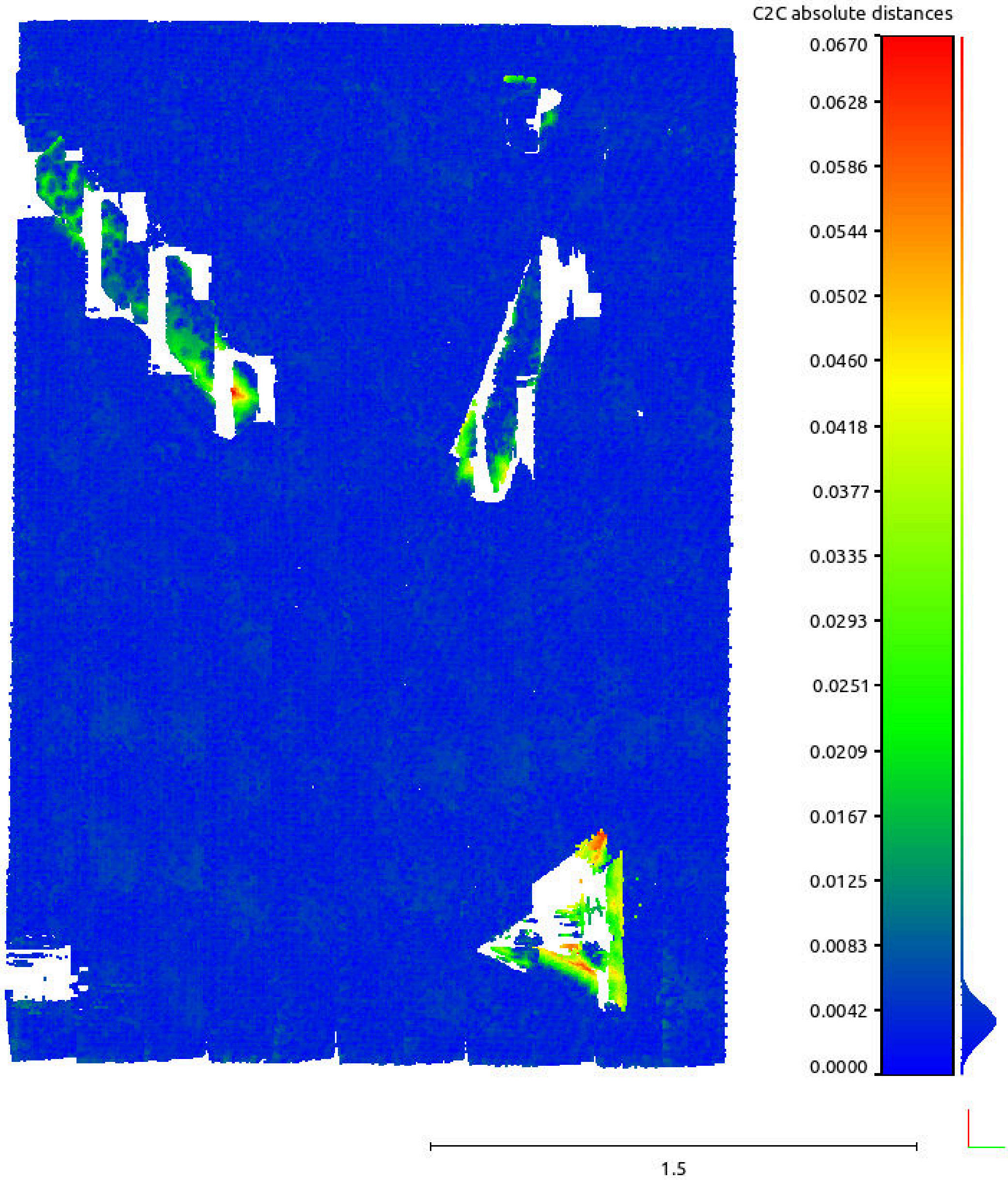
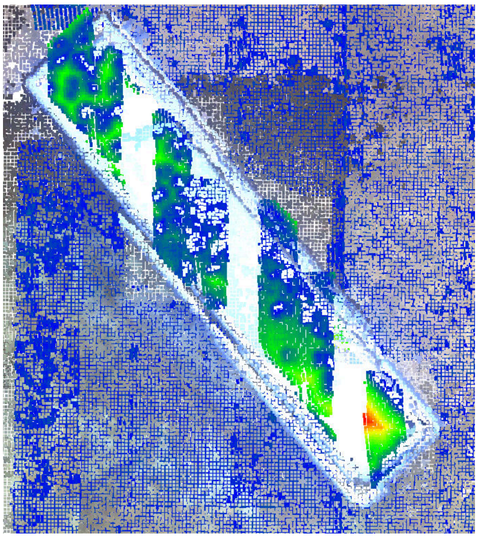
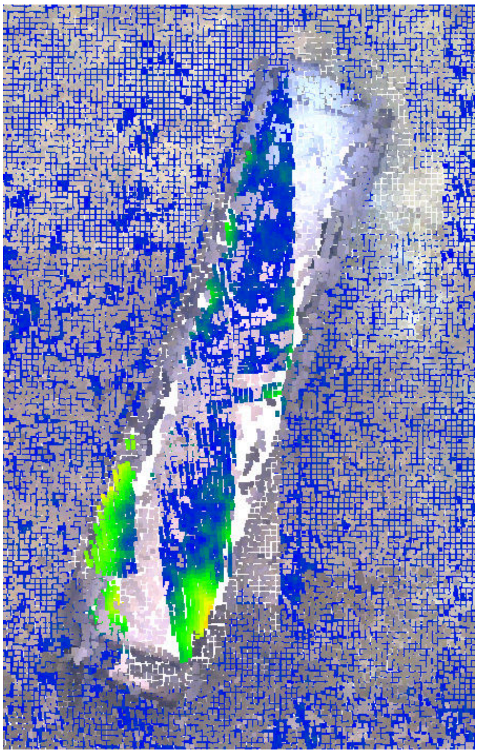


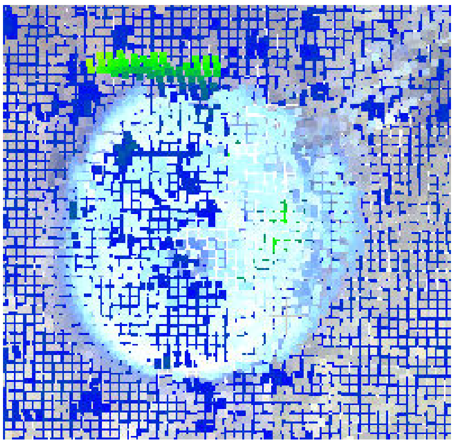
Figure 4.13: Distance from laser point cloud to stereo point cloud. Measurements in meters. Deep blue is almost zero error whilst green to red colour transition means higher error.



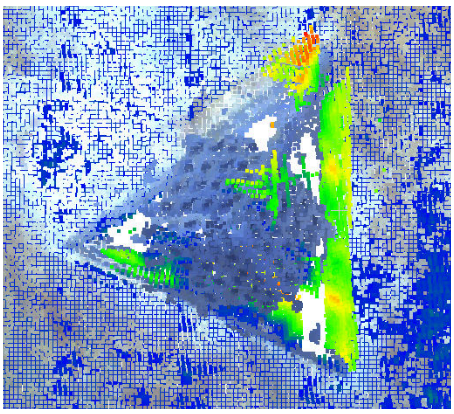
(a) Square pipe



(b) Round pipe

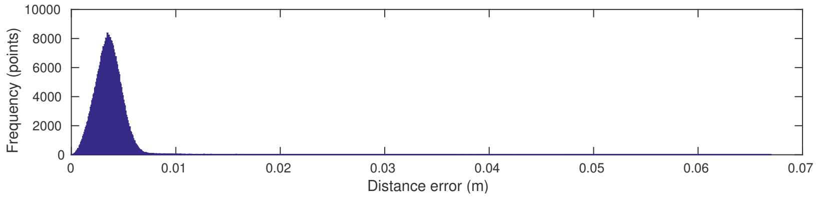


(c) Wheel

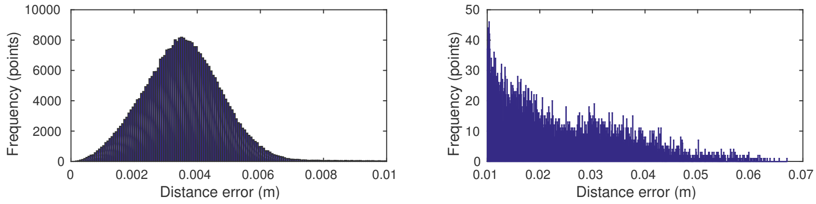


(d) Pyramid

Figure 4.14: Object distance close-ups. Deep blue is almost zero error whilst green to red colour transition means higher error.



(a) Distance error containing all the laser points.



(b) Distance error less (left) or greater (right) than 0.01 m.

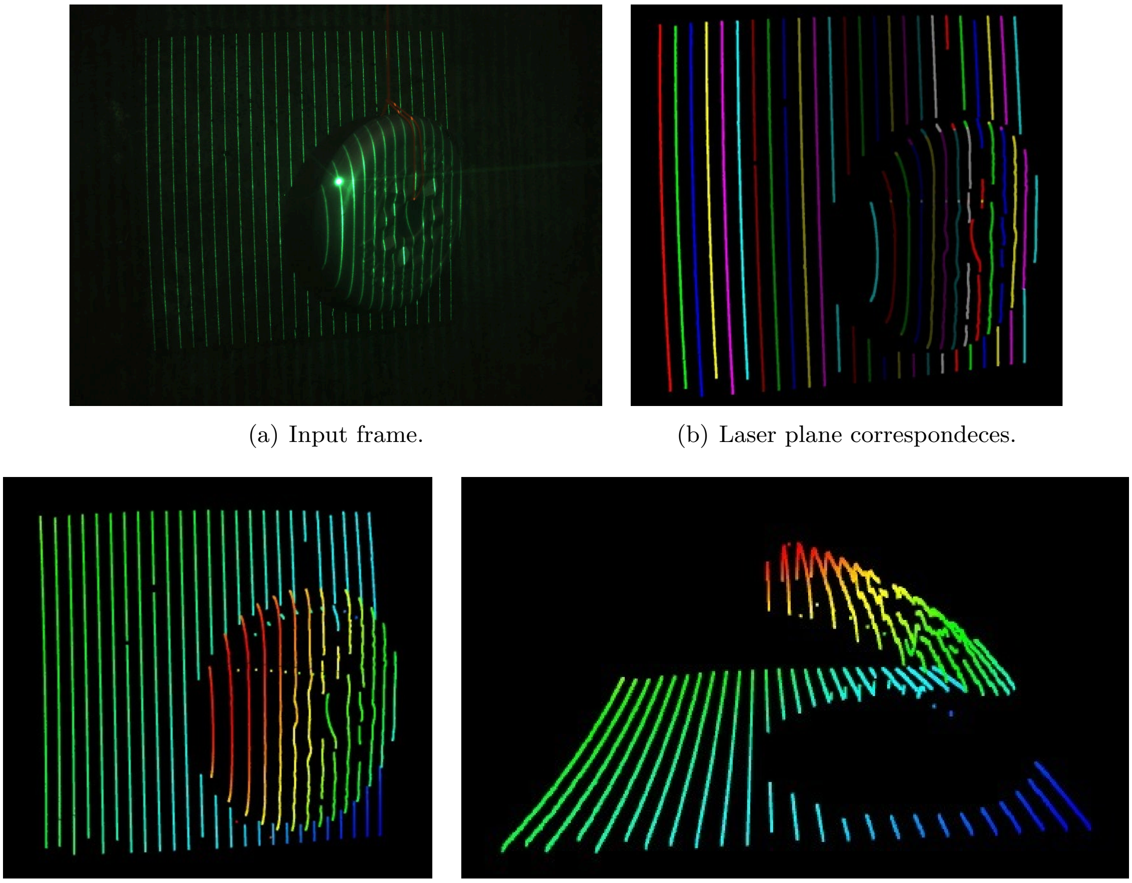
Figure 4.15: Distance error frequency plots.

correct detection of the laser depends on the contrast between the lines and the background. Low ambient light scenarios have proven best at obtaining one-shot reconstructions. The major drawback has been the low laser intensity. Fainted laser lines pose challenges for the segmentation to discern laser lines from background. Even so, the system is designed to operate at depth where no ambient light is present and artificial light can be controlled or dimmed. Therefore, there is a relationship between successful scans, laser light power (signal) and ambient light (noise). Treated as a standard signal processing framework, the signal to noise ratio has to be such that it permits correct scans. The brightness of the laser has to be clear and sharp when compared to the background light / scene. In the datasets presented in this PhD, a 20% intensity increase between laser lines and background have been successful.

The results at comparing the two point clouds are promising for the LbSLS. Notice how in table 4.2 the z dimension accuracy is better for the LbSLS than for the stereo. However, in the other two dimensions this relationship is not so clear. This may have been caused due to the weight of the pan and tilt unit that carried the sensor. Its heavy weight increased the inertia of the system, overshot it, and caused an offset from the sensor position to the Cartesian robot. This can be noted in the square pipe (figure 4.14(a)) when we compare the forward movement reconstruction to the backward movement. The four sensor transects can be paired 2 by 2, as there are 2 made in forward movement and 2 in backward. This happens throughout the course of the robot in the pool.

Regarding the comparison of the two point clouds, the distance error mean is a 0.51% of the range, and although there are points with a large error > 1 cm, their contribution is a 1.8% (6750 over 370261 points).

With these results, we can conclude that LbSLS and stereo systems can be used for 3D reconstruction taking into account several particularities. With a stereo camera one can



(a) Input frame. (b) Laser plane correspondeces.

(c) Front view. (d) Lateral view.

Figure 4.16: 3D reconstruction of a 1 kg plate.

get a good general overview of the underwater scenario, missing small details and accurate distances, whilst with LbSLS one can get sharper and clearer details, at the cost of losing colour information.

Note the top of the square pipe in the stereo dataset. There are no points due to a lack of texture. The pipe is plain white and the stereo block matching algorithm is unable to find any correspondence between left and right frames.

As a result, for long, high altitude surveys where there is enough texture and visibility, stereo data proves enough to recover the overall shape of the underwater environment. But prior to a manipulation where precise measurements and distances are desired, a sparse LbSLS with a limited field of view can provide the information required.

In the next chapter, a high altitude laser mapping system is presented and two bathymetric SLAM solutions applied to achieve self-consistent seafloor maps.

Chapter 5

Bathymetric SLAM

In this chapter, two Terrain Based Navigation solutions for underwater vehicles using a laser stripe in unconfined environments are presented and compared. First we propose a submap graph SLAM method that uses an ICP registration to improve the consistency of the bathymetric maps in section 5.2, and in section 5.2 a grid-based bathymetric SLAM is proposed to correct the final navigation estimate independently of submaps. In section 5.4 the results of both approaches are compared. Both methods achieve self-consistent maps and are capable of reconstruction featureless terrains.

5.1 Introduction

Bathymetric maps represent the height of the earth's surface underwater within a given geographical region. These kind of models can be generated by several methods, depending on the resolution and coverage desired. To provide bathymetric maps there are two options to mount a sensor, either shipborne multibeam or a sensor mounted on Unmanned Underwater Vehicles (UUVs). Ships are restricted to operating at the ocean's surface and so provided maps resolution decrease as seabed depth increases. Wave induced motion is also an issue for these systems as small misalignments in attitude turn into significantly large errors in the northing/easting coordinate of the observation.

Bathymetric maps are traditionally built using a gridded or point cloud model of the seafloor with a deterministic model of the vehicle or vessel pose estimate. In this case it is sufficient to generate the map by estimating the depth at any given location with the mean of the depth measurements observed there. However, assuming a deterministic navigation solution introduces misalignment and inconsistencies in the map if the navigation is subject to errors such as drift or biases. As will be shown in this work, such map building exercises require an accurate, reliable source of localisation in order to guarantee the consistency of the map. In the absence of such information, the only statistically consistent course of action is to maintain an estimate of the correlation between the imperfect estimates of the vehicle position and the map.

SLAM allows misalignments such as the above mentioned to be resolved by correcting the navigation solution from which the map was generated from, and can be classified as either feature-based or featureless in its approach. Using laser bathymetry, this thesis has approached featureless bathymetric SLAM. The unstructured nature of the majority of the seafloor suggests that such an approach may be more appropriate for mapping. Navigation in unconfined environments (beyond the coverage of acoustic transponder networks) remains one of the milestone problems yet to be solved. Terrain based navigation [120], with so many implementations already reported in the literature has not found its path towards routinary application. Many SLAM implementations have been recently reported in the literature using bathymetries, Roman [121] proposed a submap representation of the SONAR measurements and treated them as delayed updates in a EKF filter, and Palomer [122] further extended it to 3D sonar. Barkby [8] explored a gridded map representation and applied it to a Rao-Backwellized Particle Filter to find the best consistent map, and Vallicrosa [123] extended it using Hilbert Maps applied to a 2D occupancy grid.

5.2 Submap Bathymetric SLAM

In this section, groups of laser lines are represented as submaps using the vehicle navigation. Then, these local maps are used in a submap SLAM framework based on [121] and a g^{2}o submap pose representation [124] to handle the global map optimization.

5.2.1 Graph representation

A pose-based graph SLAM is a graphical representation of the SLAM optimization problem where poses are represented as nodes and relative transformations as edges in a graph. Let $\boldsymbol{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T)^\top$ be a vector of parameters, where $\boldsymbol{x}_i$ describes the pose of node i. Let $\boldsymbol{z}_{ij}$ and $\boldsymbol{\Omega}_{ij}$ be the mean and the information matrix of a measurement between the node i and the node j. This measurement is a transformation that makes the observations obtained from i maximally overlap with the ones acquired from j. Let $\hat{\boldsymbol{z}}_{ij}(\boldsymbol{x}_i, \boldsymbol{x}_j)$ be the measurement between nodes i and j, which usually is the relative transformation between these two. The log-likelihood l_{ij} of a measurement z_{ij} is

$$\boldsymbol{l}_{ij} \propto \boldsymbol{e}_{ij}^\top \boldsymbol{\Omega}_{ij} \boldsymbol{e}_{ij} \tag{5.1}$$

where the error function that computes the difference between the expected and the real observation is expressed in (5.2).

$$\boldsymbol{e}_{ij} = \boldsymbol{e}_{ij}(x_i, x_j, z_{ij}) = \boldsymbol{z}_{ij} - (\ominus \boldsymbol{x_j} \oplus \boldsymbol{x_i}) \tag{5.2}$$

The information matrix can be computed as the inverse of the covariance matrix

$$\boldsymbol{\Omega}_{ij} = \boldsymbol{P}_{\boldsymbol{ij}}^{\boldsymbol{-1}}. \tag{5.3}$$

For simplicity, the indices of the measurements have been encoded in the indices of the error function. The equation to minimize is the negative log-likelihood l_{ij} to achieve the goal of a maximum likelihood approach.

$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}} \sum_{(i,j)\in\mathcal{C}} \boldsymbol{e}_{ij}^\top \boldsymbol{\Omega}_{ij} \boldsymbol{e}_{ij} \tag{5.4}$$

This minimisation problem is solved using g^2o general framework for graph optimization [124]. Essentially, is is using a LevenbergâĂŞMarquardt algorithm with a damping factor to Gauss-Newton non-linear optimization to control the convergence of the non-linear system. To utilize g^2o one simply has to define the error function and a procedure for applying a perturbation to the current solution. The minimization will find the arguments $\boldsymbol{x}$ that optimize the navigation given the input measurements.

5.2.2 Navigation

Underwater vehicles use a suite of sensors to obtain a dead-reckoning navigation solution. Most commonly, Doppler Velocity Log (DVL), Inertial Magnetic Units (IMU) and depth sensors are used. Acoustic positioning is also commonly used although its frequency and accuracy is still not comparable to ground-based solutions such as GPS. Therefore, sensor filtering and fusion such as Extended Kalman Filters must be used to obtain a better position estimate. In this work, we use a generic EKF implementation [125] with a 15-dimensional state vector formed by position, orientation, linear and angular speeds and linear acceleration, as shown in (5.5). Afterwards, EKF states have been back-propagated using a Extended Kalman Smoother (EKS) [126].

$$\mathbf{x_i} = \left(x, y, z, \phi, \theta, \psi, \dot{x}, \dot{y}, \dot{z}, \dot{\phi}, \dot{\varphi}, \dot{\psi}, \ddot{x}, \ddot{y}, \ddot{z}\right)^\intercal \tag{5.5}$$

The bathymetric submaps are created as the navigation progresses. Each submap contains a set of 3D points defined with respect to a local origin located at the central vehicle pose of the submap. This origin is calculated at every submap closure, when the submap is stored in a database for future registrations. The submap pose is encoded in a position and orientation with respect to the world origin, depicted by (5.6).

$$\mathbf{x_{s_k}} = (x, y, z, \phi, \theta, \psi)^\intercal \tag{5.6}$$

5.2.3 Submap generation

As explained in section 3.2, a laser stripe together with an imaging sensor can be used to triangulate two-dimensional points, detected in an image, to the three-dimensional world. This triangulation can be seen in figure 5.1.
Since laser stripe triangulation can only produce 2D reconstructions that lie on the laser plane

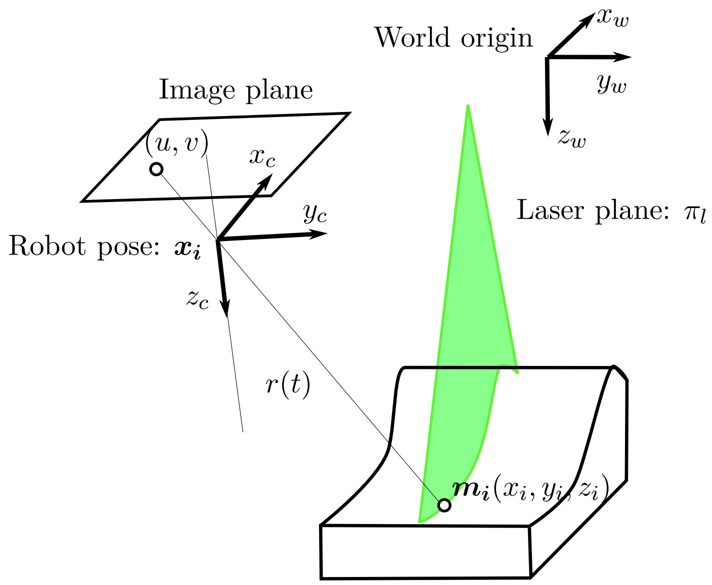


Figure 5.1: A laser plane π_l is projected into the seafloor and the plane-seafloor intersection is detected by a camera at pixel (u, v). For any laser point, the line that joins the 2D point with the camera focal point will intersect the laser plane π_l at $\boldsymbol{m_i}$.

$\boldsymbol{\pi_l}$, it is necessary to compound them with an estimate of the vehicle trajectory [127] to build a three dimensional surface patch. It is straightforward to compound the laser points with the trajectory: let $\boldsymbol{m_i} = \mathcal{N}(\boldsymbol{\hat{m}_i}, \boldsymbol{P_{m_i}})$ be one of the laser measurements already expressed in 3D coordinates (e.g. after triangulation), and $\boldsymbol{x_i} = \mathcal{N}(\boldsymbol{\hat{x}_i}, \boldsymbol{P_{x_i}})$ be the robot position at the time the image was acquired. Then the position and the uncertainty of one point $\boldsymbol{p_i} = \mathcal{N}(\boldsymbol{\hat{p}_i}, \boldsymbol{P_{p_i}})$ of the surface patch can be computed as:

$$\boldsymbol{\hat{p}_i} = \boldsymbol{\hat{x}}_i \oplus \boldsymbol{\hat{m}_i} \tag{5.7}$$

$$\boldsymbol{P_{p_i}} = \boldsymbol{J_{1\oplus}} \boldsymbol{P_{x_i}} \boldsymbol{J_{1\oplus}^\top} + \boldsymbol{J_{2\oplus}} \boldsymbol{P_{m_i}} \boldsymbol{J_{2\oplus}^\top} \tag{5.8}$$

where $\boldsymbol{J_\oplus} = [\boldsymbol{J_{1\oplus}}, \boldsymbol{J_{2\oplus}}]$ are the left and right halves (6×6) of the compounding Jacobian (6×12) [128].

5.2.3.1 Submap definition

A submap is defined as a set of 3D points belonging to different laser triangulations. Let $\boldsymbol{p^j} = (p_1^j, \ldots, p_i^j, \ldots, p_N^j)$ be the 3D points of the j-th pose expressed in a common frame, we define a submap or patch of size M as (5.9).

$$\boldsymbol{P_k} = \boldsymbol{p^j} \cup \cdots \cup \boldsymbol{p^{j+M}} \tag{5.9}$$

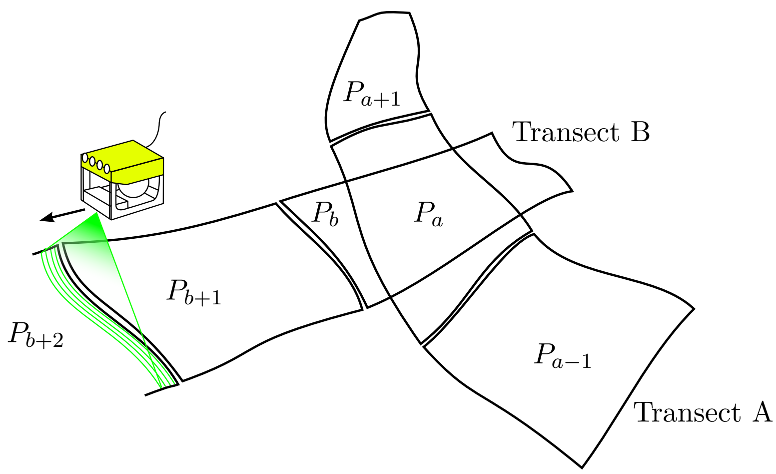
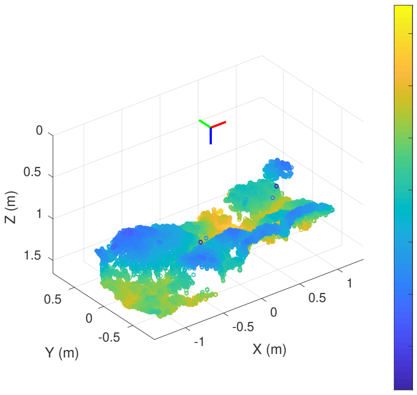


Figure 5.2: An underwater vehicle scans the seafloor using a laser stripe. The triangulated laser lines are grouped in patches depending on navigation uncertainty and patch size. Each submap is stored with known transformations for posterior optimization.

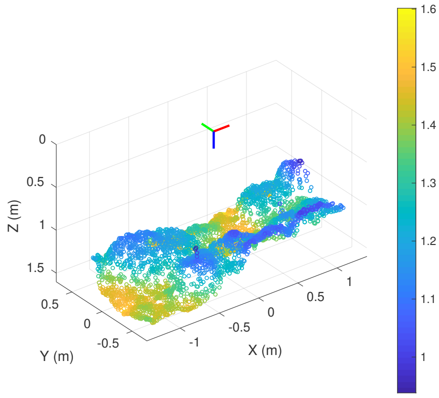
The creation over time of these submap generates discrete pointclouds that are represented as a whole rigid object, as seen in figure 5.2. The main assumption of this algorithm is that the accuracy of Dead Reckoning (DR) in a small map portion or patch is enough to represent the true world, locally. Without external ground referenced position measurements, the DR error would grow unbounded and the complete map would appear distorted. It is worth mentioning that this suggests that there may be an optimal submap size at which to break a terrain map and begin another. The selection of this break point has to take into account the following: (1) the patch will be considered rigid, therefore it has to be small enough so that the error committed by compounding DR navigation is negligible, and (2) the patch has to be large enough to contain enough three dimensional information to be unambiguously registered to another patch.

Thus we define two splitting conditions: (1) when the localization uncertainty of the last added line is bigger than a threshold and (2) when the patch is considered too big. Once the patch is closed, all the $\boldsymbol{p_i}$ points are referenced to a new coordinate frame coincident with the vehicle pose in the middle of the patch sequence to produce a more convenient uncertainty distribution among the points that form the patch [121].

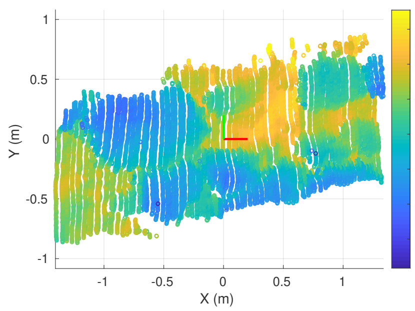
One characteristic of laser line bathymetry compounded with a movement is that depending on the forward motion (along track) the point density changes drastically. Therefore it is desirable to obtain a stable point density at every dataset. However, this is not practical for two reasons: (1) the scale may not be meaningful or (2) the time needed to slowly advance may not be suitable. Therefore it is usual to have a high resolution across track and a low resolution along track. This behaviour can be seen in figures 5.3(a) and 5.3(c), where the laser lines can be clearly distinguished from the pointcloud. There, the along track resolution is 3 cm whereas the across track is 1 mm.



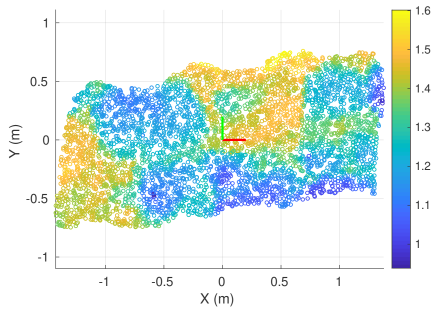
(a) Isometric view before random sampling.



(b) Isometric view after random sampling.



(c) Top view before random sampling.



(d) Top view after random sampling.

Figure 5.3: Example of random sampling over a patch for a transect in Valldemossa using Turbot AUV. The number of points is reduced from 80k to 5k without losing the overall seafloor shape. The sampling also helps ICP registration not to fall on local minima.

When it comes to point cloud registration, most metrics compute the point-to-point distance or point-to-plane distance. If the point density is not homogeneous, most algorithms will fall into local minima. To overcome this issue, each patch point cloud is triangulated using a Greedy Stepwise triangulation algorithm [129] in two dimensional space (e.g. x-y), and it is randomly sampled up to a fixed point density of 3000 points per square meter. This density is roughly a point at every 20 mm to represent high frequency surface roughness, when compared to large scale underwater features.

Each random sample is converted to three dimensional space, and its z coordinate is interpolated using the tree points that form the two dimensional triangle. This transformation from real 3D data to interpolated virtual points subsamples the 3D point structure enhancing the along track bathymetry shape. An example of this random downsampling can be seen in figures 5.3(b) and 5.3(d), where the across-track line shape is lost to favour a better point-to-point metric.

After a submap is closed, loop closure candidates are proposed. These possible relative pose links are found by looking in the 2D x-y plane for the patches that fall within a distance threshold. To set this threshold, navigation uncertainty is used. Then these point cloud candidates pass first an overlap test, where a k-Nearest Neighbour (kNN) finds the closest point of the newly closed point on every candidate cloud. Using $k = 1$ and computing the euclidean distance between point matches, an overlap metric is obtained. The clouds with higher overlap are more likely to be registered that the ones with few overlapping points.

5.2.3.2 Registration

Given two point clouds $\boldsymbol{P_a} = (\boldsymbol{p}_1^a, \cdots, \boldsymbol{p}_N^a)$ and $\boldsymbol{P_b} = (\boldsymbol{p}_1^b, \cdots, \boldsymbol{p}_M^b)$ with known poses with respect to a common origin, a transformation that registers these clouds is such that minimizes the point-to-point distance from corresponding point pairs, namely Loop Closure (LC).

$$\boldsymbol{\Delta_{ij}^{LC}} \leftarrow \arg\min_{T} \sum_i (\boldsymbol{p}_i^b - \boldsymbol{T}^* \boldsymbol{p}_i^a) \tag{5.10}$$

In this approach, Generalized Iterative Closest Point (GICP) [130] is used to minimize this transformation. The algorithm modifies the standard ICP solution with a probabilistic approach using point-to-plane association as well as point-to-point, taking advantage of surface normals. This algorithm needs an initial transformation estimate, which is obtained from the known navigation using the tail-to-tail transformation between both position estimations.

$$\hat{\boldsymbol{T}} = \ominus \boldsymbol{x}_a \oplus \boldsymbol{x}_b \tag{5.11}$$

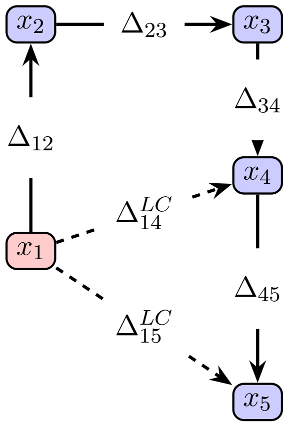


Figure 5.4: Example graph representation of 5 submaps. The first submap closes a loop with submaps 4 and 5.

5.2.4 Submap Graph SLAM

In this SLAM implementation, every node corresponds to a submap pose $\boldsymbol{x}_{\boldsymbol{s_k}} = (x, y, z, \phi, \theta, \psi)^\intercal$ chosen as a local reference frame for the patch. An edge constraint may represent either an odometry displacement or a registration result from the ICP registration. In the case of an odometry displacement $\boldsymbol{\Delta_{ij}} = \ominus \boldsymbol{x_j} \oplus \boldsymbol{x_i}$, the edge information is $\boldsymbol{z_{ij}} = \boldsymbol{\Delta_{ij}}$ with its information matrix being the inverse of the odometry uncertainty. When the edge corresponds to a loop closure registration, $\boldsymbol{z_{ij}}$ and $\boldsymbol{P_{ij}}$ correspond to $\boldsymbol{\Delta_{ij}^{LC}}$ and the inverse of the uncertainty obtained as an output of the registration algorithm. Every time a new registration edge is added to the graph, it is optimized in order to get a more accurate initial guess for the upcoming registrations. Once the last optimization is performed, the graph is updated with the new positions and the bathymetric map can be extracted using tail-to-tail operations and point cloud composition.

Figure 5.4 shows an example graph consisting of five submaps and two loop closures. In this example, graph optimization would have been triggered twice, once after the registration of node x_1 to x_4 and a second one after the registration of node x_1 to x_5.

5.3 Bathymetric Particle Filter SLAM using Grid Maps

In the previous section we proposed a submap graph SLAM reconstruction solution where the submap sections are registered to improve the global map consistence. However, the submaps themselves are not corrected and their inconsistencies will persist in the final map even if corrected. Even more, large amounts of three-dimensional data are difficult to handle and the presented method does not propose any memory handling. In this section a particle filter SLAM approach is proposed for laser-based bathymetric mapping. This approach will overcome the submap SLAM limitations previously mentioned. The work is based on [8], where the authors present BPSLAM, an algorithm targeted at acoustic bathymetric mapping. In our application, the sensing device is different from the used in the paper and we introduced three changes: (1) the observation model for the laser points is treated as Gaussian and (2) the particle movement accounts not only for x and y but also for the heading, and (3) the map representation has been altered from a grid map to a self-balancing binary search tree grid map to be more memory-efficient.

5.3.1 Particle Filter

A Particle Filter (PF) is a non-parametric implementation of the Bayes filter that can be used to approximate the probability distribution of a non-observable state. In robotics, a state x_t at time t is usually defined as the robot position in 2D or 3D space. The state x_t depends on the previous state x_{t-1} according to the probabilistic law $p(x_t|u_t, x_{t-1})$, where u_t is the control asserted in the time interval $(t-1, t]$. As the state is non-observable, the measurements z_t and their projections in the true state are used instead. The probability $p(z_t, x_t)$ is referred as the measurement model, whereas $p(x_t|u_t, x_{t-1})$ is the motion model. Using Bayes filters [131], the posterior probability can be expressed as a recursion of the previous probabilities under the initial condition $p(x_0|z_0, u_0) = p(x_0)$. This relation is expressed in equation (5.12).

$$p(x_t|z_t, u_t) = \mathrm{K} \cdot p(z_t|x_t) \int p(x_t|u_t, x_{t-1})\; p(x_{t-1}|z_{t-1}, u_{t-1})\; \mathrm{d}x_{t-1} \tag{5.12}$$

Closed form solutions to calculate (5.12) are only known for specialized cases. If $p(x_0)$ is Gaussian and $p(x_t|u_t, x_{t-1})$ and $p(z_t|x_t)$ are linear with added Gaussian noise (5.12) is equivalent to the Kalman filter. If the linearisation is obtained via a first order Taylor polynomial, the result is equivalent to an EKF.

Particle filters address the solution by using a set of hypotheses, conversely particles. A particle is a sample state drawn from a probability distribution that does not necessarily have to be Gaussian. All localisation particle filters share these same steps:

- **Initialization:** samples the M particles using $p(x_0)$.
- **Propagation:** uses a motion model described by $p(x_t|u_t, x_{t-1})$ to move the particles from their previous position to the current position with an added noise.
- **Weight:** uses a metric to evaluate the particles according to how well they match a measurement, a map, a position or any observable property, e.g. the likelihood for a particle to represent the true state.
- **Resampling:** removes low weight particles (e.g. unlikely representations) and replaces them with copies of the highest weighted particles.

Variations of this steps are available in the literature [132] [133]. One key aspect of PF is how to deal with particle depletion. Depletion happens when the set of particles is too scattered, its weights are low and a resampling stage would remove too many particles from the set. Importance resampling is a method to overcome particle depletion. The method defines an effective particle number N_{eff} as a measure for degeneration. When this number is larger than half of the particle population, resampling is performed. The metric is defined by equation (5.13)

$$N_{eff} = \frac{1}{\sum_{i=0}^{M}[w(i)]^2}, \tag{5.13}$$

where $w(i)$ is the i-th particle weight.

In a state space formed by all possible positions and maps, sampling is not efficient nor tractable. Rao-Backwellized Particle Filters (RBPF) perform a marginalisation of the probability distribution of the state space by sampling over the position, handling each particle a different map. However, handling a large number of maps is not feasible in terms of memory. Distributed Particle Mapping (DPM) enables to maintain a large number of maps efficiently by introducing particle ancestry. Each resampled particle keeps track of the changes performed in the map when compared to its parent, instead of holding an entire copy of the map.

5.3.2 Rao-Backwellized Particle Filter

Rao-Backwellized particle filters provide a framework for conducting particle-based SLAM, where each particle creates its own map. The algorithm that implements a RBPF is shown in Algorithm 5.1, extracted from [8].

Algorithm 5.1: RBPF SLAM Framework [8]

Data: N particles with poses sampled from some initial distribution and a map with prior information that may exist about the world.
Result: The best surviving particle and corresponding map.

while *not at the end of mission* **do**
  Read current observation
  for *i= 1 to N* **do**
    Propagate each particle pose to the time of the new observation by sampling from the vehicle motion model.
    Weight each particle based on how well the new observation agrees with its map.
  Resample the N particles from the current set with replacement. Perform this based on the particle weights so that particles with low weights are likely to be discarded while particles with high weights are likely to be duplicated.
  Update the N maps of the new particle set with the new observation.

Although this would be a straightforward implementation, the pipeline would require entire maps to be copied over every time a particle is resampled, which is non-efficient in terms of memory, and time-consuming. DPM addresses this issue by sharing a map among all particles, using labels to track which map sections belong and lineage to which particle.

5.3.2.1 Vehicle motion model and propagation

The vehicle motion model is split in two parts. An Extended Kalman Filter (EKF) (same as in section 5.2) shared by all the particles tracks the vehicle position, but only direct sensor observations such as depth, speed and orientations are used. The reason behind is that most alignment errors occur due to wrong vehicle position or heading. Therefore, each particle p_t^{id} holds an estimation of $\mathbf{x_t^{pf}} = (x_t, y_t, \psi_t)$, and an identifier. The rest of the state variables are obtained from an EKF shared by all the particles. Its state vector is described by equation (5.14).

$$\mathbf{x_t^{ekf}} = \left[z_t, \phi_t, \theta_t, \dot{x}_t, \dot{y}_t, \dot{z}_t, \dot{\phi}_t, \dot{\theta}_t, \dot{\psi}_t, \ddot{x}_t, \ddot{y}_t, \ddot{z}_t\right] \tag{5.14}$$

where z_t is the vehicle depth, ϕ_t, θ_t, ψ_t are roll, pitch and yaw angles respectively. The whole state is therefore a joint set of the PF and EKF states, defined by (5.15).

$$\mathbf{x_t} = (\mathbf{x_t^{pf}}, \mathbf{x_t^{ekf}}) \tag{5.15}$$

A particle set S_t is defined as (5.16)

$$\boldsymbol{S}_t = \begin{pmatrix} x_{1_t} & \cdots & x_{N_t} \\ y_{1_t} & \cdots & y_{N_t} \\ \psi_{1_t} & \cdots & \psi_{N_t} \\ id_1 & \cdots & id_N \end{pmatrix}, \tag{5.16}$$

where id_k is a particle identifier that will be used to build an ancestry tree for mapping. These particles are propagated using

$$x_{i_{t+1}} = x_{i_t} + \mathcal{N}(\mu_{\dot{x}}, \sigma_{\dot{x}}^2) \cdot \Delta t \tag{5.17}$$

$$y_{i_{t+1}} = y_{i_t} + \mathcal{N}(\mu_{\dot{y}}, \sigma_{\dot{y}}^2) \cdot \Delta t \tag{5.18}$$

$$\psi_{i_{t+1}} = \psi_{i_t} + \mathcal{N}(\mu_{\dot{\psi}}, \sigma_{\dot{\psi}}^2) \cdot \Delta t \tag{5.19}$$

The filter has to be initialized to a known position and from there the particles are propagated and their observations used for resampling, which will be explained after the grid map representation.

5.3.3 Grid Map Representation

Several map representations exists that can represent the three-dimensional structure of the seafloor. The approach used in the previous section was to store the map as a cloud of

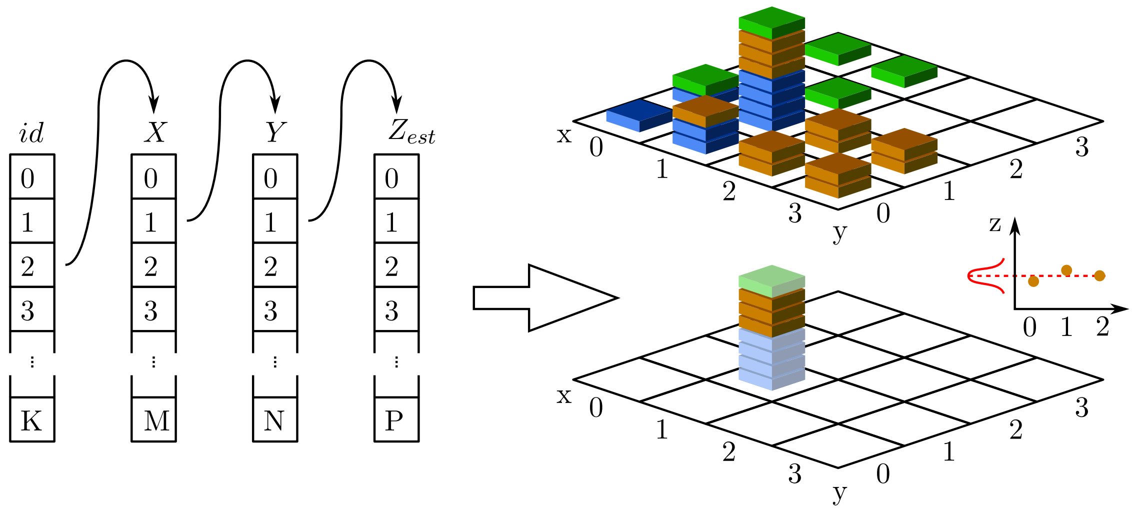


Figure 5.5: Map structure and memory storage. The four-indexed map allows to store just a copy of the complete map and allows a fast access and retrieval of the underlying depth observation. The map is first indexed by particle id, then by discrete x and y positions, and the deepest layer holds all the depth measurements available to that grid cell and particle.

3D points. Alternatively, grid-based map representations offer faster map access at the cost of losing resolution through discretization. The requirements for an underwater map can be simplified if the mapping is constrained to an elevation map, or a 2.5D map. That is, a 2D map in which each cell contains an estimate of the seafloor depth position and uncertainty. The vast majority of the seafloor can be considered 2.5D, except for subterranean cave systems, which are not in the scope of this SLAM solution. To store these maps in memory, a binary search tree has been proposed. Even with this new implementation, the memory requirements of the map scales quadratically with resolution, but the memory does not need to be preallocated to a known survey size.

At every grid cell, the depth measurement is stored as an array of observations, together with the cell mean and standard deviation that will be used to compute the resampling weight. A grid cell can be defined using (5.20)

$$\boldsymbol{G}^{\boldsymbol{k}}_{\boldsymbol{xy}} = (z_1, z_2, \ldots, z_n) \sim \mathcal{N}(\mu_z, \sigma_z), \tag{5.20}$$

where k is the k-th particle, z_i are the measurements, and (μ_z, σ_z) their Gaussian model. When a particle (representing a candidate robot path) senses the seafloor using a laser stripe, each 3D point is ray-traced to the corresponding grid of the map and stored in the grid cell's vector. This grid map representation is described in figure 5.5, where the four-indexed map structure is shown with an example of three particles and some measurements.

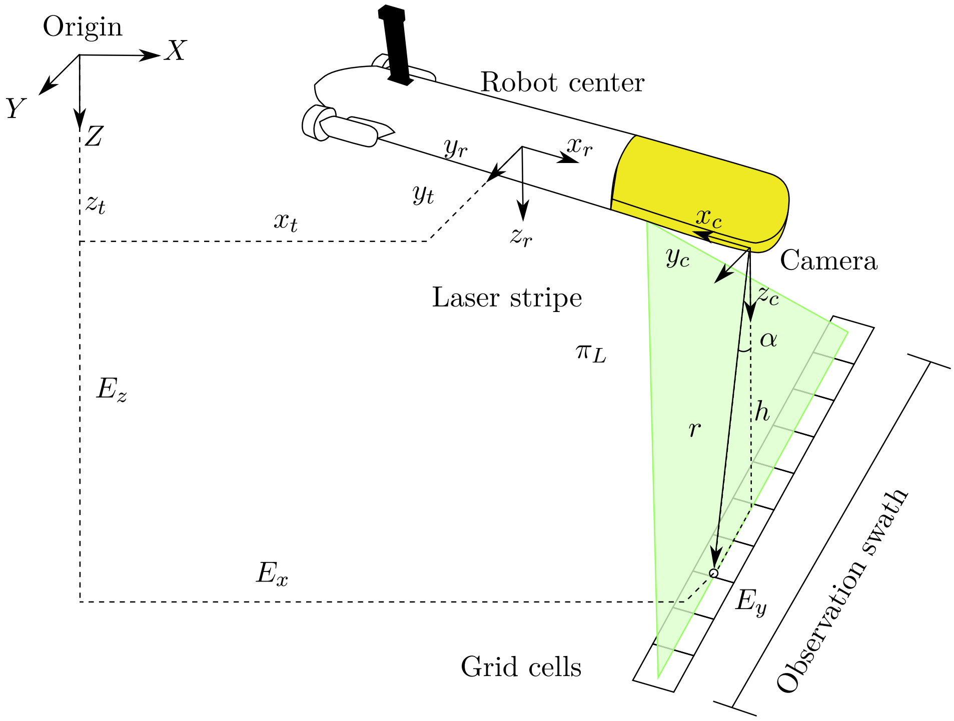


Figure 5.6: Relationship between a laser observation of a range r and a scan angle α, the state hypothesis $\boldsymbol{x_t}$ and the location of the seabed patch $\boldsymbol{E}$.

5.3.3.1 Grid map observations

At every camera frame, the triangulated laser points are projected to the grid map representation using the geometry of figure 5.6, where a grid cell (E_x, E_y, E_z) is observed from the camera at the range r and angle α. The corresponding uncertainty of the laser triangulation measurement can be computed as a function of range and angle as shown in (5.21)

$$\sigma_L = f(r, \alpha) = k_0 + k_1 \cdot r, \tag{5.21}$$

where the values k_0, k_1 have to be properly tuned to match the camera reprojection error and the depth uncertainty of the triangulation system when its geometry is taken into account, as explained in 3.1. The uncertainly is approximated by a line with slope k_1, crossing at the origin at k_0.

In such cases where two or more measurements from the same laser triangulation frame fall in the same grid cell, the observation mean and standard deviation is computed as the mixture of Gaussian distributions.

5.3.4 Particle Resampling and Lineage

Each particle maintains and builds its estimates of the seabed, observed as an altitude measurement that, with the added navigation depth, results into the seafloor depth. Prior to add the observations to the map, the particles are weighted based on how well their new seafloor measurement matches with previously stored estimates.

Using the geometry described in Figure 5.6, the laser points can be projected to the grid map and evaluated using (5.22) [134], as both the measurement and the estimation are assumed to be Gaussian.

$$weight = P\left[(\hat{z} - z_{obs}) = 0\right] = \frac{e^{-\frac{1}{2}\frac{\hat{\mu}_{est}-\mu_{obs}}{\hat{\sigma}^2_{est}+\sigma^2_{obs}}}}{\sqrt{2\pi(\hat{\sigma}^2_{est}+\sigma^2_{obs})}} \tag{5.22}$$

In a standard PF, when a particle is resampled it is normally duplicated one or more times. Instead of duplicating maps, DPM deals with this issue by creating new particles and relates them with ancestry or lineage links.

For example, a particle p of high weight survives to the resampling step. Therefore, it will be duplicated in the next iteration. The new particles resampled from p will have p as parent. The childs of p will build new map portions each. If a particle is resampled just once, the new copy is merged to its ancestor (e.g. its maps are joined).

If a set of *brothers* are resampled and all but one removed from the filter, the particle is fused to its parent. This can be also expressed as *lone-child* parents and their child are fused to lower the complexity of the ancestry tree. An example of this PF with grid maps can be seen in figure 5.7. Each particle maintains a map section generated using the observations received between two resampling event. A resampling event is triggered when a particle map estimate overlaps with itself (including ancestors) and each particle is weighted by the likelihood that it contains the true state, based on the new observation.

Every particle observation covers an entire laser swath, and each will have a different overlap with the rest of the map. A rule, already proposed in [8], that the resampled particles should have an overlap larger than a threshold γ is used. This limits the number of particles included in the resampling phase. If this threshold $\gamma = 100\%$, it would mean that a particle is resampled only if its entire swath overlaps with the already explored terrain. The particles that either do not overlap or that do not overlap enough, are considered as good as any other particle and therefore not resampled.

As discussed in [135], the resolution of the bathymetric maps is limited as the observation uncertainty can span more than just one grid. Data association techniques have not been applied and therefore the probability is constrained within grid cells. To check that this assumption is valid, a 95% confidence interval of the observation error can be used to size the minimum resolution of the grid.

5.3.5 Map and Trajectory Estimation

At the end of the trajectory, the particle with the lowest weight is queried and its map portion and its parents transform into 3D space using Algorithm 5.2. Each resulting 3D point contains

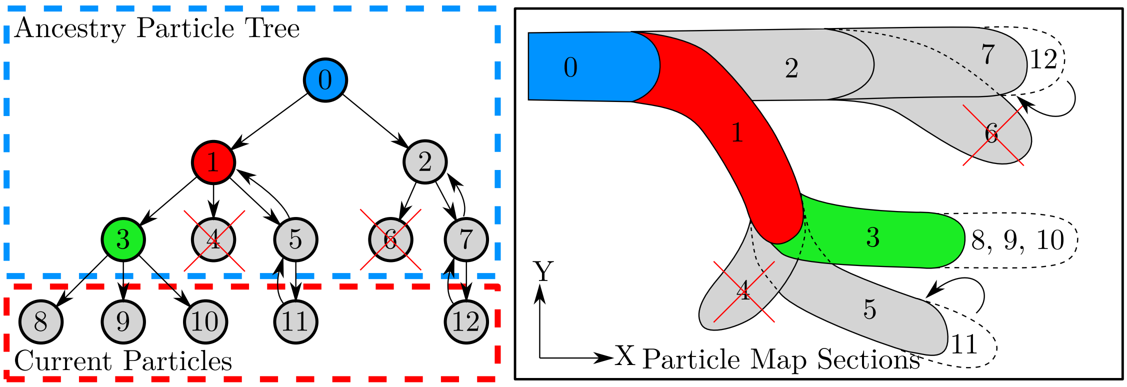


Figure 5.7: An example of the ancestry tree structure used in DPM Rao-Backwellized particle filters. The tree has had a resampling event that allowed Particles 5 and 7 to survive and Particle 3 to triplicate. These resampled particles are given new IDs and form the current particle set, indicated by the bottom layer of the tree. Particles that have not been resampled (indicated with a cross) or only possess one child (indicated by the curved arrows) are discarded and merged respectively, along with their map sections. While the particles in the new set have yet to create a map section of their own, each one inherits the map sections of its ancestors. Particle 0 is also called the root particle, as any map sections stored by this particle are available to all particles in the current set, as this particle is a common ancestor [8].

a registration error in case the grid cell that it represents has been observed more than once. The error is computed as the Consistency Based Error Evaluation (CBEE) [121], which will be explained in the following section with the experimental results.

Algorithm 5.2: Final map retrieval

Data: id^* best particle id, an ancestors vector v and a map $\boldsymbol{m}(id, x, y, \boldsymbol{z})$.
Result: The map corresponding to the particle with the best consistency.

```
Extract the list of parent particles from v → id_parents
Create an empty grid map to hold the result
for id in id_parents do
    for x in m[id] do
        for id in m[id][x] do
            Transform the grid cell to a 3D space
            Retrieve all the measurements and compute the mean → μ_z
            Set μ_z as the depth value and the maximum error as the registration error
              for that given cell
```

5.4 Experiments and results

In this section we compare and evaluate the results of the methods presented in sections 5.2 and 5.3 using two datasets gathered with different platforms.

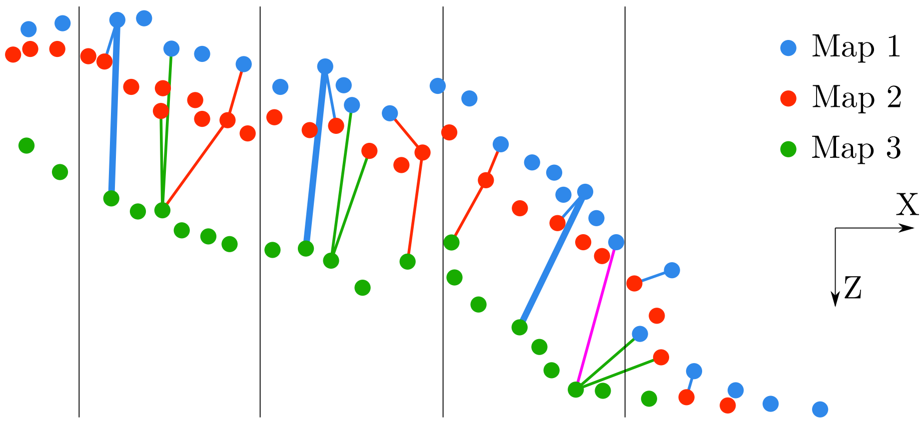


Figure 5.8: Cloud consistency error example. The colored points represent 3D points and the vertical divisions represent bins. Within each bin, a point is chosen at random from each map. The lines indicate the closest pairs of all points from the other maps. The bold blue lines indicate the maximum misregistration error within each bin. Note that when determining the closest points allowing the search to also consider points outside of the immediate bin will avoid bin size related artifacts. For example, the magenta line indicates how a nearest green-to-blue point pair would incorrectly be used if searching was only allowed inside a given bin. Finally, note that the right most bin with pairings does not show any Map 3 (green) pairings. This is because there are no Map 3 points in both surrounding bins.

To measure the performance of these methods two metrics are used: consistency and overlap. Consistency measures the agreement of the map with itself, whilst overlap measures the amount of spatial coincidence. If a map is consistent and has a high overlap ($> 60\%$) it would indicate a successful SLAM performance. If the overlap is low there could be two causes: (1) the *real* navigation path did not overlap and therefore the measure is right, or (2) the *real* navigation path did really overlap but the navigational errors have caused the path to diverge. In this case, the reconstructed terrain cannot be considered consistent.

5.4.1 Consistency assessment

One of the main concerns in unstructured environmental mapping is the lack of ground truth to validate the algorithm results. Therefore, the only option is to evaluate how does the map agree within itself. Roman and Singh [121] proposed a Consistency Based Error Evaluation (CBEE) that computes an error estimation based on the apparent thickness of the final map cloud. Ideally, the composite cloud would describe a one-point-thickness map layer with exact map registrations. However, when these transformations are not perfect, misalignments are introduced and the map will have a "thickness" whose average offset is the registration error. An example of this metric is shown in figure 5.8.

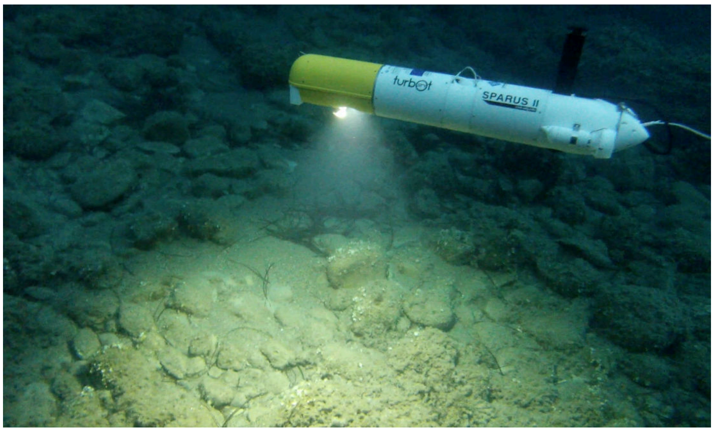


Figure 5.9: Turbot AUV diving in Port de Valldemossa. The shallow water area had to be recorded at night for the correct performance of the laser detection.

5.4.2 Laser stripe overlap

Both SLAM algorithms presented need laser lines to overlap in order to find bathymetric consistencies and correct the overall map and trajectory. The overlap that the current laser stripe presents to the known map is also a metric that tells how much information is obtainable from the ongoing transect given the settings used in the SLAM algorithm. For example, if the minimum overlap to correct a certain map is set at 60% and the overall laser stripe overlap is less than 40%, the resulting map will most likely not incorporate bathymetric knowledge due to lacking overlapping terrain. On the other hand, if the overall overlap is 80% the SLAM should have improved the map as the explored terrain has enough overlap to align and register bathymetric features.

5.4.3 Platforms

Two different AUVs have been used in this thesis to generate bathymetric maps of the seafloor: Turbot and AE2000f, which will be presented below.

5.4.3.1 Turbot AUV

Turbot is a Sparus II AUV, developed by *Universitat de Girona*, Spain. It is a 200-meter-rated, torpedo-shaped AUV with two propellers for surge and one for heave. A Teledyne Explorer DVL, an Analog Devices Adis IMU, a pressure sensor and a U-Blox GPS are its navigation sensors. Turbot is $1.6\,m$ long and weights $\sim 50\,kg$ The robot shown in figure 5.9, runs Ubuntu Linux with ROS middleware.

During this PhD programme, the robot has been equipped with two Allied Vision Technology G-283c cameras and a $1\,W$ $445\,nm$ laser stripe for seafloor reconstruction and inspection. To improve navigation accuracy, an Evologics 18/34 USBL and its corresponding modem have

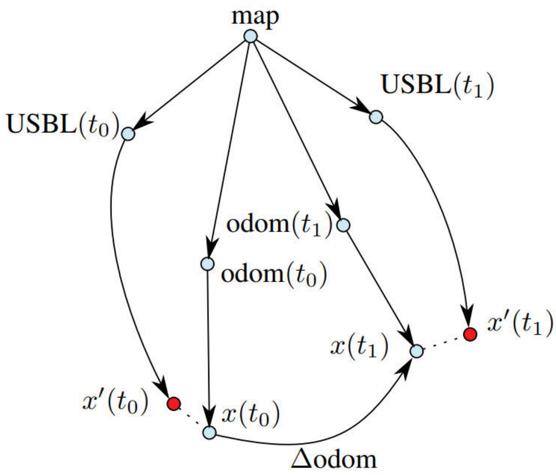


Figure 5.10: USBL architecture in Turbot AUV, presented in [9]. A fixed coordinate frame *map* references two children *odom* coordinate frames to correct for drift. This drift is calculated from a received USBL at time t_1 but measured at time t_0. The odometric displacement between t_0 and t_1 is used to update the corrected position $x'(t_1)$ by changing the transformation from *map* to *odom(t_1)*, leaving the rest unchanged.

been integrated into the vehicle navigation, the modem has been attached to Turbot and the USBL has been deployed from the shore or from a vessel depending on the experimental setup.

In regards to software, a newly developed EKF navigation architecture has been embedded in the main CPU to integrate USBL correction in real time. This EKF treats every USBL message as a delayed position update. The odometric displacement between the USBL measurement time and the current time is compounded to the USBL position measurement and handled to the EKF as an update at the current time. This method is shown in figure 5.10.

5.4.3.2 AE2000f AUV

AE2000f is a $2000\,m$ depth rated flight-style AUV instrumented with a high-altitude ($8\,m$ range) 3D imaging system and a water-column pH and temperature sensor. AE2000f operates at approximately $2\,kn$ at an altitude of $8\,m$, allowing it to visually map the seafloor at a rate of up to $\sim 40,000\ m^2/h$ at $\sim 8\,mm$ pixel resolution. AE2000f is $3\,m$ long and weighs $370\,kg$ in air.

The vehicle is equipped with a iXblue PHINS AHRS and DVL, a depth sensor, a GPS and SeaXerocks3 as high-altitude 3D imaging system. SeaXerocks3 consists of three cameras (two Xviii colour cameras for stereo mapping and a monochome Lumenera LM165), a $532\,nm$ laser stripe and four LED strobe lights.

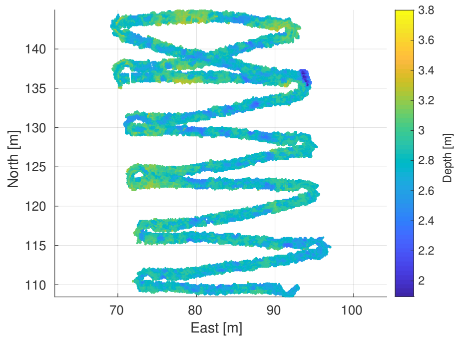
Figure 5.11: AE2000f AUV being deployed in Hydrate Ridge in front of R/V Falkor.

5.4.4 Valldemossa dataset

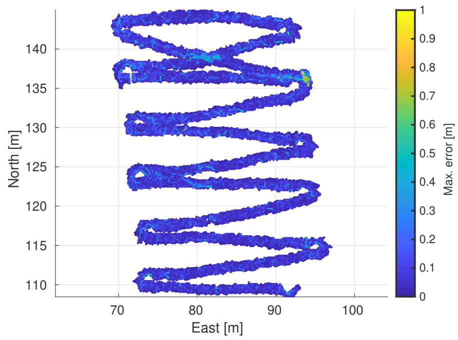
This dataset, named *Valldemossa*, was recorded by Turbot AUV. The vehicle dove in the waters of *Port de Valldemossa*, a shallow (< 5 m) rocky area next to a pebble beach. The area is partially covered in *Posidonia oceanica*, an endemic seagrass that forms meadows. The robot was programmed to perform a $25 \times 10\,m$ survey at $0.2\,m/s$ and at an altitude of $1.5\,m$ in a lawn-mower pattern with an ending diagonal transect. The cameras were shooting $960 \times 720\,px$ frames at $10\,Hz$, and at the survey altitude the laser swath was about $2.5\,m$ wide. Along track, the resolution was $1\,cm$ and $2\,mm$ accross track (e.g. one pixel footprint). In total, 12158 images were recorded in approximately 25 minutes.
The three-dimensional reconstructions of the survey are shown in figures 5.12(a) for dead reckoning and in figure 5.12(c) for an USBL-aided EKS navigation. Then the proposed SLAM reconstructions are shown in figure 5.12(e) for Submap SLAM and in figure 5.12(g) for BP-SLAM. The figures 5.12(b), 5.12(d), 5.12(f), 5.12(h) shown the CBEE of the respective navigations. Histograms for the CBEE are also available in figures 5.13(a) to 5.13(c), and the final AUV trajectory is depicted in figure 5.14.

5.4.4.1 Submap SLAM

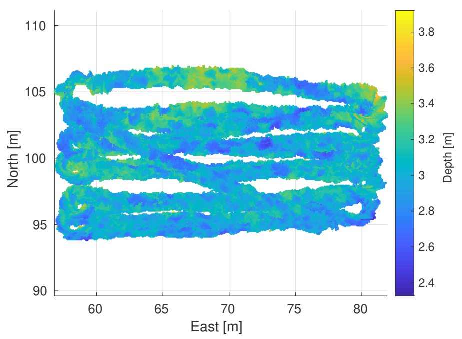
For this dataset, the patches were build for a minimum of 80 camera frames and a maximum of $3\,m$ of travelled distance. These patches were then downsampled to a maximum of $5k$ points and their normals estimated with a radius of $15\,cm$. The final grid resolution for CBEE has been $5\,cm$.
Two patches are considered loop closure candidates if they overlap more than 40% in their X-Y projection (kNN of $15\,cm$) and if they are more that five patches apart. This five-patch measure has been selected to avoid finding loop closures at the turns, when the vehicle is rolling and pitching due to its shape and dynamics. In figure 5.15 the submaps for this dataset are shown.



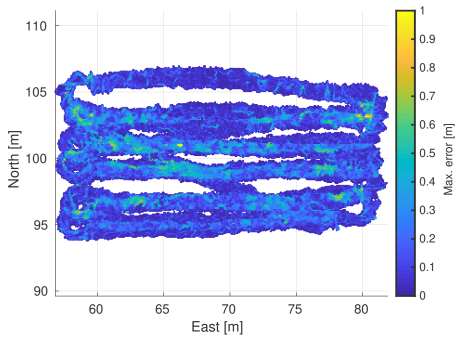
(a) Dead Reckoning Depth.



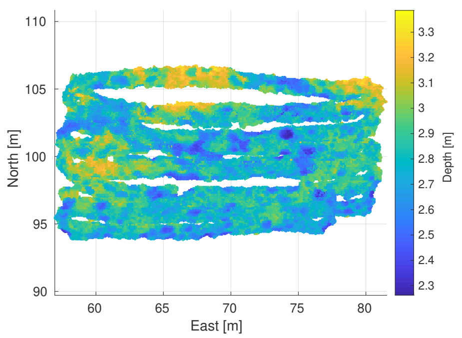
(b) Dead Reckoning CBEE.



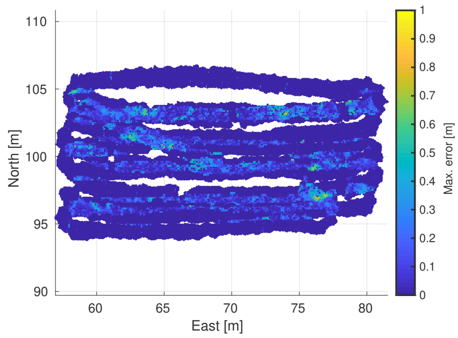
(c) USBL aided EKF Depth.



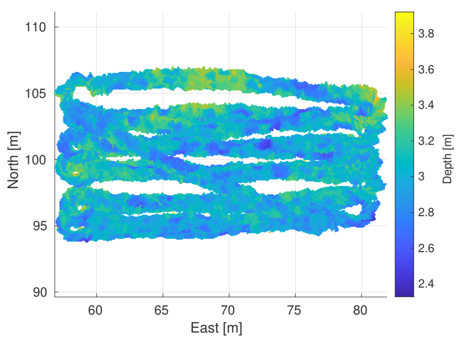
(d) USBL aided EKF CBEE.



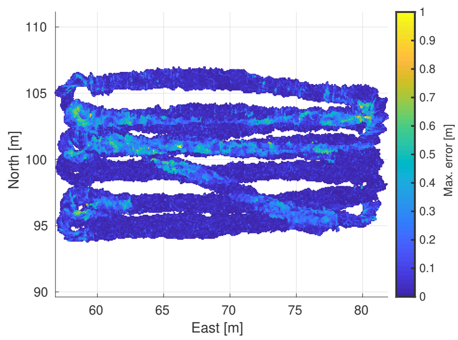
(e) Submap Depth.



(f) Submap CBEE.

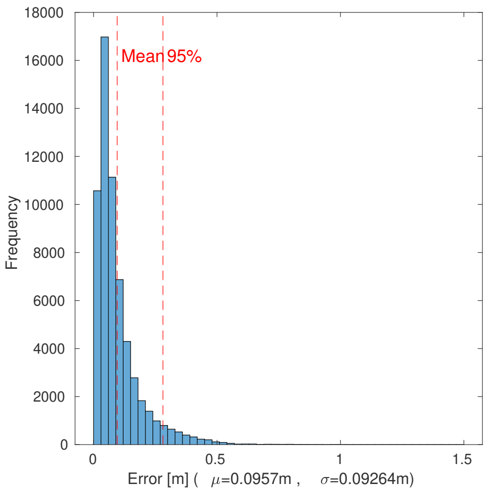


(g) BPSLAM Depth with 100 particles.

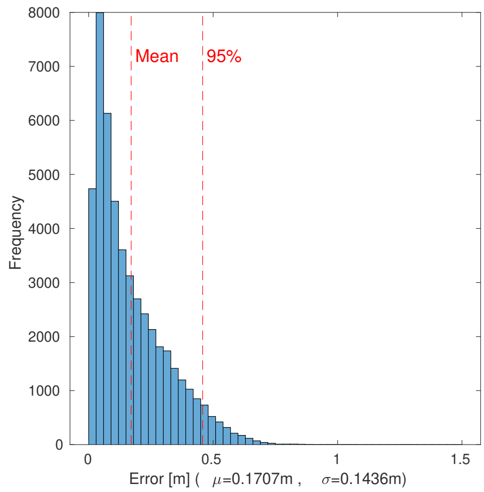


(h) BPSLAM CBEE with 100 particles.

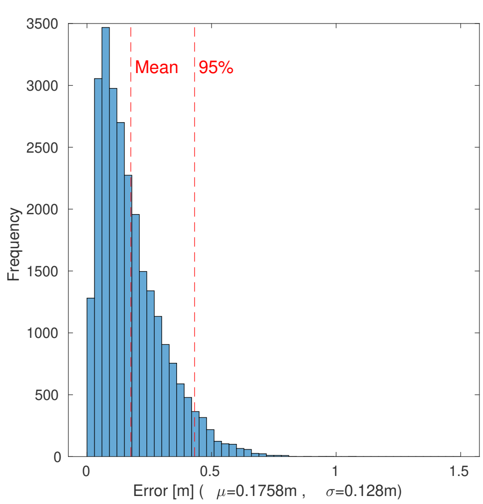
Figure 5.12: Valldemossa dataset results.



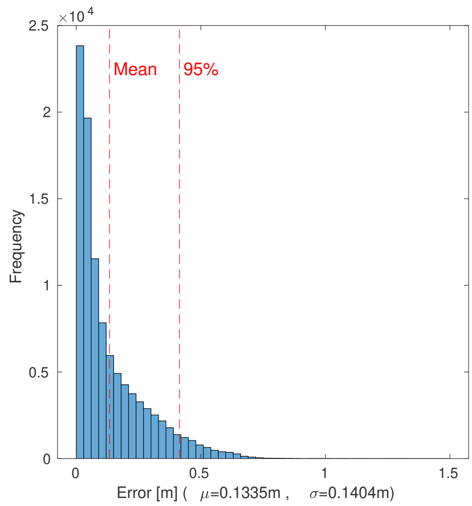
(a) Dead Reckoning CBEE histogram.



(b) USBL aided EKF CBEE histogram.



(c) Submap CBEE histogram.



(d) BPSLAM CBEE histogram with 100 particles.

Figure 5.13: CBEE Error histogram comparison for Valldemossa dataset.

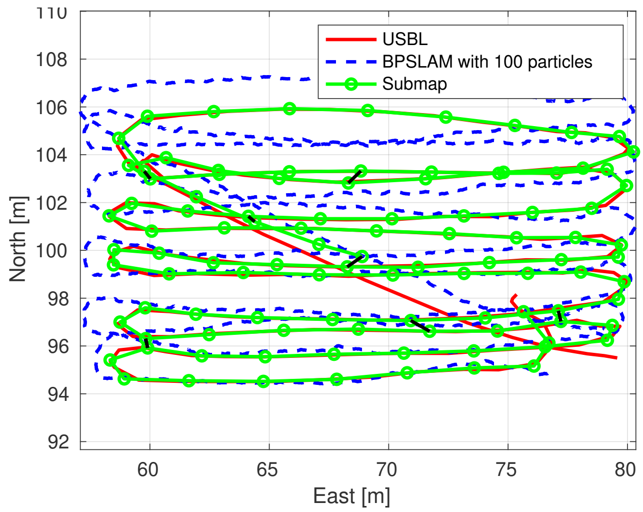


Figure 5.14: Valldemossa dataset AUV navigation path for the different solutions: USBL, BPSLAM and Submap SLAM with the graph LC edges in black. The dead reckoning solution has been omitted.

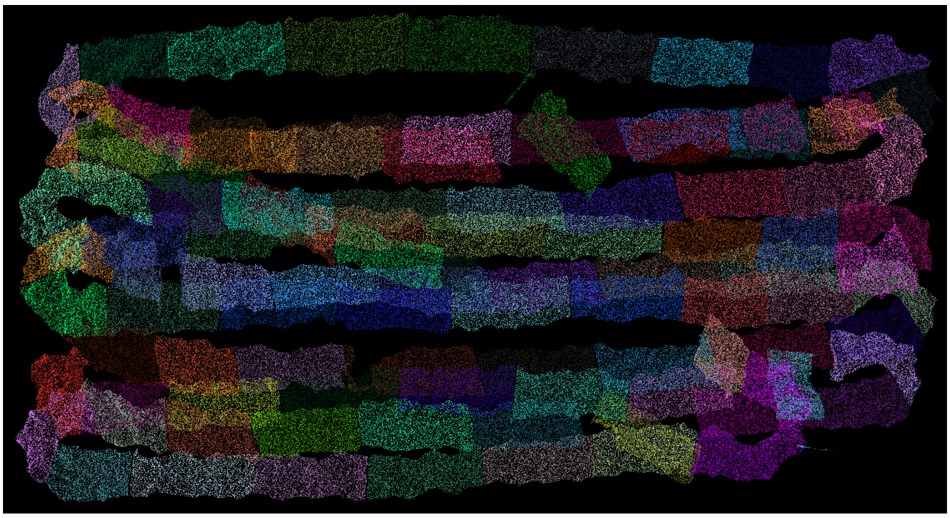


Figure 5.15: Valldemossa dataset Submap SLAM patches in random colours.

Table 5.1: Valldemossa dataset CBEE comparison between the different bathymetric solutions for an expected overlap of 60% when planning the mission.

Navigation	Overlap	CBEE μ (m)	CBEE σ (m)
Dead Reckoning	19.1%	0.0957	0.0926
USBL	72.4%	0.1707	0.1436
Submap SLAM	68.7%	0.1758	0.1280
BPSLAM 10 particles	72.2%	0.1361	0.1366
BPSLAM 25 particles	73.9%	0.1348	0.1391
BPSLAM 50 particles	74.5%	0.1344	0.1404
BPSLAM 100 particles	74.6%	0.1335	0.1404

5.4.4.2 Bathymetric Particle SLAM

The vehicle propagation noise has been chosen as $0.05\,m/s$ and $5 \cdot 10^{-3} rad/s$ for linear and angular speeds, which correspond as well to the EKF propagation noise. The minimum resample overlap has been fixed to 60% and the grid resolution to $5\,cm$.
Different number of particles have been tested for the dataset, and its results are reported in table 5.1 for $10, 25, 50$ and 100 particles.

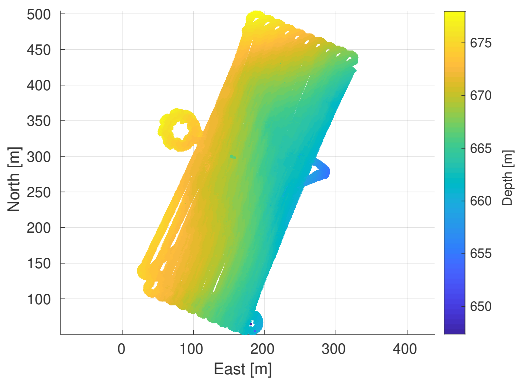
5.4.5 FK180731 dataset

The second dataset was recorded by AE2000f in Hydrate Ridge at $100\,km$ offshore of Oregon during FK180731 Adaptive Robotics cruise on board Falkor R/V [136]. The mission is a $350 \times 150\,m$ lawnmower pattern surveyed at $1\,m/s$ and at $8\,m$ of altitude. At the mission altitude, the laser swath spanned across $9.8\,m$ with an along track resolution of $60\,mm$ and $7.6\,mm$ across track. The LM165 records monochrome images at a frame rate of $15\,Hz$ and a resolution $1280 \times 1024\,px$. In total, 297000 images were recorded during more than five hours. The bathymetries of the surveyed area are shown in figures 5.16(a) for the dead reckoning solution, in figure 5.16(c) for the Submap SLAM and in 5.16(e) for BPSLAM. The CBEE overlaying the bathymetry is shown in figures 5.16(b), 5.16(d) and 5.16(f) respectively. Histograms for the CBEE are also available in figures 5.17(a) to 5.17(b), and the final AUV trajectory is retrieved in figure 5.18.

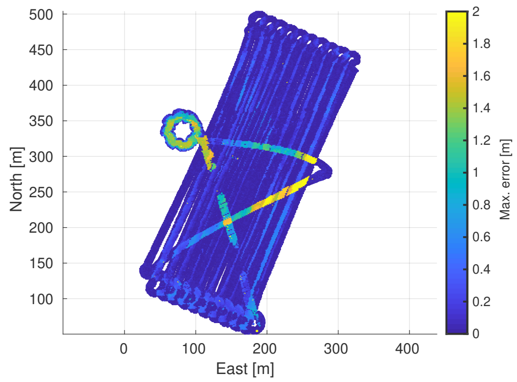
In the survey, one man-made structure was spotted, corresponding to an ODP 892 cruise in the same area. In that cruise the seafloor was drilled for gas prospection. In figures 5.19(a) and 5.19(b) a zoom-in of the structure and its capture in the AUV camera are shown.

5.4.5.1 Submap SLAM

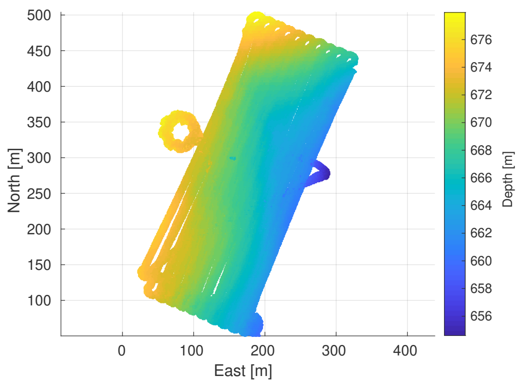
For this dataset, the patches were built for a minimum of 600 camera frames and a maximum of $30\,m$ of travelled distance. These patches were then downsampled to a maximum of $75k$



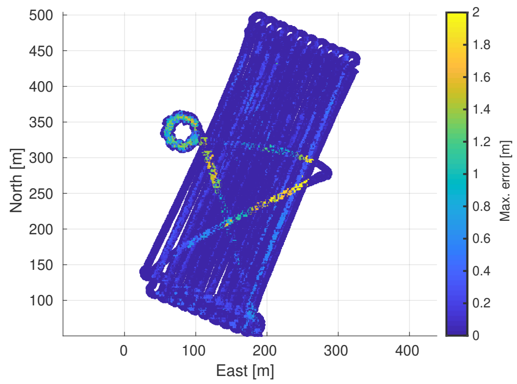
(a) Dead Reckoning Depth.



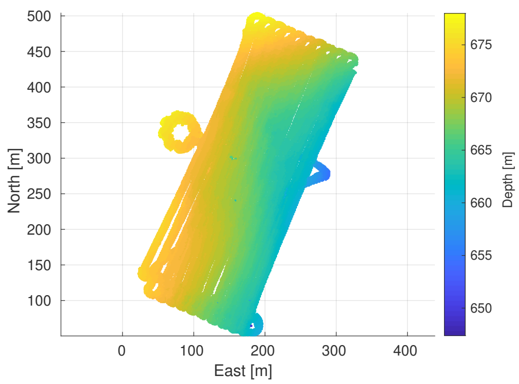
(b) Dead Reckoning CBEE.



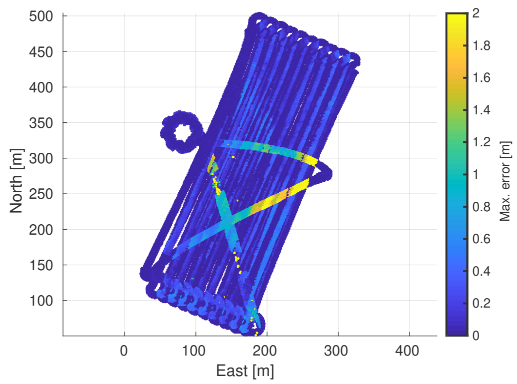
(c) Submap SLAM Depth.



(d) Submap SLAM CBEE.

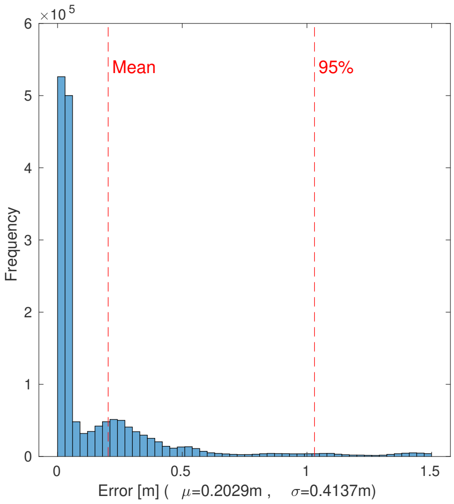


(e) BPSLAM Depth.

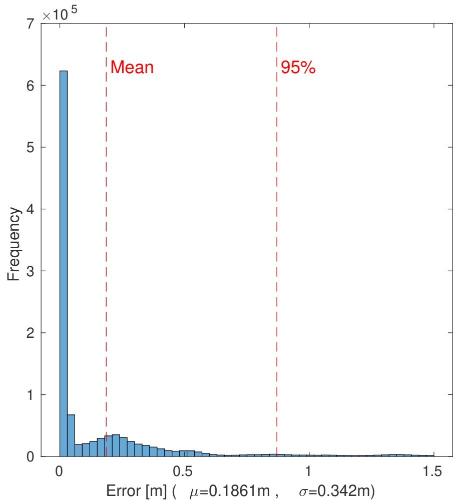


(f) BPSLAM CBEE.

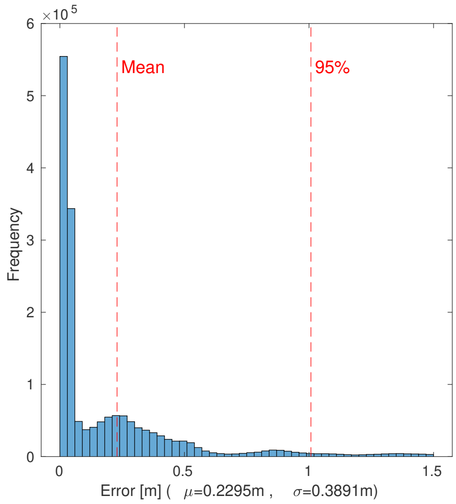
Figure 5.16: FK180731 dataset results.



(a) Dead Reckoning.

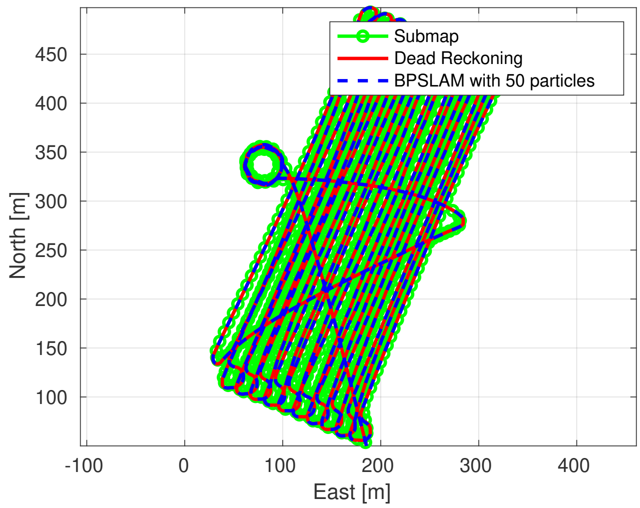


(b) Submap SLAM.

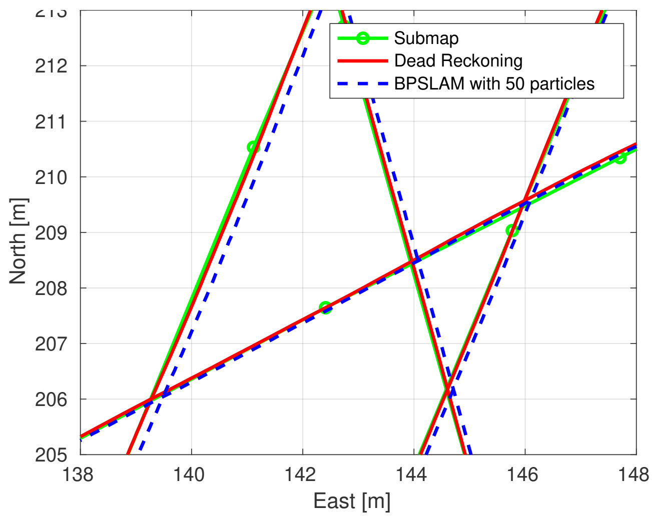


(c) BPSLAM.

Figure 5.17: CBEE Error histogram comparison for FK180731 dataset.

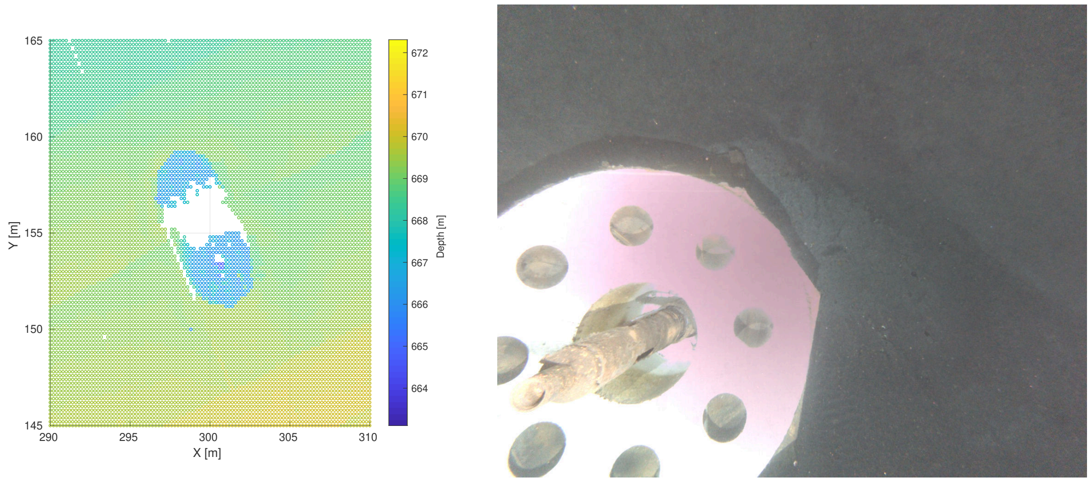


(a) Complete survey.



(b) Zoomed in region.

Figure 5.18: FK180731 dataset AUV navigation path for the different solutions: DR, BP-SLAM and Submap SLAM with the graph LC edges in black.



(a) Zoomed in region.   (b) Image of the structure.

Figure 5.19: FK180731 found drill structure in the bathymetric map and in the surveyed images.

points and their normals estimated with a radius of $40\,cm$. The final grid resolution for CBEE has been $20\,cm$ to require less memory. In order to be comparable to BPSLAM, the resolution of this dataset had to be reduced to the same resolution achievable by the particles.
As before, two patches are considered LC candidates if they overlap more than 40% in their X-Y projection (kNN of $40\,cm$) and if they are more that five patches apart. In figure 5.20 the submaps for the dataset are shown.

5.4.5.2 Bathymetric Particle SLAM

The vehicle propagation noise has been chosed as $0.05\,m/s$ and $5 \cdot 10^{-3} rad/s$ for linear and angular speeds. The minimum resample overlap has been fixed to 60% and the grid resolution to $20\,cm$.
Different number of particles have been tested for the dataset, and its results are reported in table 5.2 for $10, 25$ and 50 particles. Given the size of this dataset, a BPSLAM solution with 100 particles could not be performed due to memory restrictions.

5.4.6 Map improvement

The Dead Reckoning navigation solution has produced the maps with the largest errors in Valldemossa, and it does not correctly reproduce the surveyed area nor the path followed by the vehicle. One of the reasons of this behaviour is given by the low navigation altitude and the rocky bottom mixed with seagrass. This environment proved to be very challenging for the Turbot's DVL. In the next scenario (e.g. USBL-aided navigation), the drifting error was bounded by using a USBL as external positioning device. However, when the USBL-aided

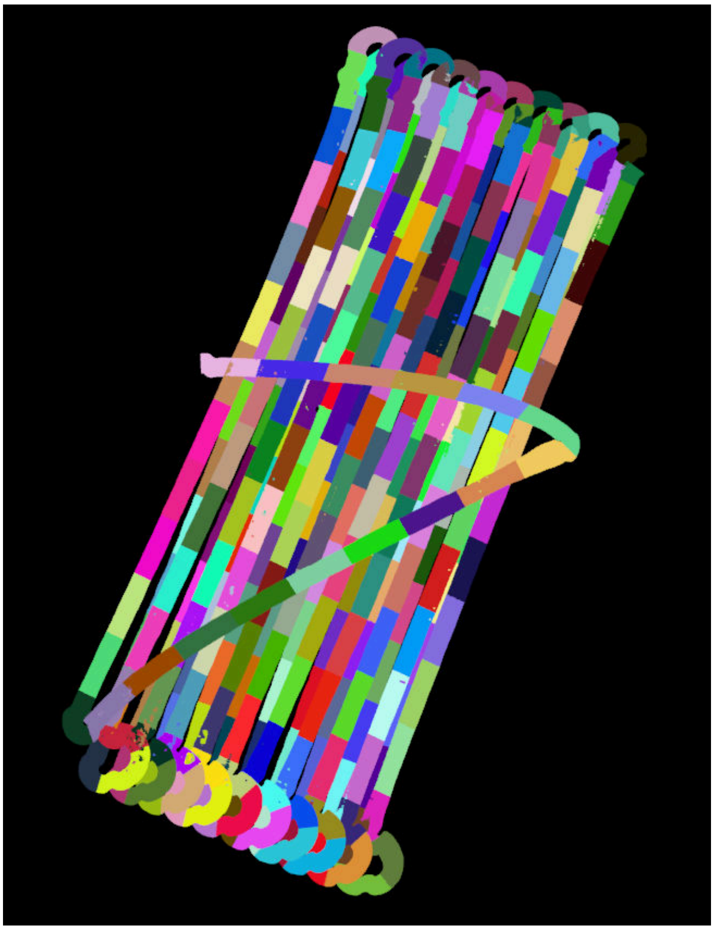


Figure 5.20: FK180731 dataset Submap SLAM patches in random colours.

Table 5.2: FK180731 dataset CBEE comparison between the different bathymetric solutions for a planned overlap of 37.2%.

Navigation	Overlap	CBEE	
		μ (m)	σ (m)
Dead Reckoning	44.8%	0.2029	0.4137
Submap SLAM	45.2%	0.1861	0.3420
BPSLAM 10 particles	44.8%	0.2295	0.3891
BPSLAM 25 particles	46.3%	0.2275	0.3833
BPSLAM 50 particles	46.7%	0.2206	0.3826

CBEE value is compared to DR value, it may look as if DR outperformed all other solutions. However, this is not the case, as DR has failed to follow the mission path and the number of overlapping points is low, almost non-existent. This lowers the overall error as there is nothing to compare to, whereas the USBL solution brings the track lines together increasing the overlap. This highlights the ability of the USBL of bounding the navigational drift.

The following figures show a clear improvement by using BPSLAM. The CBEE is not only lower, but also the bathymetry has less artefacts, although this approach still retains some error. On the other hand, the submapping solution does not improve over the USBL. There are two sources of error for this particular approach: the misregistration within the submaps is not corrected, as they are fixed, and the graph optimization 'pulls' submaps closer to reduce the error in the loop closing edges, but at the same time increases the error of the adjacent submaps, which finally translates into a larger CBEE.

In [121], a similar Submap SLAM solution was aimed for multibeam sonar. The main differences between a multibeam sonar and a laser stripe are range, resolution and number of beams (e.g. number of pixels of a laser stripe). As a matter of example, a multibeam sonar would return 128 beams acros a 120° beam width, (at $20\,m$ altitude: $70\,m$ footprint, 0.54 m across track and $0.25\,m$ along track resolution), whereas Turbot Laser triangulation is returning 960 3D points across 87°.

On Falkor dataset, the submapping navigation produced the best results and BPSLAM failed behind DR. The size of the dataset together with the low number of particles proved to be a limiting factor for the performance of the filter. Furthermore, BPSLAM only corrects the navigation in x, y, ψ, whereas submap SLAM performs a complete transformation. If we look closely at the navigation, the larger misregistration errors occur at the diagonal and perpendicular crossing, where the vehicle is travelling to its starting position and the altitude is 15 m, the double of the survey altitude. At that distance, the z errors are also magnified. With regards to the adjacency error caused by submapping, it can be seen that in FK180731 it does not happen as the seafloor is flatter and small translations do not occur to increase the error when shifting submaps for a better overlapping area.

To sum up, both submapping and BPSLAM have shown better self-consistent maps than plain navigation-only solutions. In comparison, although submapping corrects the overall error, it does not take into account the small misalignments or navigational errors within the patch, whereas BPSLAM corrects the overall navigation (northings, eastings and heading) and the reconstruction.

Chapter 6

Conclusions

This chapter concludes the thesis by presenting a summary of completed work in Section 6.1. The main contributions are reviewed under Section 6.2 and compelling areas for future work are outlined in Section 6.3.

6.1 Summary

This thesis has addressed the topic of underwater three-dimensional reconstruction using laser light. Laser light presents several benefits over classic imaging and floodlight solutions whilst at the same time, enables terrain measurements in poor visibility or featureless environments. We have focused in two approaches, one-shot reconstruction for close range measurements and laser bathymetry for longer distances.

In Chapter 2, we have reviewed and set the background on the available literature for the different imaging techniques and sensors used in underwater reconstruction. We summarized the most important aspects, their benefits and drawbacks.

In Chapter 3, we have provided the reader a preliminary background and a methodology to map laser image pixels to three-dimensional space, and proposed two calibration methods for a laser plane, which is easily extendable to other complex laser shapes.

In Chapter 4, we presented a novel one-shot laser-based structured light sensor, its implementation and evaluation in underwater conditions, turbidity and compared agains stereoscopy. The results prove the methodology presented in chapter 3 and showcase the usefulness of such device in murky waters or in dynamic environments.

In Chapters 5 we adapted two state of the art frameworks from multibeam sonar to the novel domain of laser bathymetry, and we validated them using two different datasets. These demonstrate how navigation and mapping corrections can be achieved when only sparse bathymetry overlap is available. Finally, these methods have been tested and validated in two real environments, with two different types of AUV.

6.2 Contributions

This dissertation has advanced the current state-of-the-art for underwater laser-based structured light sensors, providing one-shot and multishot solutions. We can break down this general contribution into the following items:

Complete review: A literature review of the most important contributions to the state of the art on imaging based underwater reconstruction and its sensors during the last fifteen years.

***In situ* calibration algorithm:** We have improved an on-site laser-to-camera calibration method already available [137]. In detail, our contribution focuses on using stereoscopy to our calibration pipeline.

One-shot reconstruction: A one shot laser pattern reconstruction pipeline in underwater environments with a proposed calibration method has been studied and validated over different visibility scenarios and compared to a standard stereo rig. The experiments performed have shown that it is capable of performing a sparse reconstruction in featureless and murky environments.

Bathymetric laser SLAM We have proposed a submap graph-SLAM solution and proved its effectiveness with a quantitative evaluation over two datasets, and we have studied as well a multishot laser registration using a known Bathymetric SLAM framework. It is worth mentioning that this framework was originally described for sonar-based bathymetry, but in this thesis it has been demonstrated its potential application with laser data. The experiments show an improvement on the overall reconstructed map.

6.3 Future work

The topic of the thesis has opened new research areas to further explore. In order of importance, the following tasks are planned to be carried out

- Develop a Bathymetric SLAM based on quadtree representation using Gaussian Processes to correct non-overlapping navigation whilst efficiently handling memory. With this type of approach the map corrections would not depend on the amount of overlap between transects. Available recorded data could be post-processed to obtain better quality seafloor maps.

- To progress in a hybrid mapping solution using laser bathymetry and visual features. This future work could enable to close loops using both shape and colour. Areas of the ocean have different colours / bacterial mats or different substrates that would not be clear in their shape, but easy to pick up using colour information.

- For the one-shot laser, overcome the laser correspondence limitations by increasing the number of cameras (e.g. stereo and laser). A second camera would overcome simple occlusion scenarios and would be beneficial to correctly label laser lines.
- Extend our laser pipeline solution to support different types of laser patterns and wavelengths, as well as aid the correspondence of points to the pattern to autocalibrate the multiline laser projector. Extending the work of this PhD for autocalibration would be a natural path to follow as work has been presented for a single line (descending helix).

Bibliography

[1] F. R. Dalgleish, F. M. Caimi, W. B. Britton, and C. F. Andren, “Improved LLS imaging performance in scattering-dominant waters,” in *Proc. SPIE 7317, Ocean Sensing and Monitoring*, W. W. Hou, Ed., vol. 7317, 5 2009, pp. 73 170E–73 170E–12.

[2] L. J. Mullen, V. M. Contarino, A. Laux, B. M. Concannon, J. P. Davis, M. P. Strand, and B. W. Coles, “Modulated laser line scanner for enhanced underwater imaging,” in *Proc. SPIE. 3761, Airborne and In-Water Underwater Imaging*, G. D. Gilbert, Ed., vol. 3761. SPIE, 10 1999, pp. 2–9.

[3] N. Törnblom, “Underwater 3D Surface Scanning using Structured Light,” Ph.D. dissertation, Uppsala Universitet, 2010, available at http://www.diva-portal.org/smash/get/diva2:378911/FULLTEXT01.pdf.

[4] F. Bruno, A. Gallo, M. Muzzupappa, B. Davidddde Petriaggi, and P. Caputo, “3D documentation and monitoring of the experimental cleaning operations in the underwater archaeological site of Baia (Italy),” *Digital Heritage International Congress (DigitalHeritage)*, pp. 105–112, 2013.

[5] F. Bonin-Font, A. Cosic, P. L. Negre, M. Solbach, and G. Oliver, “Stereo slam for robust dense 3d reconstruction of underwater environments,” in *OCEANS 2015 - Genova*, May 2015, pp. 1–6.

[6] Wikipedia contributors, “Bayer filter — Wikipedia, the free encyclopedia,” https://en.wikipedia.org/w/index.php?title=Bayer_filter&oldid=867116795, 2018, [Online; accessed 12-February-2019].

[7] E. Garcia-Fidalgo, A. Ortiz, F. Bonnin-Pascual, and J. P. Company, “Vessel Visual Inspection: A Mosaicing Approach,” Department of Mathematics and Computer Science, University of the Balearic Islands, Palma de Mallorca, Tech. Rep. A-01-2015, Mar. 2015. [Online]. Available: http://dmi.uib.es/~egarcia/TRA012015.pdf

[8] S. A. Barkby, S. B. Williams, O. Pizarro, and M. V. Jakuba, “A featureless approach to efficient bathymetric SLAM using distributed particle mapping,” *Journal of Field Robotics*, vol. 28, no. 1, pp. 19–39, jan 2011. [Online]. Available: http://doi.wiley.com/10.1002/rob.20382

[9] E. Guerrero-Font, M. Massot-Campos, P. L. Negre, F. Bonin-Font, and G. O. Codina, “An USBL-aided multisensor navigation system for field AUVs,” in *Multisensor Fusion*

and Integration for Intelligent Systems (MFI), 2016 IEEE International Conference on. IEEE, 2016, pp. 430–435.

[10] "A decade of compiling the sharpest mars map," https://mars.nasa.gov/resources/8333/a-decade-of-compiling-the-sharpest-mars-map/, accessed: 2018-12-15.

[11] "How much of the ocean have we explored?" https://oceanservice.noaa.gov/facts/exploration.html, accessed: 2018-12-15.

[12] D. J. Lockwood, *Rayleigh and Mie Scattering.* New York, NY: Springer New York, 2016, pp. 1097–1107. [Online]. Available: https://doi.org/10.1007/978-1-4419-8071-7{_}218

[13] Y. Y. Schechner and N. Karpel, "Recovery of underwater visibility and structure by polarization analysis," *IEEE Journal of Oceanic Engineering*, vol. 30, no. 3, pp. 570–587, July 2005.

[14] D. Fornas, J. Sales, A. Peñalver, J. Pérez, J. J. Fernández, R. Marín, and P. J. Sanz, "Fitting primitive shapes in point clouds: a practical approach to improve autonomous underwater grasp specification of unknown objects," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 28, no. 1-2, pp. 369–384, 2016. [Online]. Available: https://doi.org/10.1080/0952813X.2015.1024495

[15] N. Fairfield and D. Wettergreen, "Active localization on the ocean floor with multibeam sonar," in *OCEANS 2008*, 2008.

[16] "Phins subsea specifications," https://www.ixblue.com/products/phins-subsea#specifications, accessed: 2019-07-12.

[17] M. Massot-Campos and G. Oliver-Codina, "Optical Sensors and Methods for Underwater 3D Reconstruction," *Sensors*, vol. 15, no. 12, pp. 31 525–31 557, dec 2015. [Online]. Available: http://www.mdpi.com/1424-8220/15/12/29864/htm

[18] F. Bonin-Font, G. Oliver, S. Wirth, M. Massot-Campos, P. L. Negre, and J. P. Beltran, "Visual Sensing for Autonomous Underwater Exploration and Intervention Tasks," *Ocean Engineering*, vol. 93, pp. 25–44, 2015.

[19] M. Leat, A. Bodenmann, M. Massot-Campos, and B. Thornton, "Analysis of measurement uncertainty in high-resolution laser-scanned bathymetric measurements of seafloor topology using an autonomous underwater vehicle," in *Autonomous Underwater Vehicles (AUV), 2018 IEEE/OES.* IEEE, 2018.

[20] M. Massot-Campos, G. Oliver, A. Bodenmann, and B. Thornton, "Submap bathymetric SLAM using structured light in underwater environments," in *Autonomous Underwater Vehicles (AUV), 2016 IEEE/OES.* IEEE, 2016, pp. 181–188.

[21] M. Massot-Campos, F. Bonin Font, P. L. Negre Carrasco, E. Guerrero, A. Martorell, and G. Oliver Codina, "A 3D mapping, obstacle avoidance and acoustic communication payload for the SPARUS II," in *Instrumentation Viewpoint*, no. 19. SARTI, 2016, pp. 31–33.

[22] M. Massot-Campos, G. Oliver, H. Kemal, Y. Petillot, and F. Bonin-Font, "Structured light and stereo vision for underwater 3D reconstruction," in *IEEE/MTS Oceans'15*, 2015.

[23] M. Massot-Campos and G. Oliver-Codina, "One-Shot Underwater 3D Reconstruction," in *Proc. 19th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA 2014)*, 2014.

[24] M. Massot-Campos and G. Oliver-Codina, "Underwater laser-based structured light system for one-shot 3d reconstruction," in *SENSORS, 2014 IEEE*, Nov 2014, pp. 1138–1141.

[25] M. Massot-Campos and G. Oliver-Codina, "Survey on Underwater 3D Reconstruction," in *Technical Reports*. Universitat de les Illes Balears, 2014, no. A-04-2014, pp. 1–10.

[26] J. Sales, M. Massot-Campos, P. Lluís Negre, G. Oliver, and P. J. Sanz, "Integración de técnicas ópticas de reconstrucción 3D para mejorar la planificación de agarres en tareas de manipulación arqueológica subacuática," in *XXXVI Jornadas Nacionales de Automatica*, 2015.

[27] J. S. Jaffe, K. Moore, J. McLean, and M. Strand, "Underwater Optical Imaging: Status and Prospects," *Oceanography*, vol. 14, no. 3, pp. 64–75, 2001.

[28] B. Foley and D. Mindell, "Precision Survey and Archaeological Methodology in Deep Water," *MTS Journal*, vol. 36, no. 3, pp. 13–20, 2002.

[29] D. M. Kocak, F. R. Dalgleish, and F. M. Caimi, "A focus on recent developments and trends in underwater imaging," *Marine Technology Society Journal*, vol. 42, no. 1, pp. 52–67, 2008.

[30] D. M. Kocak and F. M. Caimi, "The Current Art of Underwater Imaging With a Glimpse of the Past and Vision of the Future," *Marine Technology Society Journal*, vol. 39, no. 3, pp. 5–26, 2005.

[31] F. M. Caimi, D. M. Kocak, F. Dalgleish, and J. Watson, "Underwater imaging and optics: Recent advances," in *Proceedings of MTS/IEEE Oceans*, vol. 45. IEEE, 2008, pp. 1–9.

[32] F. Bonin-Font and A. Burguera, "Imaging systems for advanced underwater vehicles," *Journal of Maritime Research*, vol. VIII, no. 1, pp. 65–86, 2011.

[33] G. Bianco, A. Gallo, F. Bruno, and M. Muzzupappa, "A comparative analysis between active and passive techniques for underwater 3D reconstruction of close-range objects." *Sensors (Basel, Switzerland)*, vol. 13, no. 8, pp. 11 007–11 031, 1 2013.

[34] M. Eric, R. Kovacic, G. Berginc, M. Pugelj, Z. Stopinsek, and F. Solina, "The impact of the latest 3D technologies on the documentation of underwater heritage sites," in *Digital Heritage International Congress (DigitalHeritage)*. IEEE, 10 2013, pp. 281–288.

[35] A. Jordt-Sedlazeck, "Underwater 3D Reconstruction Based on Physical Models for Refraction and Underwater Light Propagation," Ph.D. dissertation, Kiel University, Kiel, 2014, available at http://www.inf.uni-kiel.de/de/forschung/publikationen/kcss.

[36] J. S. Jaffe, "Computer modeling and the design of optimal underwater imaging systems," *IEEE Journal of Oceanic Engineering*, vol. 15, no. 2, pp. 101–111, 4 1990.

[37] K. Pathak, A. Birk, and N. Vaskevicius, "Plane-based registration of sonar data for underwater 3D mapping," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 10 2010, pp. 4880–4885.

[38] N. Hurtos, X. Cufi, and J. Salvi, "Calibration of optical camera coupled to acoustic multibeam for underwater 3D scene reconstruction," in *Proceedings of MTS/IEEE Oceans*. IEEE, 5 2010, pp. 1–7.

[39] Y. Guo, "3D underwater topography rebuilding based on single beam sonar," in *2013 IEEE International Conference on Signal Processing, Communication and Computing (ICSPCC 2013)*, vol. 1. IEEE, 8 2013, pp. 1–5.

[40] T. Blueview, "3D Mechanical Scanning Sonar," http://www.blueview.com/products/3d-multibeam-scanning-sonar, accessed: 2015-09-30.

[41] E. Coiras, Y. Petillot, and D. M. Lane, "Multiresolution 3-D reconstruction from side-scan sonar images." *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 16, no. 2, pp. 382–390, 2 2007.

[42] N. Brahim, D. Gueriot, S. Daniel, and B. Solaiman, "3D reconstruction of underwater scenes using DIDSON acoustic sonar image sequences through evolutionary algorithms," in *Proceedings of MTS/IEEE Oceans*. IEEE, 6 2011, pp. 1–6.

[43] S. Negahdaripour, H. Sekkati, and H. Pirsiavash, "Opti-acoustic stereo imaging: on system calibration and 3-D target reconstruction." *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 18, no. 6, pp. 1203–1214, 6 2009.

[44] M. Babaee and S. Negahdaripour, "3-D Object Modeling from Occluding Contours in Opti-Acoustic Stereo Images," in *Proceedings of MTS/IEEE Oceans*, vol. 1, 2013, pp. 1–8.

[45] S. Negahdaripour, "On 3-D reconstruction from stereo FS sonar imaging," in *Proceedings of MTS/IEEE Oceans*. IEEE, 9 2010, pp. 1–6.

[46] M. Aykin and S. Negahdaripour, "Forward-Look 2-D Sonar Image Formation and 3-D Reconstruction," in *Proceedings of MTS/IEEE Oceans - San Diego, 2013*. IEEE, 9 2013, pp. 1–4.

[47] L. Rosenblum and B. Kamgar-Parsi, "3D reconstruction of small underwater objects using high-resolution sonar data," *Proceedings of the 1992 Symposium on Autonomous Underwater Vehicle Technology*, pp. 228–235, 1992.

[48] B. D. Reineman, L. Lenain, D. Castel, and W. K. Melville, "A Portable Airborne Scanning Lidar System for Ocean and Coastal Applications," *Journal of Atmospheric and Oceanic Technology*, vol. 26, no. 12, pp. 2626–2641, 12 2009.

[49] N. Cadalli, P. J. Shargo, D. C. Munson, Jr., and A. C. Singer, "Three-dimensional tomographic imaging of ocean mines from real and simulated lidar returns," in *Proc. SPIE 4488, Ocean Optics: Remote Sensing and Underwater Imaging*, R. J. Frouin and G. D. Gilbert, Eds., 1 2002, pp. 155–166.

[50] F. Pellen, V. Jezequel, G. Zion, and B. L. Jeune, "Detection of an underwater target through modulated lidar experiments at grazing incidence in a deep wave basin." *Applied optics*, vol. 51, no. 31, pp. 7690–7700, 11 2012.

[51] L. Mullen, A. Vieira, P. Herezfeld, and V. Contarino, "Application of radar technology to aerial lidar systems for enhancement of shallow underwater target detection," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 43, no. 9, pp. 2370–2377, Sep 1995.

[52] L. J. Mullen and V. M. Contarino, "Hybrid LIDAR-radar: Seeing through the scatter," *IEEE Microwave Magazine*, vol. 1, no. 3, pp. 42–48, 2000.

[53] K. Moore, J. S. Jaffe, and B. Ochoa, "Development of a new underwater bathymetric laser imaging system: L-bath," *Journal of Atmospheric and Oceanic Technology*, vol. 17, pp. 1106–1117, 2000.

[54] K. Moore and J. S. Jaffe, "Time-evolution of high-resolution topographic measurements of the sea floor using a 3-D laser line scan mapping system," *IEEE Journal of Oceanic Engineering*, vol. 27, no. 3, pp. 525–545, 7 2002.

[55] D. McLeod and J. Jacobson, "Autonomous Inspection using an Underwater 3D LiDAR," in *Proceedings of MTS/IEEE Oceans*, 2013.

[56] M. Imaki, H. Ochimizu;, H. Tsuji;, S. Kameyama;, T. Saito;, S. Ishibashi;, and H. Yoshida, "Underwater three-dimensional imaging laser sensor with 120-deg wide-scanning angle using the combination of a dome lens and coaxial optics," *Optical Engineering*, vol. 56, no. 3, p. 031212, 2016.

[57] B. Cochenour, L. J. Mullen, and J. Muth, "Modulated pulse laser with pseudorandom coding capabilities for underwater ranging, detection, and imaging." *Applied optics*, vol. 50, no. 33, pp. 6168–78, 11 2011.

[58] L. Rumbaugh, Y. Li, E. Bollt, and W. Jemison, "A 532 nm Chaotic Lidar Transmitter for High Resolution Underwater Ranging and Imaging," in *Proceedings of MTS/IEEE Oceans*, 2013, pp. 1–6.

[59] L. de Dominicis, G. Fornetti, M. Guarneri, M. F. de Collibus, M. Francucci, M. Nuvoli, A. Al-Obaidi, D. Mcstay, *et al.*, "Structural Monitoring Of Offshore Platforms By 3d Subsea Laser Profilers." in *Offshore Mediterranean Conference.* Offshore Mediterranean Conference, 2013.

[60] A. Gordon, "Use of Lases Scanning Systems on Mobile Underwater Platforms," in *AUV Conf.*, 1992, pp. 202–205.

[61] 3d at Depth, "SL1 High Resolution Subsea Laser Scanner," http://www.3datdepth.com/sl1overview/, 2015-09-30.

[62] J. Salvi, S. Fernandez, T. Pribanic, and X. Llado, "A state of the art in structured light patterns for surface profilometry," *Pattern Recognition*, vol. 43, no. 8, pp. 2666–2680, 8 2010.

[63] J. Salvi, J. Pagès, and J. Batlle, "Pattern codification strategies in structured light systems," *Pattern Recognition*, vol. 37, no. 4, pp. 827–849, 2004. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0031320303003303

[64] Q. Zhang, Q. Wang, Z. Hou, Y. Liu, and X. Su, "Three-dimensional shape measurement for an underwater object based on two-dimensional grating pattern projection," *Optics & Laser Technology*, vol. 43, no. 4, pp. 801–805, 6 2011.

[65] F. Bruno, G. Bianco, M. Muzzupappa, S. Barone, and A. Razionale, "Experimentation of structured light and stereo vision for underwater 3D reconstruction," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 66, no. 4, pp. 508–518, 7 2011.

[66] S. Narasimhan and S. Nayar, "Structured Light Methods for Underwater Imaging: Light Stripe Scanning and Photometric Stereo," in *Proceedings of MTS/IEEE Oceans*. IEEE, 2005, pp. 1–8.

[67] J. S. Jaffe and C. Dunn, "A Model-Based Comparison Of Underwater Imaging Systems," *Proceedings of the Ocean Optics*, vol. IX, no. 925, pp. 344–350, 1988.

[68] S. Tetlow and J. Spours, "Three-dimensional measurement of underwater work sites using structured laser light," *Measurement Science and Technology*, vol. 1162, 1999.

[69] H. Kondo, T. Maki, T. Ura, Y. Nose, T. Sakamaki, and M. Inaishi, "Structure tracing with a ranging system using a sheet laser beam," in *Proceedings of the 2004 International Symposium on Underwater Technology (IEEE Cat. No.04EX869)*. IEEE, 2004, pp. 83–88.

[70] M. Hildebrandt, J. Kerdels, J. Albiez, and F. Kirchner, "A practical underwater 3D-Laserscanner," in *Proceedings of MTS/IEEE Oceans*. IEEE, 2008, pp. 1–5.

[71] A. Bodenmann, B. Thornton, T. Nakatani, and T. Ura, "3d colour reconstruction of a hydrothermally active area using an underwater robot," in *OCEANS 2011*, Sept 2011, pp. 1–6.

[72] A. Bodenmann, B. Thornton, S. Hara, K. Hioki, M. Kojima, T. Ura, M. Kawato, and Y. Fujiwara, "Development of 8m long range imaging technology for generation of wide area colour 3D seafloor reconstructions," in *Proceedings of MTS/IEEE Oceans*. IEEE, 10 2012, pp. 1–4.

[73] A. Bodenmann, B. Thornton, and T. Ura, "Development of long range color imaging for wide area 3D reconstructions of the seafloor," *2013 IEEE International Underwater Technology Symposium (UT)*, pp. 1–5, 3 2013.

[74] A. Bodenmann, B. Thornton, R. Nakajima, H. Yamamoto, and T. Ura, "Wide Area 3D Seafloor Reconstruction and its Application to Sea Fauna Density Mapping," in *Proceedings of MTS/IEEE Oceans*, 2013, pp. 4–8.

[75] T. Nakatani, S. Li, T. Ura, A. Bodenmann, and T. Sakamaki, "3D visual modeling of hydrothermal chimneys using a rotary laser scanning system," *2011 IEEE Symposium on Underwater Technology and Workshop on Scientific Use of Submarine Cables and Related Technologies*, pp. 1–5, 4 2011.

[76] J. Liu, A. Jakas, A. Al-Obaidi, and Y. Liu, "Practical issues and development of underwater 3D laser scanners," in *2010 IEEE 15th Conference on Emerging Technologies & Factory Automation (ETFA 2010)*. IEEE, 9 2010, pp. 1–8.

[77] S. Smart Light Devices, "3DLS Underwater 3D Laser Imaging Scanner," http://www.smartlightdevices.co.uk/products/3dlaser-imaging/, accessed: 2015-09-30.

[78] M. Prats, J. J. Fernandez, and P. J. Sanz, "An approach for semi-autonomous recovery of unknown objects in underwater environments," in *2012 13th International Conference on Optimization of Electrical and Electronic Equipment (OPTIM)*. IEEE, 5 2012, pp. 1452–1457.

[79] M. Prats, J. J. Fernandez, and P. J. Sanz, "Combining template tracking and laser peak detection for 3D reconstruction and grasping in underwater environments," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 10 2012, pp. 106–112.

[80] P. J. Sanz, A. Penalver, J. Sales, D. Fornas, J. J. Fernandez, J. Perez, and J. Bernabe, "GRASPER: A Multisensory Based Manipulation System for Underwater Operations," *2013 IEEE International Conference on Systems, Man, and Cybernetics*, pp. 4036–4041, 10 2013.

[81] C. Cain and a. Leonessa, "Laser based rangefinder for underwater applications," *2012 American Control Conference (ACC)*, pp. 6190–6195, 6 2012.

[82] M. Caccia, "Laser-Triangulation Optical-Correlation Sensor for ROV Slow Motion Estimation," *IEEE Journal of Oceanic Engineering*, vol. 31, no. 3, pp. 711–727, 7 2006.

[83] Y. Yang, B. Zheng, and H. Zheng, "3D reconstruction for underwater laser line scanning," in *Proceedings of MTS/IEEE Oceans*, 2013, pp. 2008–2010.

[84] A. Bodenmann, B. Thornton, and T. Ura, "Generation of High-resolution Three-dimensional Reconstructions of the Seafloor in Color using a Single Camera and Structured Light," *Journal of Field Robotics*, vol. 34, no. 5, pp. 833–851, 2017. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/rob.21682

[85] C. Tsiotsios, M. Angelopoulou, T.-K. Kim, and A. Davison, "Backscatter compensated photometric stereo with 3 sources," in *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, June 2014, pp. 2259–2266.

[86] AgiSoft, "PhotoScan Professional (Version 1.2.6) (Software)." 2016. [Online]. Available: http://www.agisoft.com/downloads/installer/

[87] A. Jordt-Sedlazeck, K. Koser, and R. Koch, "3D reconstruction based on underwater video from ROV Kiel 6000 considering underwater imaging conditions," in *Proceedings of MTS/IEEE Oceans*. IEEE, 5 2009, pp. 1–10.

[88] O. Pizarro, R. M. Eustice, and H. Singh, "Large Area 3-D Reconstructions From Underwater Optical Surveys," *IEEE Journal of Oceanic Engineering*, vol. 34, pp. 150–169, 2009.

[89] A. Meline, J. Triboulet, and B. Jouvencel, "Comparative study of two 3D reconstruction methods for underwater archaeology," *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 740–745, 10 2012.

[90] D. McKinnon, H. He, B. Upcroft, and R. Smith, "Towards automated and in-situ, near-real time 3-d reconstruction of coral reef environments," in *OCEANS 2011*, Sept 2011, pp. 1–10.

[91] A. Jordt-Sedlazeck and R. Koch, "Refractive Structure-from-Motion on Underwater Images," *2013 IEEE International Conference on Computer Vision*, pp. 57–64, 12 2013.

[92] S. Cocito, S. Sgorbini, a. Peirano, and M. Valle, "3-D reconstruction of biological objects using underwater video technique and image processing," *Journal of Experimental Marine Biology and Ecology*, vol. 297, no. 1, pp. 57–70, 12 2003.

[93] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, ISBN: 0521540518, 2004.

[94] N. S. Ku3 and R. Kumar P., "Design & development of autonomous system to build 3D model for underwater objects using stereo vision technique," in *2011 Annual IEEE India Conference*, vol. 1. IEEE, 12 2011, pp. 1–4.

[95] P. Jasiobedzki, S. Se, M. Bondy, and R. Jakola, "Underwater 3D mapping and pose estimation for ROV operations," in *Proceedings of MTS/IEEE Oceans*. IEEE, 2008, pp. 1–6.

[96] P. Nurtantio Andono, E. Mulyanto Yuniarno, M. Hariadi, and V. Venus, "3D reconstruction of under water coral reef images using low cost multi-view cameras," in *2012 International Conference on Multimedia Computing and Systems*. IEEE, 5 2012, pp. 803–808.

[97] V. Schmidt and Y. Rzhanov, "Measurement of micro-bathymetry with a gopro underwater stereo camera pair," in *Oceans, 2012*, Oct 2012, pp. 1–6.

[98] T. D. Dao, "Underwater 3D reconstruction from stereo images," Msc Erasmus Mundus in Vision and Robotics, University of Girona (Spain), University of Burgundy (France), Heriot Watt University (UK), 2008, available at https://fb.docs.com/VA95.

[99] V. Brandou, a. G. Allais, M. Perrier, E. Malis, P. Rives, J. Sarrazin, and P. M. Sarradin, "3D Reconstruction of Natural Underwater Scenes Using the Stereovision System IRIS," in *Proceedings of MTS/IEEE Oceans*. IEEE, 6 2007, pp. 1–6.

[100] C. Beall, B. J. Lawrence, V. Ila, and F. Dellaert, "3D reconstruction of underwater structures," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 0448111. IEEE, 10 2010, pp. 4418–4423.

[101] J. Servos, M. Smart, and S. L. Waslander, "Underwater stereo SLAM with refraction correction," *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3350–3355, 11 2013.

[102] A. Hogue, A. German, and M. Jenkin, "Underwater environment reconstruction using stereo and inertial data," in *Systems, Man and Cybernetics, 2007. ISIC. IEEE International Conference on*, Oct 2007, pp. 2372–2377.

[103] P. L. Negre Carrasco, F. Bonin-Font, and G. Oliver-Codina, "Stereo Graph-SLAM for Autonomous Underwater Vehicles," in *Proceedings of the 13th International Conference*

on Intelligent Autonomous Systems (IAS 2014), E. M. et at., Ed. Springer International Publishing, 2014, pp. 351–360.

[104] G. Inglis and C. Roman, "Sonar constrained stereo correspondence for three-dimensional seafloor reconstruction," in *Proceedings of MTS/IEEE Oceans*. IEEE, 5 2010, pp. 1–10.

[105] "Echoscope 4g, coda octopus," https://www.codaoctopus.com/products/3d/echoscope4G, accessed: 2019-07-12.

[106] "Blueview multibeam scanning sonar, teledyne," http://www.teledynemarine.com/blueview-3d-multibeam-scanning-sonar/?BrandID=3, accessed: 2019-07-12.

[107] Teledyne CDL, "INSCAN 3D Scanning Subsea Laser," http://teledyne-cdl.com/events/inscan-demonstration-post-press-release, accessed: 2015-09-30.

[108] S. Hannon, "Underwater Mapping," *LiDAR Magazine*, vol. 3, no. 1, pp. 1–4, 2013.

[109] 2G Robotics, "ULS-100 Underwater Laser Scanner for Short Range Scans," http://www.2grobotics.com/products/underwater-laser-scanner-uls-100/, pp. 1–2, accessed: 2015-09-30.

[110] 2G Robotics, "ULS-200 Underwater Laser Scanner for Mid Range Scans," http://www.2grobotics.com/products/underwater-laser-scanner-uls-200/, accessed: 2015-09-30.

[111] 2G Robotics, "ULS-500 Underwater Laser Scanner for Long Range Scans," http://www.2grobotics.com/products/underwater-laser-scanner-uls-500/, pp. 2–3, accessed: 2015-09-30.

[112] Tritech, "Sea Stripe ROV/AUV Laser Line Generator," http://www.tritech.co.uk/product/rov-auv-laser-line-generator-seastripe, accessed: 2015-09-30.

[113] C. Roman, G. Inglis, and J. Rutter, "Application of structured light imaging for high resolution mapping of underwater archaeological sites," in *Proceedings of MTS/IEEE Oceans*. IEEE, 5 2010, pp. 1–9.

[114] M. Massot-Campos and G. Oliver-Codina, "Evaluation of a laser based structured light system for 3D reconstruction of underwater environments," in *Martech 2013 5th International Workshop on Marine Technology*. Girona: SARTI, 2013, p. 63.

[115] J. Forest Collado, J. Salvi, and E. Cabruja, "Laser stripe peak detector for 3D scanners. A FIR filter approach," in *Proceedings of the 17th International Conference on Pattern Recognition (ICPRâĂŹ04)*, 2004, pp. 2–5.

[116] R. Fisher and D. Naidu, "A comparison of algorithms for subpixel peak detection," *Advances in Image Technology*, 1996.

[117] G. Inglis, C. Smart, I. Vaughn, and C. Roman, "A pipeline for structured light bathymetric mapping," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, no. Figure 1. IEEE, 10 2012, pp. 4425–4432.

[118] W. Brink, A. Robinson, and M. Rodrigues, "Indexing Uncoded Stripe Patterns in Structured Light Systems by Maximum Spanning Trees," in *British Marine Vision Conference*, 2008.

[119] "CloudCompare: 3d point cloud and mesh processing software," http://www.cloudcompare.org/, accessed: 2017-03-20.

[120] J. Melo and A. Matos, "Survey on advances on terrain based navigation for autonomous underwater vehicles," *Ocean Engineering*, vol. 139, no. September 2015, pp. 250–264, 2017. [Online]. Available: http://dx.doi.org/10.1016/j.oceaneng.2017.04.047

[121] C. Roman and H. Singh, "A Self-Consistent Bathymetric Mapping Algorithm," *Journal of Field Robotics*, vol. 24, no. 1-2, pp. 23–50, jan 2007. [Online]. Available: http://doi.wiley.com/10.1002/rob.20164

[122] A. Palomer, P. Ridao, and D. Ribas, "Multibeam 3D underwater SLAM with probabilistic registration," *Sensors (Switzerland)*, vol. 16, no. 4, 2016.

[123] G. Vallicrosa and P. Ridao, "H-SLAM: Rao-Blackwellized particle filter SLAM using Hilbert Maps," *Sensors (Switzerland)*, vol. 18, no. 5, 2018.

[124] R. Kummerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard, "G2o: A general framework for graph optimization," in *2011 IEEE International Conference on Robotics and Automation.* IEEE, may 2011, pp. 3607–3613. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5979949

[125] T. Moore and D. Stouch, "A generalized extended kalman filter implementation for the robot operating system," in *Proceedings of the 13th International Conference on Intelligent Autonomous Systems (IAS-13).* Springer, July 2014.

[126] B. M. Yu, K. V. Shenoy, and M. Sahani, "Derivation of Extended Kalman Filtering and Smoothing Equations," University of Stanford, Department of Electrical Engineering, Tech. Rep., 2004.

[127] S. Zandara, P. Ridao, A. Mallios, and D. Ribas, "MBpIC-SLAM: Probabilistic Surface Matching for Bathymetry Based SLAM," in *IFAC Workshop on Navigation, Guidance and Control of Underwater Vehicles (NGCUV'2012)*, no. 2007, 2012, pp. 1–6.

[128] R. Smith, M. Self, and P. Cheeseman, "Estimating uncertain spatial relationships in robotics," in *Proceedings. 1987 IEEE International Conference on Robotics and Automation*, vol. 4, March 1987, pp. 850–850.

[129] Z. C. Marton, R. B. Rusu, and M. Beetz, "On fast surface reconstruction methods for large and noisy point clouds," in *2009 IEEE International Conference on Robotics and Automation*, May 2009, pp. 3218–3223.

[130] A. Segal, D. Hähnel, and S. Thrun, "Generalized-ICP," in *Proc. of Robotics: Science and Systems*, 2009.

[131] Z. Chen, "Bayesian Filtering: From Kalman Filters to Particle Filters, and Beyond," *Statistics*, vol. 182, 2003.

[132] A. Ortiz, J. Antich, and G. Oliver, "A particle filter-based approach for tracking undersea narrow telecommunication cables," *Machine Vision and Applications*, vol. 22, no. 2, pp. 283–302, 2011.

[133] G. Vallicrosa, A. Palomer, D. Ribas, and P. Ridao, "Realtime AUV Terrain Based Navigation with Octomap in a Natural Environment," in *ROBOT2013: First Iberian Robotics Conference*, vol. 252, 2014, pp. 41–53. [Online]. Available: http://link.springer.com/10.1007/978-3-319-03413-3

[134] S. A. Barkby, S. B. Williams, O. Pizarro, and M. V. Jakuba, "An efficient approach to bathymetric SLAM," in *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*, 2009, pp. 219–224.

[135] S. A. Barkby, "Efficient and Featureless Approaches to Bathymetric Simultaneous Localisation and Mapping," Ph.D. dissertation, 2011. [Online]. Available: http://ses.library.usyd.edu.au/handle/2123/7919

[136] Schmidt Ocean Institute Cruise., "Adaptive Robotics at Barkley Canyon and Hydrate Ridge," https://schmidtocean.org/cruise/adaptive-robotics-at-barkley-canyon-and-hydrate-ridge/, accessed: 2018-09-30.

[137] M. Leat, "Analysis of measurement uncertainty in high-resolution laser-scanned bathymetric measurements of seafloor topology using an autonomous underwater vehicle," Master's thesis, University of Southampton - United Kingdom, 2018.