

**FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO**



# **Tri-dimensional Reconstruction of Submerged Objects through an Active Eye-in-Hand Perception System**

**João Pedro da Silva Esteves**

Mestrado em Engenharia Eletrotécnica e de Computadores

Supervisor: Andry Maykol Gomes Pinto

October 13, 2022





# Resumo

A exploração dos ambientes marítimos, bem como os seus recursos, tem um grande valor económico cuja existência não segue indetetável nos dias de hoje. Por esse motivo, há atualmente uma grande aposta na expansão da indústria marítima, em particular em tarefas de operação e manutenção (O&M) para apoiar a produção de energia renovável em alto mar.

Soluções autónomas, como Veículos Autónomos Subaquáticos (AUVs) e Veículos Autónomos de Superfície (ASVs), tornam-se tecnologias importantes para realizar atividades de O&M. Por contribuírem para a análise do ambiente em redor, os sistemas de perceção visual definem os níveis de autonomia das embarcações, apesar de enfrentarem desafios complexos provenientes dos fenómenos físicos que ocorrem entre as interações luz e água, que degradam significativamente os dados adquiridos. No entanto, quando a intervenção é necessária, perceção tridimensional completa e precisa do ambiente revela-se insuficiente. Manipuladores robóticos, agora muitas vezes incluídos em veículos autónomos, tornam-se uma extensão da embarcação e possibilitam a interação com o ambiente, reduzindo drasticamente a necessidade de intervenção humana bem como consequências de erros inerentes a humanos.

A combinação de um sistema de perceção visual tridimensional com um manipulador robótico é o principal tema de interesse desta dissertação, em particular numa abordagem *Eye-in-Hand*. Esta metodologia é alcançada com a fixação do sistema de perceção no próprio manipulador, de modo a explorar os seus graus de liberdade adicionais para ajustar, com maior precisão, a distância entre o sistema de perceção e o alvo a capturar.

Esta dissertação utiliza a configuração *Eye-in-Hand* para realizar reconstruções 3D de objetos submersos, através de um sistema de perceção híbrido, formado pelo módulo de perceção ativo, composto por dois lasers, e pelo módulo de perceção passivo, composto por duas câmaras numa configuração estéreo. A abordagem desenvolvida consiste num algoritmo de deteção de objetos para detetar a posição do objeto no meio envolvente, num algoritmo de planeamento e execução de trajetória do manipulador que calcula pontos de vista interessantes do objeto e, por fim, um algoritmo de aquisição e processamento de dados que captura pontos tridimensionais nos vários pontos de vista, através de uma versão simulada do sistema híbrido de perceção, que resulta em duas nuvens de pontos distintas, em cada reconstrução, com propriedades diferentes.

A análise quantitativa do sistema foi descrita com métricas como distância de *Chamfer* e percentagem de conclusão da *point cloud*. Os resultados das experiências demonstram que a abordagem que seleciona vários pontos de vista ao redor do alvo atinge uma maior cobertura da superfície do objeto. Com esta abordagem, foram obtidas *point clouds* com distância de *Chamfer* média de 0.0107 m e 37.5% de percentagem de conclusão, utilizando o módulo de perceção ativo sem ruído, bem como distância de *Chamfer* de 0.0293 m e 59.5% de conclusão de *point cloud*, usando o módulo de perceção passivo sem ruído. Embora o módulo passivo atinge maiores percentagens de conclusão de *point cloud*, o módulo ativo apresenta um erro significativamente menor na aquisição de informação tridimensional.



# Abstract

The exploitation of the maritime environments, as well as its resources, have a high economic value whose existence is no longer underestimated in today's world. For that reason, a big expansion in the field of maritime industry, in particular for Operations and Maintenance (O&M) tasks to support green energy production in an offshore setting, is ongoing.

Autonomous solutions, such as Autonomous Underwater Vehicles (AUVs) and Autonomous Surface Vehicles (ASVs), are now becoming important technologies to perform such O&M activities. Visual perception systems greatly affect these vessels' autonomy levels, due to their contributions to situational awareness, although facing complex challenges regarding the physical phenomena that occur with light and water interactions which degrade acquired data. However, when intervention is necessary, a complete and accurate tridimensional perception of the surroundings is not enough. Robotic manipulators, now often included in autonomous vehicles, can operate as an extension of the vessel and bring the ability to interact with the environment, thus drastically reducing human intervention in dangerous settings as well as avoiding consequences of inherent human errors.

The combination of a tridimensional visual perception system with a robotic manipulator is the main topic of interest of this research, in particular an Eye-in-Hand approach. This methodology is achieved by attaching the perception system to the manipulator, in order to explore its additional degrees-of-freedom for better control of the distance between the perception system and the target.

This dissertation uses the Eye-in-Hand configuration to perform 3D reconstructions of submerged objects, using an hybrid perception system, formed by the active perception module, composed of two lasers, and the passive perception module, composed of two cameras in a stereo configuration. The developed approach consists of an object detection algorithm to detect the object's position in the surrounding space, a path planning and execution algorithm that controls the movement of the manipulator in order to achieve interesting viewpoints of the object and, finally, a data acquisition and processing node that captures tridimensional points from one or multiple viewpoints, using a simulated version of the hybrid perception system, which results in two distinct point clouds.

A quantitative analysis of the system's performance was accomplished through metrics such as Chamfer distance and point cloud completion. The results of the experiments demonstrate that, using an approach that selects several viewpoints around the target, it is possible to obtain a greater object surface coverage. With this approach, the resulting evaluation on point clouds got an average Chamfer distance of 0.0107 m and point cloud completion of 37.5%, using the noiseless active perception module, and an average Chamfer distance of 0.0293 m and point cloud completion of 59.5%, using the noiseless passive perception module. Although the passive module allows for greater point cloud completion, the active module has a significant lower error in acquisition of three-dimensional information.



# Agradecimentos

Ao meu orientador, Prof. Dr. Andry Maykol Pinto, que me deu a oportunidade de trabalhar na dissertação com a equipa inovadora do CRAS, agradeço toda a orientação, paciência e *feedback* dado que fizeram toda a diferença na concretização desta dissertação. Ao meu co-orientador, Pedro Nuno Leite, pela total disponibilidade para me auxiliar com os problemas que iam surgindo e também pela importante orientação na escrita.

Aos meus amigos da faculdade, um agradecimento a todos por formarem comigo um grupo de amigos com quem posso descontraír, conviver e me divertir, sempre que preciso.

Aos meus pais e irmão, pelo constante apoio que me ajudou a atingir os meus objetivos. São os meus maiores motivadores e espero que um dia possa ser eu a concretizar alguns dos vossos sonhos.

Termino com um agradecimento aos professores por toda a transmissão de conhecimento e pelo importante contributo na formação dos melhores engenheiros possíveis.

João Pedro da Silva Esteves



# Institucional Acknowledgements

This work was carried out with the support of the host institution INESC TEC and supervised at the institution by Pedro Nuno Barbosa Leite.

This work is co-financed by the ERDF – European Regional Development Fund through the Operational Programme for Competitiveness and Internationalisation - COMPETE 2020 under the PORTUGAL 2020 Partnership Agreement, and through the Portuguese National Innovation Agency (ANI) as a part of project NESSIE: POCI-01-0247-FEDER-039817.

João Pedro da Silva Esteves





*“The key to realizing a dream is to focus not on success but significance – and then even the small steps and little victories along your path will take on greater meaning.”*

Oprah Winfrey



# Contents

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Introduction</b>  | <b>1</b>  |
| 1.1      | Context and Motivation . . . . .   | 1         |
| 1.2      | Objectives . . . . .   | 3         |
| 1.3      | Structure . . . . .  | 3         |
| <b>2</b> | <b>Literature review</b>   | <b>5</b>  |
| 2.1      | The Challenges in Underwater Perception . . . . .                        | 5         |
| 2.2      | Underwater Perception Systems and 3D Reconstruction Techniques . . . . . | 8         |
| 2.2.1    | Passive Perception Systems . . . . .                                     | 9         |
| 2.2.2    | Active Perception Systems . . . . .                                      | 13        |
| 2.2.3    | Hybrid Perception Systems . . . . .                                      | 18        |
| 2.3      | Eye-in-Hand Approaches . . . . .   | 19        |
| 2.4      | Critical Analysis . . . . .  | 24        |
| <b>3</b> | <b>A 3D Reconstruction System for Submerged Objects</b>                  | <b>27</b> |
| 3.1      | Introduction . . . . .   | 27        |
| 3.1.1    | Simulation Setup . . . . .   | 27        |
| 3.1.2    | Manipulator Reach Bravo 7 . . . . .                                      | 28        |
| 3.1.3    | MARESyE Perception System . . . . .                                      | 29        |
| 3.2      | 3D Reconstruction Methodology . . . . .                                  | 31        |
| <b>4</b> | <b>Results</b>   | <b>39</b> |
| 4.1      | Introduction . . . . .   | 39        |
| 4.2      | 3D Reconstruction Evaluation . . . . .                                   | 40        |
| 4.3      | Experimental Results . . . . .   | 41        |
| 4.3.1    | Analysis of Resulting Point Clouds . . . . .                             | 42        |
| 4.3.2    | Reconstructed Meshes . . . . .   | 49        |
| 4.3.3    | Conclusions . . . . .  | 51        |
| <b>5</b> | <b>Conclusions and Future Work</b>                                       | <b>53</b> |
|          | <b>References</b>  | <b>55</b> |



# List of Figures

|      |  |    |
|------|--|----|
| 2.1  | Absorption coefficients of the spectrum of light in the water medium. Extracted from [14]. . . . .   | 6  |
| 2.2  | Energy values received by the sensor (camera) throughout time portraying the scattering problem and range-gate solution. Extracted from [17]. . . . .  | 7  |
| 2.3  | Difference of light rays' refraction on a flat enclosure (left) and on a dome-shaped enclosure (right). Extracted from [20]. . . . .   | 9  |
| 2.4  | The compact and portable AQUASENSOR device with its housing to its right. Extracted from [22]. . . . .   | 10 |
| 2.5  | ROV with a stereo camera setup composed by two enclosed cameras and a main piloting camera. Extracted and adapted from [28]. . . . .   | 11 |
| 2.6  | Stereo correspondence based on triangulation for 3D point computation. Extracted from [29]. . . . .  | 12 |
| 2.7  | Disparity map result from the adapted SGM (left); 3D reconstruction result (right). Extracted and adapted from [31]. . . . .   | 12 |
| 2.8  | Optical system composed of two laser beam sensors, laterally placed to a SXGA Basler camera, inside watertight enclosures. Extracted from [38]. . . . .  | 14 |
| 2.9  | Input captured image of a pipe with projected parallel lines (left); top view (middle) and lateral view (right) of resulted 3D reconstruction of the pipe. Extracted from [40]. . . . .  | 15 |
| 2.10 | Model of an optical scanner system composed of a camera and line laser, mounted on top of a rotating structure. Extracted and adapted from [44]. . . . .   | 16 |
| 2.11 | 3D laser scanner mounted on a tripod (left); resulting point cloud coloured by depth (right). Extracted from [46]. . . . .   | 17 |
| 2.12 | ToF sensors principle of operation. Extracted from [47]. . . . .   | 17 |
| 2.13 | Triangulation principle on pattern projection methods. Extracted from [48]. . . . .  | 18 |
| 2.14 | MARESyE concept model represented by a watertight enclosure containing the PS module, composed of 2 cameras and LEDs, and the LSR module, composed of a red and green coloured laser sensor. Extracted and adapted from [10]. . . . .                | 19 |
| 2.15 | One of the objects used for laboratory testing (left); resulting 3D reconstruction using PS module, in low turbidity conditions (middle) and resulting 3D reconstruction using LSR module, in low turbidity conditions. Extracted from [10]. . . . . | 19 |
| 2.16 | Recovery operation: SAUVIM AUV, with robotic arm and sensory unit attached, navigating towards the target (left); manipulator detected the object and begins to grasp it (right). Extracted and adapted from [50]. . . . .                           | 21 |
| 2.17 | Configuration of the arm, gripper and camera, during development phase. Extracted and adapted from [51]. . . . .   | 21 |
| 2.18 | Configuration of the arm, laser emitter and camera for object scanning. Extracted from [53]. . . . .   | 22 |

|      |   |    |
|------|---|----|
| 2.19 | Laser scanning of an amphora (left); projected image in camera view (middle); side view of the generated raw point cloud of the amphora scan (right). Extracted from [53]. . . . .  | 22 |
| 2.20 | Eye-in-Hand configuration, with the camera and laser stripe emitter attached to the Arm 5E manipulator. Extracted from [55]. . . . .  | 23 |
| 2.21 | Skull as the simulated target object for reconstruction (left); result of the reconstruction using one-view approach (middle); result of the reconstruction using proposed multi-view approach (right). Extracted from [55]. . . . .  | 23 |
| 2.22 | Custom end-effector with grasping abilities (left) with embedded camera (right). Extracted from [56]. . . . .   | 24 |
| 3.1  | Proposed pipeline for the complete 3D reconstruction process. . . . .   | 28 |
| 3.2  | Gazebo setup used for simulation. . . . .   | 28 |
| 3.3  | Reach Bravo 7, extracted and adapted from 11 <sup>1</sup> (left), and simulated model in Gazebo (right), with arrows representing corresponding revolute joints rotations. . . . .  | 29 |
| 3.4  | Simulated model of the MARESyE hybrid imaging system in Gazebo. . . . .   | 30 |
| 3.5  | Operation of the perception system: modules' projections on the scenario's floor when no object is being captured (left); modules' projection on a cube object (right). White points correspond to data being captured by the PS module while red and green points correspond to data being captured by the LSR module. . . . . | 31 |
| 3.6  | Diagram of the object detection algorithm. . . . .  | 32 |
| 3.7  | Manipulator in initial position with cage's reachable surface and perception system introduced as collision elements in the surrounding planning scene. . . . .   | 33 |
| 3.8  | Manipulator's movement throughout detection path, repeated at every 45° increment, until reaching a complete 270° rotation of the manipulators' base footprint joint. . . . .   | 33 |
| 3.9  | Overview diagram of the path planning and execution algorithm. . . . .  | 34 |
| 3.10 | Desired forward, right and up vector representations so that the camera directly looks at a target object. . . . .  | 35 |
| 3.11 | Interpolated (red) and selected (green) viewpoints from a vertical plane containing the object's position and the manipulator's z-axis (top); interpolated from the surface of 3 cylinders of increasing radius, centred on the target, with desired number of final viewpoints equal to three (bottom). . . . .                | 36 |
| 3.12 | Scan sweep. . . . .   | 37 |
| 3.13 | Overview diagram of the 3D data acquisition & processing algorithm. . . . .   | 37 |
| 3.14 | Captured PS point clouds before being box filtered (left); box filtered PS point cloud (right). . . . .   | 38 |
| 4.1  | Three different target objects selected for 3D reconstruction experimentations. . . . .   | 39 |
| 4.2  | ICP alignment effect on a resulting point cloud from the PS module, in red, with object's template point cloud, in white, as a target. The aligned point clouds (right) will lead to a better application of quantitative metrics for 3D reconstruction evaluation. . . . .   | 41 |
| 4.3  | Viewpoints generated in each 3D reconstruction experiment, represented as green arrows, looking at the object's detected position. . . . .  | 43 |
| 4.4  | LSR-obtained point clouds, with no noise, using first approach (left) and second approach (right) of a treasure chest (top), anchor (middle) and coral (bottom) type objects. . . . .   | 44 |

|     |   |    |
|-----|---|----|
| 4.5 | PS-obtained point clouds, with no noise, using first approach (left) and second approach (right) of a treasure chest (top), anchor (middle) and coral (bottom) type objects. . . . .  | 46 |
| 4.6 | PS-obtained point clouds, with noise considered, using first approach (left) and second approach (right) of a treasure chest (top), anchor (middle) and coral (bottom) type objects. . . . .  | 48 |
| 4.7 | Meshes computed from LSR-obtained point clouds, with no noise, using a first path planning approach (top row) and second path planning approach (bottom row) of a treasure chest (left), anchor (middle) and coral (right) type objects. . . .  | 50 |
| 4.8 | Meshes computed from PS-obtained point clouds, with no noise, using a first path planning approach (top row) and second path planning approach (bottom row) of a treasure chest (left), anchor (middle) and coral (right) type objects. . . . . | 51 |
| 4.9 | Matching points' distances of the template point cloud to the obtained LSR (left) and PS (right) point clouds using first (top) and second (bottom) path planning approach, of an anchor type object. . . . .                                   | 52 |





# List of Tables

|     |   |    |
|-----|---|----|
| 3.1 | Reach Bravo 7 properties. . . . .   | 29 |
| 3.2 | LSR simulation default parameters. . . . .  | 30 |
| 3.3 | PS simulation default parameters. . . . .   | 31 |
| 4.1 | Euclidean distance errors of the three detected object's position in experiments with three different values for PS noise standard deviation (std) . . . . .  | 42 |
| 4.2 | Metrics results on 3D reconstructions of the three different objects, using the LSR active perception module and no noise considerations, with first and second viewpoint interpolation approaches. . . . . | 45 |
| 4.3 | Metrics results on 3D reconstructions of the three different objects, using the PS active perception module and no noise considerations, with first and second viewpoint interpolation approaches. . . . .  | 47 |
| 4.4 | Metrics results on 3D reconstructions of the three different objects, using the PS active perception module and added noise, with first and second viewpoint interpolation approaches. . . . .              | 49 |



# Acronyms and Symbols

|       |  |
|-------|--|
| 2D    | Two-Dimensional  |
| 3D    | Three-Dimensional                                      |
| ASV   | Autonomous Surface Vehicle                             |
| AUV   | Autonomous Underwater Vehicle                          |
| AUVSI | Association for Unmanned Vehicle Systems International |
| CCD   | Charge-coupled devices                                 |
| CD    | Chamfer distance                                       |
| CLAHE | Contrast Limited Adaptive Histogram Equalization       |
| DOE   | Degrees-of-Freedom                                     |
| DNN   | Deep Neural Network                                    |
| EiH   | Eye-in-Hand  |
| FoV   | Field-of-View  |
| HSV   | Hue-Saturation-Value                                   |
| I-AUV | Intervention - Autonomous Underwater Vehicle           |
| ICP   | Iterative Closest Point                                |
| LED   | Light Emitting Diode                                   |
| LiDAR | Light Detection And Ranging                            |
| LSR   | Light Stripe Ranging                                   |
| NTU   | Nephelometric Turbidity Units                          |
| O&M   | Operation & Maintenance                                |
| OMPL  | Open Motion Planning Library                           |
| PCL   | Point Cloud Library                                    |
| PS    | Photometric Stereo                                     |
| RADAR | Radio Detection and Ranging                            |
| RGB   | Red-Green-Blue   |
| ROS   | Robot Operating System                                 |
| ROV   | Remotely Operated Vehicle                              |
| RVIZ  | ROS visualizer   |
| SAD   | Standard Absolute Differences                          |
| SfM   | Structure-from-Motion                                  |

|        |  |
|--------|--|
| SGM    | Semi-Global Matching                               |
| SLS    | Structured Light System                            |
| SLS-LP | Structured Light System - Line Projection          |
| SNR    | Signal-Noise-Ratio                                 |
| SONAR  | Sound Navigation and Ranging                       |
| SSIM   | Structural Similarity Index Measurements           |
| ToF    | Time-of-Flight                                     |
| UTOFIA | Underwater Time Of Flight Image Acquisition System |
| UNION  | UNderwater Intelligent Operation and Navigation    |

# Chapter 1

## Introduction

### 1.1 Context and Motivation

The maritime environment and its resources have recently been garnering quite a lot of attention from several industries. All economic activities related to oceans, seas and coasts can be grouped into what is called "Blue Economy", though this term varies among organizations. Activities such as shipping, aquaculture, renewable energy production, ports, shipyards, production, and research contribute to this massive potential of economic power <sup>1</sup>. Particularly, offshore renewable energy production is becoming increasingly more important, due to high depletion of natural reserves of fossil fuels as energy sources, being wind power the most mature and low cost type of technology. When it comes to onshore and offshore wind power, the latter, being produced in the ocean or on the offshore islands, is the preferred solution for many countries with coastlines, since its main advantage is the abundant wind energy resources, though more expensive to build [1].

Innovative robotic solutions for maritime operations go back as early as 1997 with the fully teleoperated manipulation system, developed under project UNION (UNderwater Intelligent Operation and Navigation) and presented by V. Rigaud *et al.* [2]. This system, composed by the VORTEX Remotely Operated Vehicle (ROV) with the 7 degrees-of-freedom (DOF) Pa10 robotic arm, was mainly used to perform tasks that involve underwater manipulation and navigation. With the increased interest in maritime resources and potential economic value, more industries are investing in research that promotes more sustainable solutions and increased autonomy levels, specially for tasks of operations and maintenance (O&M) in offshore infrastructures, such as the DIIUS project [3] (2021), that aims to improve perceptual capabilities of robotic systems by proposing a distributed perception solution for inspection of aquatic infrastructures, and the ATLANTIS project [4] (2021), that promotes automatization of offshore operations by providing a pioneer pilot infrastructure that allows for testing, validation and demonstration of new robotic solutions.

Along the years, technological advances in obstacle avoidance algorithms [5, 6] and docking approaches [7, 8] improved the performance of Autonomous Surface Vehicles (ASVs) and Autonomous Underwater Vehicles (AUVs) in many sea related missions. These vessels represent

---

<sup>1</sup><https://unric.org/en/blue-economy-oceans-as-the-next-great-economic-frontier/>

a great solution to substitute human labour in tasks that could potentially be dangerous, such as deep ocean exploration and environmental monitoring, O&M of offshore platforms or ship hull inspection. In fact, according to X. Chen *et al.* (2021) [9], AUVs have come quite close to reach a fully autonomous level in order to completely execute missions without any human intervention, marking an important positive progress mostly due to the fact that, from 1996 to 2000, most of the failures of AUVs were direct causes of human errors. The authors reviewed risk analysis research regarding operations of AUVs and created a complete table identifying human factors that cause AUV failures, such as the operator's ability to timely identify errors and critical situations during missions, human fatigue, communication effectiveness and the level of the operator's belief in the autonomous capabilities of the AUV.

According to an article written in the website of the Association for Unmanned Vehicle Systems International <sup>2</sup> (AUVSI), a "*successfully integrated AUV must be adaptable to varying environmental and internal state changes, with top consideration being accuracy, bias and complexity*". To accomplish all the essential aspects that an AUV must follow, a perception system is inevitable to contribute for situational awareness, one of the human factors that lead to the AUV failure, and adaptation of the vehicle for different missions and behaviour parameters. The usage of perception systems in aerial or terrestrial environments has been largely exploited over the last decades, but not much research has been done with the integration of these systems in an underwater context. One of the topics of this dissertation refers to the challenges that these systems face when operating in subsea conditions, mainly absorption and scattering while light propagates in the water medium, and how to improve the performance of the optical system for more accurate and precise results. Additionally, to further emulate human's interaction capability with the surroundings, a manipulator is also often needed to be integrated, particularly when intervention tasks on underwater structures are required, which is why this dissertation's proposed approach could be an interesting solution for applications that require both manipulators and perception systems.

This dissertation proposes an Eye-in-Hand (EiH) configuration approach to perform tridimensional (3D) reconstructions of objects in an underwater context. This type of configuration, obtained by rigidly mounting the perception system on the manipulator, allows to explore the extra degrees of freedom of the robotic arm, which should result in more complete and faithful reconstructions of target objects. This methodology has applications on operations such as ocean exploration, underwater inspection and offshore O&M. The perception system used is MARESys [10], developed by A. Pinto et al, which is composed by two modules, the photometric stereo (PS) and multiple laser stripe range (LSR), both equipped with technology that reduces the impact of underwater imaging challenges. The perception system was attached to the Reach Bravo 7, a manipulator specially designed for inspection and intervention applications.

The proposed algorithm includes an initial detection of the target's position followed by a path planning and execution algorithm that iterates and moves through several viewpoints, where a scan sweep is performed. During the scan sweep, a 3D data acquisition algorithm is triggered, which captures point clouds from both PS and LSR modules. These point clouds are compensated

---

<sup>2</sup><https://www.auvsi.org/industry-news/industry-news/assured-onboard-autonomy-architecture-auvs>

with the manipulators movement between triggers and combined with point clouds from previous iterations. This algorithm finishes with point cloud box filtering and an outlier removal process. In order to evaluate the performance of the reconstruction, an Iterative Closest Point (ICP) algorithm is used to align the obtained point clouds with a template. For a quantitative analysis, two metrics are used. The 3D Chamfer distance (CD) defines how different a source and target point cloud are and is computed using the euclidean distances between matching points. The last metric is point cloud completion, which is the ratio between total number of matching points in the template's point cloud and obtained point clouds, and total number of points in the template's point cloud.

## 1.2 Objectives

The dissertation's main goal is to study the EiH configurations, where a perception system is attached to a manipulator, allowing to take advantage of the additional degrees-of-freedom to acquire multimodal information, such as images or point clouds, with a high signal-to-noise ratio. Therefore, the objectives of the dissertation are as follows:

- Study different state of the art technologies of perception systems and EiH configurations;
- Develop an object detection algorithm that detects the target's position;
- Develop algorithms to plan a suitable trajectory that ensures different viewpoints and to perform acquisition of both dense and sparse tridimensional information, using the PS and LSR modules, respectively;
- Obtain dense and accurate reconstructions of the targets by combining the tridimensional information acquired in different viewpoints;
- Evaluate the obtained reconstructions by comparing it to a template mesh, using metrics such as the 3D Chamfer Distance (CD) and point cloud completion.

## 1.3 Structure

This document is divided in five chapters in which each is divided in a few sections. Chapter 1 is the introduction and it provides the context and motivation to develop this dissertation. Furthermore, it also states the objectives that should be accomplished with this work. Chapter 2 corresponds to the state of the art. It is divided in four sections that cover topics such as underwater imaging technologies and challenges in the field of study of this dissertation, using scientific work as references throughout the text. Chapter 3 presents relevant topics of the developed 3D reconstruction system, such as the setup used for experimentation purposes, alongside the proposed algorithms for object detection, path planning and execution, and 3D data acquisition. In chapter 4, several evaluation metrics used to perform a quantitative analysis on resulting point clouds are presented. Finally, in chapter 5, there is a brief overview of the developed work and what was accomplished, with additional suggestions on ways to improve for future work that could be done.





## Chapter 2

# Literature review

The aim of this chapter is to provide the reader with an overview of what underwater perception encompasses, in addition to applications that benefit from this capability and how the Eye-in-Hand (EiH) configuration can contribute to an improved efficiency of such approaches.

In the maritime industry, most of the emerging autonomous and remote technologies largely exploit robotic perception. According to F. Bonin-Font *et al.* (2015) [11], vision sensors are regarded as one of the most prominent and powerful types of devices in underwater robotics, in terms of temporal and spatial resolution capabilities. However, it is important to take into account the numerous challenges that degrade information retrieved from these visual sensors, e.g. inherent attributes such as range and resolution or poor illumination conditions [12]. The underwater environments present even more challenges to these types of sensors that also need to be tackled. Throughout section 2.1, the phenomena that occur during the interaction between light and water, along with their impact on the 3D information gathered from the imaging system, are discussed. Underwater imaging systems are introduced in section 2.2, along with methods often used with those systems to extract 3D information. Section 2.3 refers to previous research work that utilizes the EiH approach. Finally, section 2.4 summarizes most of the knowledge resulting from the literature review, thus presenting as a critical analysis of the whole chapter.

### 2.1 The Challenges in Underwater Perception

Perception has been an important tool for many undersea applications, in particular for exploration and intervention tasks, such as maintenance and construction of infrastructures, wreck retrievals, rescue missions or search and identification of submerged objects [11]. Nevertheless, optic sensors, despite being powerful and useful, may have performance issues that are aggravated in an underwater context and that need to be handled. Besides temporal and resolution limitation, poor visibility conditions and lack of surface textures, two of the most prominent challenges occur while light travels through water. These phenomena correspond to the absorption and scattering of light.

**Absorption** is characterized as the attenuation of light as it propagates through the medium, since the electromagnetic radiation progressively loses power [13]. The level of absorption is defined by the absorption coefficient, which is heavily related to the refractive index of the medium and the wavelength of the spectrum of light, as it is portrayed in figure 2.1. The shift in colour towards blue, in underwater scenes, is a consequence of this phenomenon, since longer wavelengths (e.g., red light) are strongly attenuated, in comparison to shorter wavelengths (e.g., blue light) [13, 14, 15]. This means that warmer coloured lights and infrared (IR) electromagnetic waves travel shorter distances, which explains why IR technologies are obsolete in an underwater operational setting.

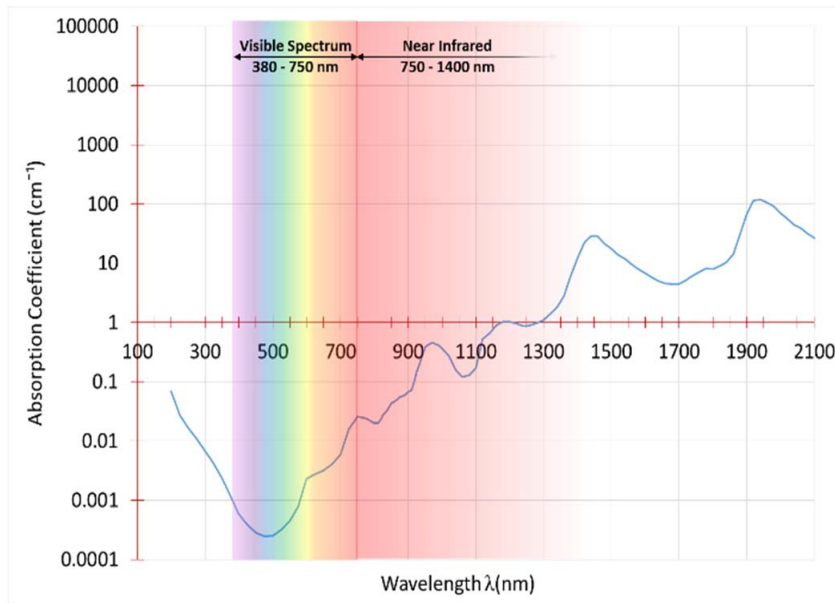


Figure 2.1: Absorption coefficients of the spectrum of light in the water medium. Extracted from [14].

There are many solutions used to correct light absorption's negative effects on resulting images. These include image processing methods, such as image restoration or image enhancement techniques, to fix the colour attenuation caused by absorption [13, 14, 16], or the use of artificial illuminants. The work of I. Vasilescu *et al.* (2010) [15] proposes an artificial illuminant that is used for compensation based on the computed light loss given the distance to the imaged object, and thus not needing any post-processing work on the captured images.

**Scattering** corresponds to the deviation from the projected straight-line path of light propagation which also hinders underwater visibility. This deflection is wavelength independent and occurs due to the presence of suspended particles and organisms randomly distributed in the water, commonly named as "marine snow". This phenomenon can be further classified into forward-scattering or backscattering. Forward-scattering occurs when light is deviated from its trajectory between the scenario and the optical sensor and generally leads to a blurry effect on the image features, whereas backscattering refers to the reflection of projected light on suspended particles,

that occurs before it actually reaches the scenario, limiting the contrast of the resulting image [13].

Post-processing techniques are also common and easy solutions to mitigate scattering effects. The work of A. Khan *et al.* (2016) [14] proposes a wavelet based fusion of images, where the colour fading caused by absorption of light can be corrected through histogram stretching, in the hue-saturation-value (HSV) colour space, and the limitation of contrast caused by scattering can be enhanced using the contrast limited adaptive histogram equalization (CLAHE) technique. Nevertheless, several works propose a more computationally demanding solution to tackle the scattering problem that consists of using a range-gated synchronized receiver approach [17, 18, 10]. Visually represented in figure 2.2, the range-gated approach's principle begins with the accumulation of the reflected light source pulsed from a LED or laser system, in the optical sensor, over a specified time. The camera shutter only opens after a short period of time, when triggering signals are received [10]. This way, it is possible to differentiate between the backscattered noise and the light reflected by the target. It is also possible to reduce the amount of light scattered back from suspended particles by increasing the baseline, which is the distance that separates the light source and the sensor.

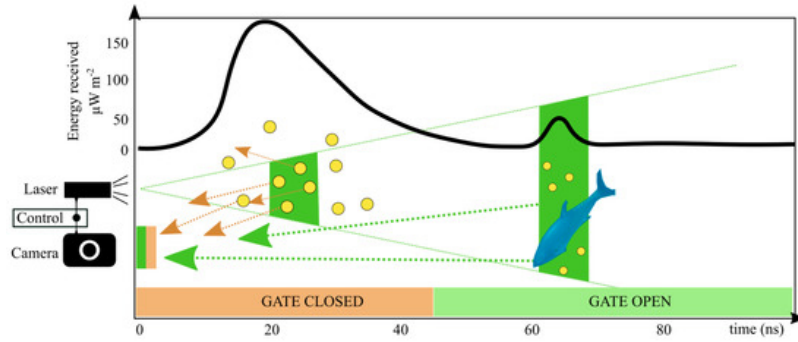


Figure 2.2: Energy values received by the sensor (camera) throughout time portraying the scattering problem and range-gate solution. Extracted from [17].

The absorption and scatter phenomena can be used as input for the Lambert-Beer empirical law, for further analysis [13, 10]. Assuming an isotropic and homogenous medium, the irradiance  $E$  at position  $r$  can be modelled as:

$$E(r) = E(0)e^{-ar}e^{-br} \quad (2.1)$$

where  $a$  and  $b$  are inherent properties of the medium and correspond to the absorption and scattering coefficients, respectively. Considering the image formation model of Jaffe-McGlamery [13], the total irradiance received by the optical sensor ( $E_T$ ) can be divided into three components: the direct component ( $E_d$ ) that relates to the light being directly reflected by the object, and the forward-scattered and backscattered components ( $E_f$  and  $E_b$  respectively). Therefore, the total

irradiance is calculated as follows:

$$E_T = E_d + E_f + E_b. \quad (2.2)$$

The aforementioned solutions for scattering aim to increase the robustness of the image acquisition process by minimizing the influence of  $E_f$  and  $E_b$ .

As it is possible to conclude, the conditions of the surrounding underwater environment have a direct impact on the issues related to light propagation in water. In order to study that influence, a proper evaluation of the water's state is important. There are two concepts which are commonly used to characterize the underwater surroundings when testing an underwater sensor [18]: turbidity and attenuation length. By measuring the degree of loss of transparency of water due to the presence of suspended particles, is possible to obtain the water's **turbidity** value. It's measured in Nephelometric Turbidity Units (NTU) using a device called nephelometer or turbidimeter. In lakes, the turbidity can also be measured using a Secchi disk <sup>1</sup>. The final metric, **attenuation length** of light, is defined as the distance where the intensity of the light beam has dropped to  $\frac{1}{e}$ , or about 63% of the light particles have been stopped.

Lastly, another aspect that influences the quality of the data acquired by the perception system is the type of watertight housing structure that is used to protect the electronics. The enclosure is responsible for two refractions of the incident light, that occur between the two interfaces (air-glass-water), following Snell's law [18, 19, 20]. There are two options for housing types often used, a flat-shaped or dome-shaped port. As it is shown in the work done by T. Treibitz (2011) *et al.* [21], when using a flat surface enclosure, the rays coming from different objects, refracted by the medium, appear as if they are imagined from different points of view since there is no alignment between the object point in water, the centre of projection of the camera and the image point, completely negating the assumption of the single viewpoint model [19, 21]. Figure 2.3 (left) demonstrates how the light rays are refracted when a flat surface enclosure is used to house the optical system. In order to minimize the non-collinearity effect caused by the medium's refractions, a dome-shaped enclosure should be used with an exact alignment between the camera's projection centre and the enclosure's centre point. With this configuration, there are, theoretically, no significant refractions, concluding that the interface normal and the incoming rays are closely aligned, and thus viable to assume a single viewpoint model [18, 20, 21], as it is demonstrated in figure 2.3 (right).

## 2.2 Underwater Perception Systems and 3D Reconstruction Techniques

In recent years, an increasing interest in autonomous agents have led to an advancement in perception systems. These include optical systems, which are a common solution when it is nec-

---

<sup>1</sup> Secchi disk is a black and white circular disk that is lowered inside the water to measure the transparency (Secchi depth), which is related to water turbidity level.

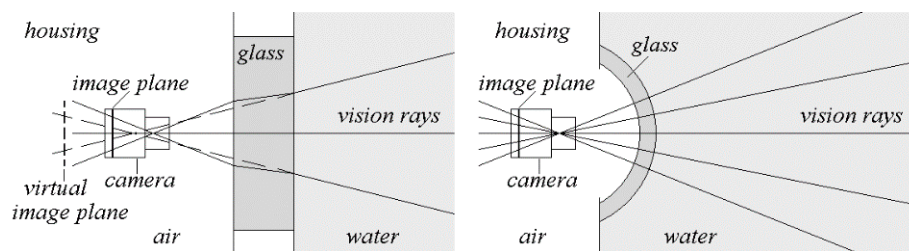


Figure 2.3: Difference of light rays' refraction on a flat enclosure (left) and on a dome-shaped enclosure (right). Extracted from [20].

essary to gather 3D information for reconstruction purposes. Depending on the type of optical sensor that is integrated, these systems can be separated into two categories: passive and active (further explained in sections 2.2.1 and 2.2.2 respectively). The combination of both types compose an hybrid approach, presented in section 2.2.3.

### 2.2.1 Passive Perception Systems

Passive perception systems utilize optical sensors that directly measure natural energy (radiation) that is emitted or reflected by the object or scene being observed. When the imaging is done at deeper waters, light sources (i.e., spotlights and LEDs) could be necessary for scene illumination.

A. Houge *et al.* (2007) [22] developed AQUASENSOR, a stereo vision-inertial sensing device that combines the 3D information obtained using stereo vision with a 3 DOF inertial sensor. Represented in figure 2.4, the device was designed to be sufficiently portable to be held by a single diver and to later be integrated within a visually guided autonomous robot. The authors also report that AQUASENSOR was used for wreck and coral reef reconstruction, with offline 3D data processing, and tested operation depths of up to 30 m while sealed within the custom underwater housing. Later on, in 2014, K. Williams *et al.* [23], presented the TrigCam stereo system, for the purpose of capturing biological activity in the underwater environment. The authors designed the optic system using off-the-shelf cameras and computer components for a low cost solution. Illumination is chosen in order to be insensitive and unobtrusive to the fish and, thus, the area is illuminated with a far red light emitting diode (LED) while no life is detected. The particularity of the authors' approach is a triggering method. The presence of fish switches the far red light to white illumination, in order to acquire a high-resolution full colour stereo image of the environment.

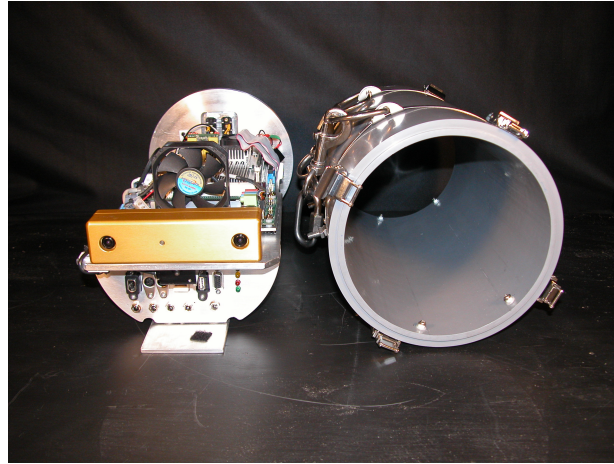


Figure 2.4: The compact and portable AQUASENSOR device with its housing to its right. Extracted from [22].

Under the MARIS project [24], F. Oleari *et al.* (2015) [25] developed a stereo rig composed of two industrial AVT Mako G125C GigE cameras that implements an underwater stereo vision system. This work system is designed to support object grasping and manipulation from a floating base, by implementing an underwater 3D object detection solution. A Standard Absolute Differences (SAD) correlation method was used to evaluate the vision system's performance to reconstruct a 1292x964 resolution disparity map, achieving a throughput of 12.5 frames per second. In research work authored by F. Bonin-Fort *et al.* (2015) [11], the Fugu-f advanced vision system, was designed as a flexible visual perception solution for an AUV, formed by two external Bumblebee stereo rigs, capable of performing at 250m of depth. The 3D reconstruction performance test of this system resulted in dense point clouds of the robot's environment of up to 800.000 points from one shot. After the merge of successive point clouds, it was stated that it resulted in high processing times and usage of memory resources, which were resolved with reducing the rate of the processed stereo pairs to 1 per second and down-sampling the data to only one point per voxel. The main advantage of this system was its modularity and adaptability, that enabled this system to integrate two different AUVs: the Girona 500 [26] and Nessie VI [27].

In more recent work, by M. Rossi *et al.* (2018) [28], a stereo RGB camera rig, composed by two standard industrial machine vision cameras enclosed in underwater housings, was mounted on the front of ROVs, as represented in figure 2.5. The cameras produced RGB images with a resolution of 960x600 pixels at 22 frames per second with the 3D models being built and updated online, in clear and murky water conditions. The authors adapted the KinectFusion algorithm, which uses the optical sensor Kinect to acquire 3D information, to the stereo camera setup, presenting and experimenting the underwater StereoFusion, an algorithm for real time 3D dense reconstruction and camera tracking.

One of the strong points of passive sensors is the ability to obtain highly dense point clouds that can be extracted using perception systems that integrate them [29, 30]. Besides that, some



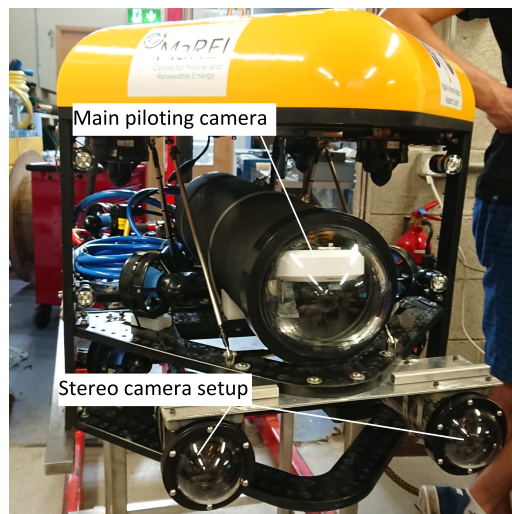


Figure 2.5: ROV with a stereo camera setup composed by two enclosed cameras and a main piloting camera. Extracted and adapted from [28].

other benefits include lower required energy consumption to operate such devices, lower cost and theoretically higher lateral resolution (capability of discerning objects that are lateral to each other) [18]. The main drawbacks consist of the increased computational power required and dependency of the target's texture in order to extract 3D features from the image [29, 30].

### 2.2.1.1 3D Reconstruction with Passive Perception Systems

3D reconstruction techniques using passive sensors rely on photogrammetry, the science of making measurements from photographs, since the 3D information is retrieved from overlapping images captured at different viewpoints [10, 29, 30]. Using these extracted images, point clouds generation techniques include:

- Stereo Matching Techniques;
- Structure-from-Motion (SfM).

#### Stereo Matching Techniques

These techniques refer to image acquiring systems that usually include two or more synchronized cameras offset by a fixed distance (baseline), resulting in a large field of view (FoV) overlap. The stereo 3D reconstruction is based on the triangulation method of corresponding points on the two views, in the case of two cameras, portrayed in figure 2.6. This requires to find, in both images, common 2D points and match them (correspondence problem), determining the plane that contains the 3D point on the object [29]. The work done by G. Huo *et al.* (2018) [31] consists on applying a semi-global stereo matching technique on a underwater 3D reconstruction system based on binocular vision. The authors propose an improved algorithm based on Semi-Global Matching (SGM), that is described in their article, that is able to provide a time-efficient way of accurately

calculating disparity maps. The disparity values are later used in the triangulation process for 3D reconstruction of objects. The authors report on an improved accuracy on 3D reconstruction by using their adapted algorithm, while working within 2 m distance from the target. Resulting disparity maps and 3D reconstruction are portrayed in figure 2.7.

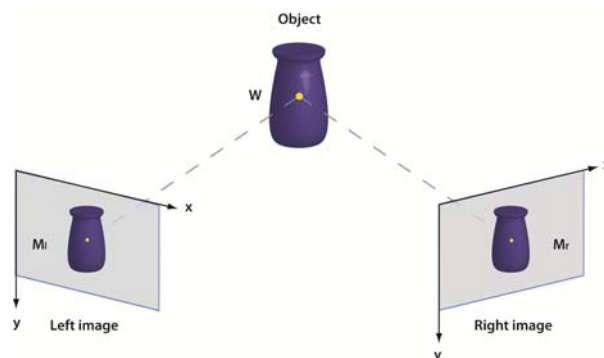


Figure 2.6: Stereo correspondence based on triangulation for 3D point computation. Extracted from [29].

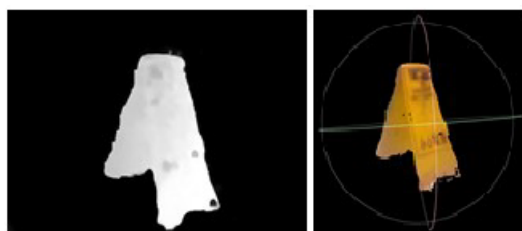


Figure 2.7: Disparity map result from the adapted SGM (left); 3D reconstruction result (right). Extracted and adapted from [31].

With recent developments on deep learning approaches, some researches integrate new developed algorithms in stereo vision-based systems, for underwater 3D reconstruction. In 2019, K. Skinner *et al.* [32] proposed an unsupervised learning method, named UW StereoNet, for underwater applications. An unsupervised deep neural network (DNN) takes input raw colour underwater stereo imagery and outputs dense, gray scaled disparity maps and colour corrected images of underwater environments that enable later triangulation.

### Structure-from-Motion

3D reconstruction can also be accomplished using a Structure-from-Motion (SfM) approach, where several images of an object taken from a monocular camera, in different points of view, provide information for detecting and matching image features, by overlapping them. The 3D coordinates of each individual feature can later be computed using triangulation.



Taking into consideration image degradation and refraction in the underwater setting, X. Qiao *et al.* [33] (2019) proposed a systematic SfM pipeline that ends with the 3D reconstruction of the scene by triangulation methods.

Although the SfM technique is quite simple to implement, it is greatly affected by inaccurate navigation data. Thus, some published work attempt to solve this problem by estimating the scale required for the 3D models to accurately represent the targeted scene or object. In 2020, K. Istenič *et al.* [34] improved 3D modelling based on structure-from-motion by using laser scalers for automatic scale estimation. The authors proposed two novel approaches which are suitable for different laser scaler configurations: a flexible method that enables the use of arbitrary laser setups and a partially constrained method that reduces the need for calibration. In the real scenario, the fully flexible and unconstrained approach was able to produce computed values of the model with a standard deviation of 0.3%. Finally, the proposed approaches were compared with other methods usable for the tested dataset and it was reported that the consistency of scales obtained using these approaches is much higher (0.6% compared to 4%).

### 2.2.2 Active Perception Systems

Optic systems that utilize coherent light (lasers) as a tool to get information from the environment correspond to active perception systems. These systems integrate optical sensors that mainly use structured-light, including laser beams, laser beam with a Diffractive Optical Element (DOE) and projectors which project a pattern or Light Stripe Ranging (LSR) which, by intersecting a projected plane onto the scene, create an illuminated line [10, 18, 29, 30]. Other technologies that can also operate as active perception systems are time-of-flight (ToF) based sensors, such as LiDAR, RADAR and SONAR.

Active techniques have been vastly explored in underwater applications, more particularly structured-light based optical sensors. An older work done by A. Yamashita *et al.* (2004) [35] follows the principle of projecting a laser line that is captured by a charge-coupled device (CCD) camera. The 3D measurement technique, using triangulation, takes into consideration the refraction of the light to achieve better accuracy results of about 1 mm for objects located at about 400 mm. Optical systems that have a single line laser would often result in very sparse point-cloud extraction. In order to minimize that effect, multiple lasers started being integrated in the perception systems in order to acquire richer 3D information from the scene. However, the usage of more lasers requires increased power consumption, which is a significant downside when applied to AUVs or any other robotic platform. Recent work done by N. Hansen *et al.* (2015) [36] developed a minimalistic two laser line sensors optical system, with one camera, for underwater inspection by AUVs or any other robotic platforms. 3D data acquisition is based on triangulation of points extracted from the image using the Hough Transformation. The optical system was evaluated on both simulated and real OpenROV, with no water turbidity variations, and presented approximately 0.04 m of drift in the measurements. F. Lopes *et al.* (2015) [37] also took on a similar approach, with an optical system that includes two line lasers (green and red coloured, with different beam wavelength for experimentation purposes) and a SXGA Basler camera between them. The devices

were coupled together in a watertight housing, represented in figure 2.8. The authors also implemented two different structured light system (SLS) calibration techniques. In the experimentation phase, the optical system was used to gather 3D data from a submerged scallop, which resulted in 2% error on the object's size using the structured light system line projection (SLS-LP) calibration method.



Figure 2.8: Optical system composed of two laser beam sensors, laterally placed to a SXGA Basler camera, inside watertight enclosures. Extracted from [38].

Another approach often used is projecting patterns onto the scene. However, in order to achieve a balance between power autonomy and point cloud density, this approach can also be implemented using lasers with a DOE for less energy demanding operations. Both of these methods are particularly interesting since the projected patterns accentuate surface textures. In 2010, Q. Zhang *et al.* [38] developed a 3D shape measurement technique that uses a CCD camera and a digital projector that illuminates the target with a 2D sinusoidal fringe. Firstly, a phase tracking algorithm is used to identify homologous points in phase distributions of the deformed fringe capture and in the fringe pattern being projected. Secondly, a ray tracing algorithm is responsible for tracing the propagation path of each ray and to calculate the 3D coordinates of each point on the target's surface, so that the shape of the tested object can be later reconstructed. The authors also state that the main advantages of this approach lay on the simplistic configuration, compared to scanning devices, that is able to gather data from the whole measured surface simultaneously and with great speed, with the cost of resulting in lower accuracy results. The authors F. Bruno *et al.* [39] (2011) have designed and experimented on a structured light and stereo vision setup for underwater 3D reconstruction. This system projects a binary pattern onto the target, used to establish correspondences between projector and projections, and later recovers 3D information by means of triangulation, using the configuration of camera positioning. Their reported accuracy is of 0.2 mm at a working distance of 1200 mm. M. Massot-Campos *et al.* (2014) [40, 41] used the single laser combined with DOE approach. The authors presented a one-shot sensor for un-

derwater 3D reconstruction that is composed of a RGB CCD camera and a green laser of 532 nm wavelength, with a DOE attached to it in order to project a pattern of parallel lines into the scene. The camera is used to capture the projection of the laser lines on the target, which is then segmented and later used for the process of solving the correspondence problem, where the projected stripes are matched with the captured stripes. Finally, each 3D point is computed using triangulation (result portrayed in figure 2.9). Their work also states that the DOE refracts the laser beam in 25 parallel lines to provide sufficient data from a single camera frame for 3D reconstruction. This optical approach was tested in variant water turbidity levels and it was reported that with low to medium turbidity, a high number of 3D points were successfully extracted to a non-dense point cloud. However, with high turbidity, most laser lines were completely absorbed in the water and only the central beam, which had the highest energy, is captured.

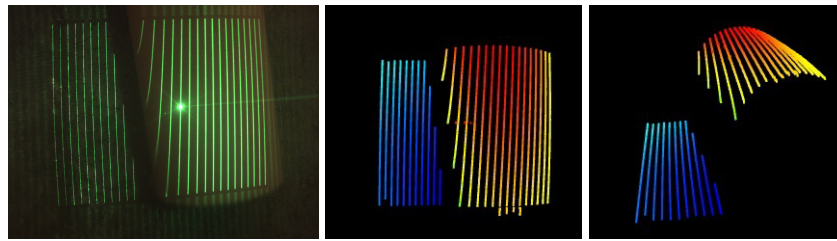


Figure 2.9: Input captured image of a pipe with projected parallel lines (left); top view (middle) and lateral view (right) of resulted 3D reconstruction of the pipe. Extracted from [40].

In order to obtain more faithful 3D representation of reconstructed objects or scenarios, translation and rotation motions were required to enable lasers lines to swipe through an area and ultimately increase point-cloud densities. S. Tetlow and J. Spours [42] (1999) report on a laser stripe scanning system with an automatic threshold setup for the camera, increasing the robustness against pixel saturation if the laser reflection is too intense. The author's also present a flow diagram for the whole reconstruction process, which includes the triangulation method used to compute the distance to the target for each scanner position. The evaluation of the system resulted in resolutions of up to 5 mm at a 3 m distance from the target. When the vehicles, coupled with this device, lock onto the work site, it is reported that the reconstruction is done with 1 cm of accuracy in turbid water. A laser line scanner was developed by S. Jiang *et al.* (2017) [43] to perform underwater 3D reconstruction of objects. The system is composed of a fixed colour camera with a fish-eye lens and a green line laser that moved at a constant speed on an installed rail on top of a water tank. Later, the authors compare their 3D reconstruction method to an experiment using SfM and concluded that, although the latter achieved better edge reconstruction accuracy of the object, their proposed technique was more detailed. Q. Xue *et al.* (2021) [44] recently developed a high-precision rotating line laser 3D scanner optical system, composed of a low illumination underwater camera and green line laser, in separate watertight enclosures, mounted on top of a rotating structure, portrayed in figure 2.10. Able to do underwater 3D data acquisition in the range of field of view of 50° vertical and 360°, this system's 3D reconstruction error results are reported

to be less than 0.6 mm with a working distance of 140 mm to 2500 mm.

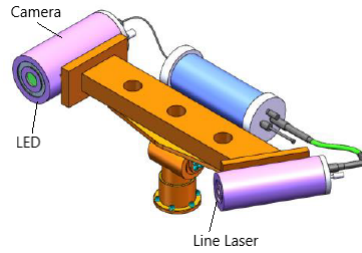


Figure 2.10: Model of an optical scanner system composed of a camera and line laser, mounted on top of a rotating structure. Extracted and adapted from [44].

Although rotating and translating structures with the laser sensors attached are capable of creating a scan motion, some researches took on another approach, that is fit to be deployed on lower power consumption platforms. In 2016, S. Chi *et al.* [45] developed a laser line auto-scanning system that was designed to perform underwater close-range 3D reconstructions with high accuracy and resolution. The auto-scan is performed with a laser-galvanometer motion module and is captured by a CCD industrial-grade low-illumination analog camera, with a resolution of 752x582. The galvanometer can rotate a mirror in the range of  $\pm 20^\circ$  with voltage in the range of  $\pm 10V$  controlled from a D/A conversion card in the computer. The authors report good accuracy in 3D reconstructing smaller objects, under ideal laboratory conditions, and acknowledge that the main source of error is the calibration factor, additionally stating that, in real underwater conditions, more errors would be introduced that weren't considered in these trials. A. Palomer *et al.* (2018) [46] follow a similar approach, developing a 3D laser scanner (figure 2.11) stored in a watertight housing with two viewports: one for a unit formed by a laser line projector and a galvanometer steering a mirror and another for the camera. The galvanometer position and camera are electronically synchronized in a way that in every capture, one single line is projected. Additionally, the authors also reported that this system can operate from 0.5 m to 5 m depending on turbidity levels and surface light absorption, at a frame rate of 0.09 Hz at full resolution (equivalent to  $0.008^\circ$  of galvanometer rotation between captures) up to 6 Hz at lower resolution and is able to measure with errors smaller than 1 mm and  $0.1^\circ$ .

Overall, this type of systems require complex optic setups, has higher acquisition times when not using a scan motion, has lower performance under bright sunlight and usually has higher energy consumption but is an efficient way of reducing the scattering component that appear in the FoV of cameras or other image acquisition devices, since the beam of light is more focused, and is not so dependent on the target's texture.

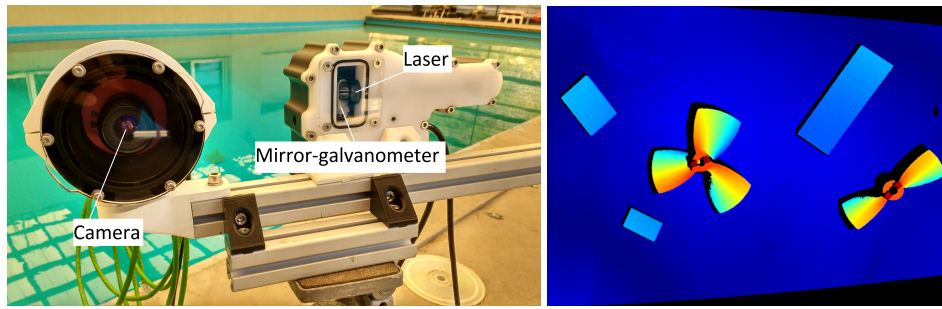


Figure 2.11: 3D laser scanner mounted on a tripod (left); resulting point cloud coloured by depth (right). Extracted from [46].

### 2.2.2.1 3D Reconstruction with Active Perception Systems

When using active imaging, 3D information can be extracted using methodologies such as:

- Time-of-flight (ToF);
- Pattern Projection.

#### Time-of-flight

Although less used in underwater settings, since ToF sensors often use IR waves which are prone to higher attenuation while propagating in the water (explained in section 2.1), these methods have been quite studied throughout the years. Time-of-flight techniques are based on controlling the travel time of a signal, and thus operate by transmitting a light pulse and detecting the light reflected from the objects within the receiver's FoV, as it is portrayed in figure 2.12.

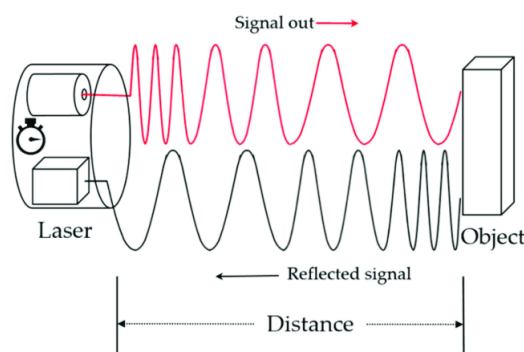


Figure 2.12: ToF sensors principle of operation. Extracted from [47].

Range information,  $d$ , is computed by measuring the time difference between the transmitted and received light signal,  $\Delta t$ , considering the speed of light in the medium,  $c_m$ , that is calculated as the quotient between the speed of light in vacuum ( $c$ ) and the refraction index of the medium ( $\eta_m$ ), according to the following formulas [18]:

$$d = \frac{1}{2} \Delta t c_m \quad (2.3)$$

$$c_m = \frac{c}{\eta_m} \quad (2.4)$$

These methods achieve somewhat long distances, although changes in the water's properties must be accounted for. For short distances, small inaccuracies could present big changes in the time measurements, which could lead to high resulting errors [19]. Ultimately, 3D information could be interpreted as depth measurements from the sensors that use this methodology. SONAR, RADAR and LiDAR are some of the sensors that utilize this principle of measurement.

### Pattern Projection

These methods refer to optical devices that use structured-light as a means for extraction of 3D information, such as laser lines, with and without DOE, and projectors. The extraction of highly precise 3D points is executed by exploiting a known relative pose between camera and laser or projection emitter, by means of triangulation, as represented in figure 2.13.

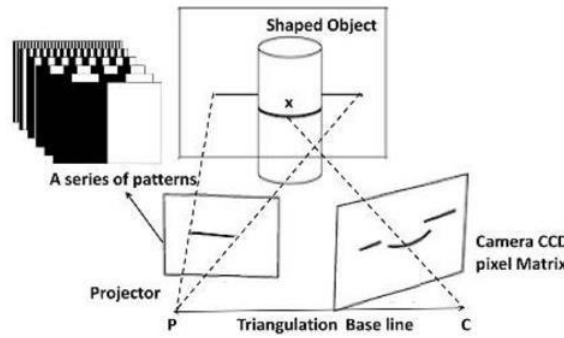


Figure 2.13: Triangulation principle on pattern projection methods. Extracted from [48].

### 2.2.3 Hybrid Perception Systems

An innovative research done by A. Pinto and A. Matos (2020) [10] presents an advanced underwater imaging system for underwater robotic application with a hybrid approach for 3D data acquisition, named MARESyE. This proposed methodology differentiates itself from previous mentioned work and other conventional visual sensors by taking into account the photometric limitations that affect optic systems operating in underwater environments. The imaging system, represented in figure 2.14 is composed of three modules: a passive module with a photometric stereo (PS) configuration (two cameras and a high intensity LED illumination unit), an active module with a light stripe range (LSR) configuration (two laser stripes of red and green colour) and a range-gated synchronizer, that allows to decrease the signal-to-noise ratio (SNR), by mitigating the effects of backscattering (refer to section 2.1).



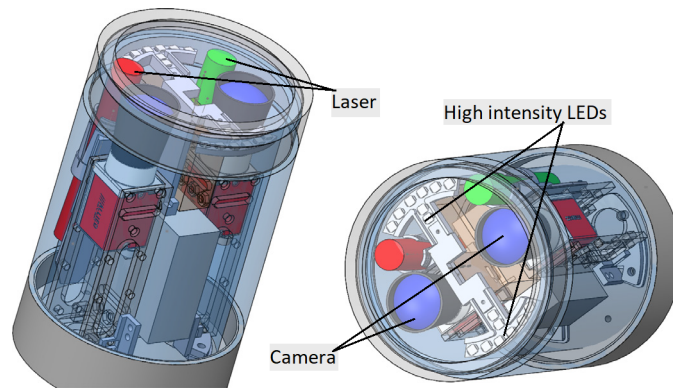


Figure 2.14: MARESys concept model represented by a watertight enclosure containing the PS module, composed of 2 cameras and LEDs, and the LSR module, composed of a red and green coloured laser sensor. Extracted and adapted from [10].

The hybrid approach exploits the combination of both imaging modules through a data-driven formulation to increase the measurement range and produce high density 3D estimations, making it a very flexible solution for AUVs or ROVs operating under dynamic or non-dynamic underwater environments that have varying photometric conditions. The authors stated that the hybrid system was evaluated under realistic sub-sea conditions, from a low baseline level turbidity, with resulting 3D reconstructions presented in figure 2.15, to high turbidity, and reported perceptual qualities with structural similarity index measurements (SSIM) above 0.655, while a conventional system obtained a value of 0.328, with highly accurate and precise reconstruction, since the Euclidean error was lower than 0.025 m and 0.010 m for the PS and LSR modules, respectively.

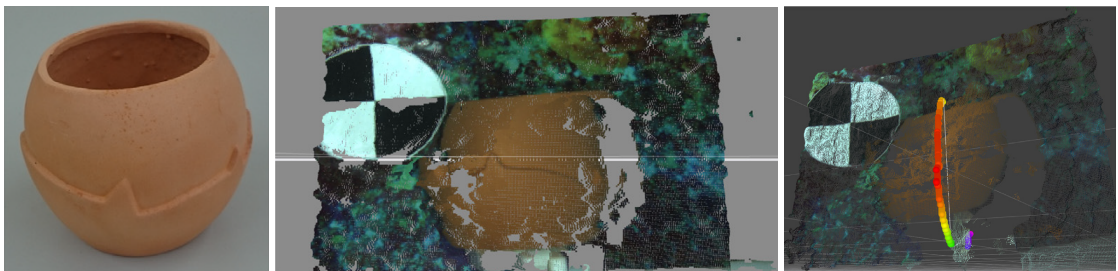


Figure 2.15: One of the objects used for laboratory testing (left); resulting 3D reconstruction using PS module, in low turbidity conditions (middle) and resulting 3D reconstruction using LSR module, in low turbidity conditions. Extracted from [10].

## 2.3 Eye-in-Hand Approaches

As it was previously seen, perception systems have a plethora of applications in sub-sea environments. In particular, when it comes to O&M tasks in an offshore setting, vision perception is an important tool for successful operations of inspection and intervention. Thus, it's inevitable that

the perception system must be integrated within the vehicle itself, either AUVs or ROVs. However, most underwater vehicles with a manipulator attached often present a poor configuration that leads to occlusions produced by the robotic arm. This study intends to present an Eye-in-Hand approach, where the perception system is attached to the end-effector of the manipulator instead of the vehicle's body, in order to avoid these occlusions and to take full advantage of the extra DOF in which the manipulator is able to operate, so that the image acquisition can be performed closer to the target and to obtain the maximum level of detail as possible. This configuration has been vastly used in robotic arms, although it is still quite scarce when it comes to underwater intervention and inspection scenarios.

The recent work done by S. Aldhaferi *et al.* (2022) [49] provides a good state-of-art on projects of underwater vehicles where a robotic manipulator and visual perception are required. The following section explores the Eye-in-Hand topic, where several research work that benefits from this configuration is presented.

In 2008, G. Marani *et al.* [50] developed SAUVIM (Semi Autonomous Underwater Vehicle for Intervention Mission), an AUV equipped with a 7 DOF robotic manipulator (MARIS 7080), which has a fully autonomous mode with the aid of a camera unit coupled to the robotic arm and other perception sensors. In a search and manipulate experiment, a remote user is able to command the AUV to search a hook when it reaches the target site, prompting the robotic arm to scan the immediate area. Once the target is detected, the robot attempts to hook the cable to a known point of the object, in order to bring it to the surface. The detection operation is achieved through optical feedback from a video camera located on the arm's wrist, as it is portrayed in figure 2.16. The target is detected based on its geometry, through the following pipeline: image filtering, edge extraction using Canny filter to the colour image and using the colour contrast gradient and finally circle extraction, using the line segments found in digital images. Once the target has been detected, the arm enters a tracking state (visual servoing) and proceeds to grasp the object, while it is not moving. In terms of accuracy detection, it was reported as under 1 cm for the translational part, sufficient for the target to be grasped, with a slightly higher error in the orthogonal direction of the camera view.

In 2010, P. Sanz *et al.* [51] presented a reconfigurable autonomous underwater vehicle for survey and intervention purposes, developed during the RAUVI project. The authors present their I-AUV concept model, Girona 500, for a maximum operating depth of up to 500 m, and the CSIP Arm5E manipulator, with 4 DOF, represented in figure 2.17. The authors take on the Eye-in-Hand approach, mounting a high-resolution colour CCD camera, valid up to 100 m underwater, on the last link of the robot arm, for 2D visual perception.

After the successful first experiments with RAUVI [51], M. Prats *et al.* (2012) integrated some aspects of their previous work into the TRIDENT project [52]. The authors report on the integration of the Girona 500 AUV with the Eye-in-Hand configuration composed of the Arm5E light-weight manipulator and a stereo system in a watertight enclosing, formed by two high resolution cameras, for 3D information acquisition of the scene in order to perform survey and intervention operations. For the object recovery experiment, the visual perception system is pivotal in the sur-



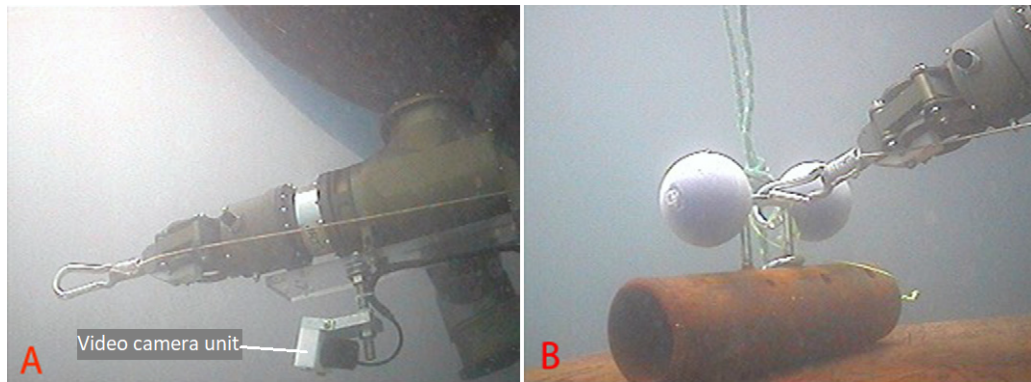


Figure 2.16: Recovery operation: SAUVIM AUV, with robotic arm and sensory unit attached, navigating towards the target (left); manipulator detected the object and begins to grasp it (right). Extracted and adapted from [50].

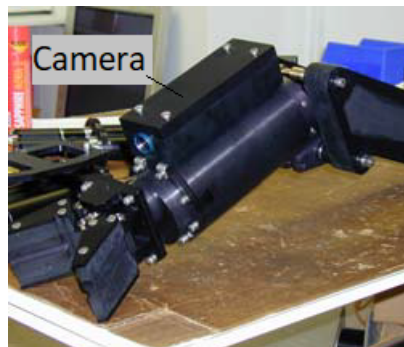


Figure 2.17: Configuration of the arm, gripper and camera, during development phase. Extracted and adapted from [51].

vey phase, in which a orthophotomosaic of the seafloor is setup with the imagery gathered. When the target object is detected in the mosaic, the intervention phase begins, with feedback aid of the optical system for navigation purposes, where the vehicle is guided towards the neighbourhood of the target object in order to recover it.

In 2012, M. Prats *et al.* [53] developed a different setup for 3D reconstruction and grasping in underwater environments, using the robotic manipulator Arm 5E with an active perception system, where the structured laser beam is attached to the forearm of the manipulator and a digital video camera is integrated in the AUV, looking downwards in order to capture laser projections, in a configuration represented in figure 2.18. The authors incorporated a structured-light based sensor and a camera to perform triangulation in order to recover the 3D structure of unknown objects lying on the seafloor. The projected images are segmented, by applying a threshold on the calculated euclidean distances (in RGB space) of each pixel with a reference, in order to detect projection pixels and then, by means of triangulation, extract a point cloud. The whole process is represented in figure 2.19, where after the scanning process, the acquired images of the laser projections are used for 3D reconstruction of the scenario. The generated point cloud is then used

to aid in grasp planning of a target object that is executed autonomously by the robot after a simple user command.

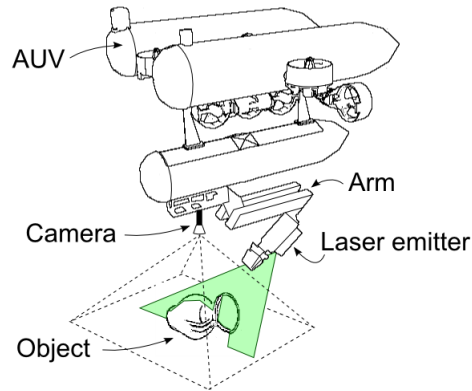


Figure 2.18: Configuration of the arm, laser emitter and camera for object scanning. Extracted from [53].

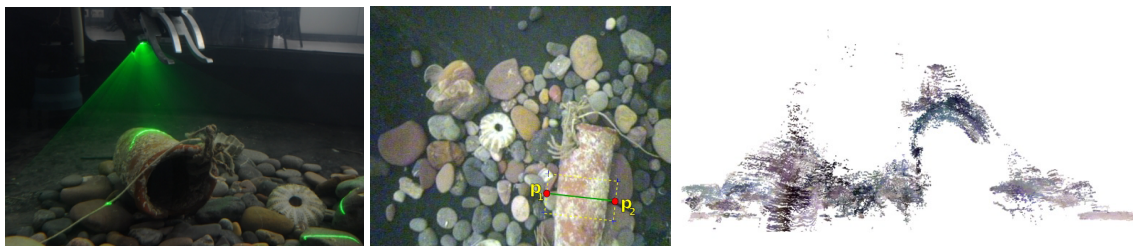


Figure 2.19: Laser scanning of an amphora (left); projected image in camera view (middle); side view of the generated raw point cloud of the amphora scan (right). Extracted from [53].

Under the project titled TRITON [54], in 2013, P. Sanz *et al.* developed GRASPER, a multi-sensory based manipulation system for underwater operations, that applies the approach developed by M. Prats *et al.* [53], in 2012, but in a different scenario, where the Arm 5E robotic manipulator is attached to a floating platform for intervention purposes, particularly tested for situations such as the search and recovery of an object of interest and the intervention of an underwater panel in a permanent observatory.

Later, in 2015, A. Peñalver *et al.* [55] changed the approach from previous research work by attaching both laser stripe emitter and camera to the forearm of the Arm 5E robotic arm, as it is portrayed in figure 2.20. With this Eye-in-Hand configuration, the authors stated that with the camera coupled to the manipulator, it was possible to reconstruct the scene from different views, since the camera was no longer in a fixed position, and from a position that was closer to the target object, obtaining more precision in the reconstruction. This study also includes a comparison between the one-view method, where the camera is placed in a fixed position, and the proposed multi-view, with both resulting reconstructions represented in figure 2.21, clearly showing a more complete and defined reconstruction with the latter.

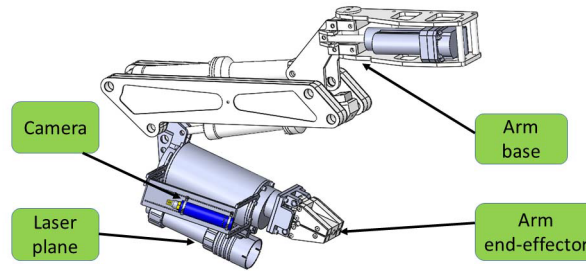


Figure 2.20: Eye-in-Hand configuration, with the camera and laser stripe emitter attached to the Arm 5E manipulator. Extracted from [55].



Figure 2.21: Skull as the simulated target object for reconstruction (left); result of the reconstruction using one-view approach (middle); result of the reconstruction using proposed multi-view approach (right). Extracted from [55].

A similar Eye-in-Hand configuration is applied to recently developed AUVs, in a project named TWINBOT (2021) [56], for autonomous underwater cooperative transportation. R. Pi *et al.* presented a multi-agent system solution, with a group of cooperating I-AUVs, composed by two Girona 500 lightweight modular I-AUVs, where each vehicle had a ECA 5E manipulator attached. Additionally, each vehicle also had three cameras, an analog camera attached to the body of the I-AUV, forward-looking for localization purposes, another analog camera embedded in the custom end effector to provide visual feedback for grasping purposes, portrayed in figure 2.22, and finally a high-performance megapixel industrial camera, down-looking, attached to the lower side of the I-AUV, to get seafloor image updates for navigation purposes. The authors' article also include images of a successful pipe pick-up and transportation process.

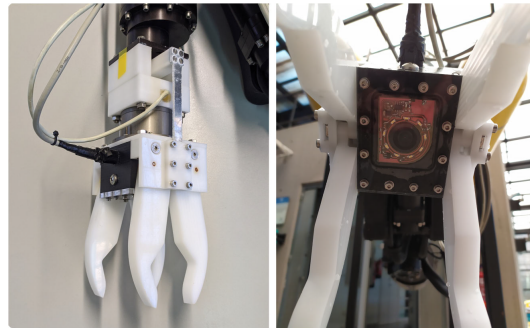


Figure 2.22: Custom end-effector with grasping abilities (left) with embedded camera (right). Extracted from [56].

## 2.4 Critical Analysis

Perception systems are a pivotal component when it comes to acquiring 3D data from the surrounding scene, specially for inspection and intervention tasks, often required in offshore O&M scenarios. These systems enabled surface and underwater vehicles to achieve higher levels of autonomy and thus, to remove the human factor from the equation, which is often important, not only to protect lives but also to prevent human prone errors in these operations.

Nevertheless, it is crucial to take into account and develop strategies to mitigate the negative effects that water, as a medium of propagation, causes on optic systems, including light absorption and scattering. For light absorption, post-processing image methods are often used for colour correction. Scattering effects on image acquisition can also be minimized with post-processing techniques to increase image contrast or with a range-gated approach to filter most of the captured backscattered component. It is also important to evaluate which type of watertight housing for the perception system works best, since dome-shaped ports, unlike flat-shaped, minimize the distortion of scenarios, since, in ideal conditions, the refracted light and the interface normal are closely aligned.

Active and passive optical sensors are the core of perception systems and thus, several of these have been developed with the variety of sensors that fall under these two categories. When using natural or artificial lighting to illuminate the scene, stereo-based configuration of two cameras can achieve a very dense point cloud of the target, though with low accuracy, since these sensors are highly dependent on the level of texture of the captured object to later extract 3D information. Another technique often used to extract 3D information from just cameras is structure-from-motion, using a single camera to capture the target from different viewpoints. The obvious advantage is the need of only one camera, making this a more affordable solution, although it is required to move the camera to another positions in order to extract any 3D information and is highly dependent on navigation data. Systems that incorporate active sensors mainly use structured light projected onto the scene, such as LSR. From these optical systems, it is possible to extract very accurate but sparse 3D point clouds. By applying translation or rotational movement with these techniques, it is possible to fully scan the target and thus obtain more complete and detailed 3D

reconstructions. Because structured light is a great cheap alternative and works best in static environments, it makes it a great candidate for this dissertation's scenario, where a 3D reconstruction of an underwater object is the end goal.

As it was previously mentioned, many of these technologies were adapted to underwater environments in order to do tasks that involve inspections and interventions. In order to do so, researchers have attached these optic elements and systems to robotic manipulators. Although with scarce application to this day, particularly in an underwater context, this method, named Eye-in-Hand, as proven to be highly efficient by providing a way to avoid possible occlusions that the robotic arm could cause and to fully explore the extra degrees-of-freedom that the arm enables to achieve. In the context of this dissertation, this configuration would be an appropriate solution to get the perception system closer to the target object and achieve interesting points of view in order to get a richer, more complete and detailed 3D reconstruction of the submerged object, that could be used either for information on the object's pose, characteristics or as an important grasp planning aid.



## Chapter 3

# A 3D Reconstruction System for Submerged Objects

### 3.1 Introduction

Recent technology advances allowed to exploit the maritime environment in several ways. One of the most prominent industries that operate in offshore environments is the energy production, particularly those that pursue green energy sources, such as the wind farm industry. The large infrastructures installed at sea for that purpose require constant need for maintenance, which is why investment in technologies to support inspection and intervention tasks in subsea environments is at an all time high. For such operations, this dissertation reports on the development of a system for 3D reconstruction of submerged objects utilizing the Eye-in-Hand configuration.

The following chapter will present a description of the developed system for 3D reconstruction using the active EiH approach. The proposed architecture, fully depicted in figure 3.1, includes an object detection node that acquires the 3D position of the target object, a pivotal parameter to be used to compute several viewpoints of the target, a path planning and execution node, that computes reachable viewpoints for the manipulator and executes them, and a 3D data acquisition and processing module, composed of two nodes, one for each perception unit (LSR and PS), that captures the 3D information, filters and combines it, in a point cloud form. The developed system also includes an evaluation module that takes these two point clouds as input and utilizes several metrics to determine the quality of the 3D reconstruction of the object.

#### 3.1.1 Simulation Setup

The simulation environment used to test the proposed 3D reconstruction method was constructed using the 3D simulator Gazebo. The scenario, fully represented in figure 3.2, is composed by a cage with dimensions 2x2x1 m. Besides working as a support for the manipulator, it also limits the space in which the target object to be reconstructed should be located. Additionally, it is also composed by the manipulator, which is fixed at the top of the cage, along with the coupled perception system, completing the Eye-in-Hand configuration. Although the maritime conditions

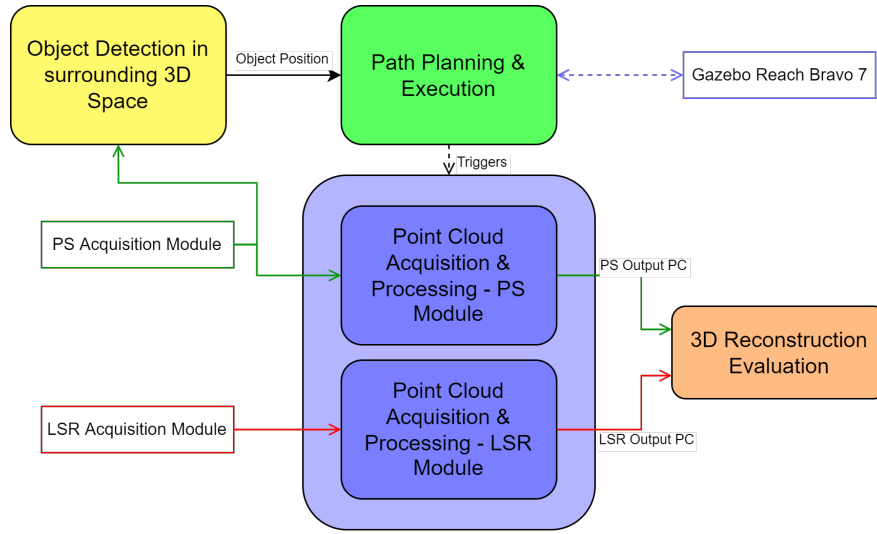


Figure 3.1: Proposed pipeline for the complete 3D reconstruction process.

are not being simulated in this Gazebo setup, the underwater challenges mentioned in 2.1 would need to be taken into account when testing in real underwater scenarios.

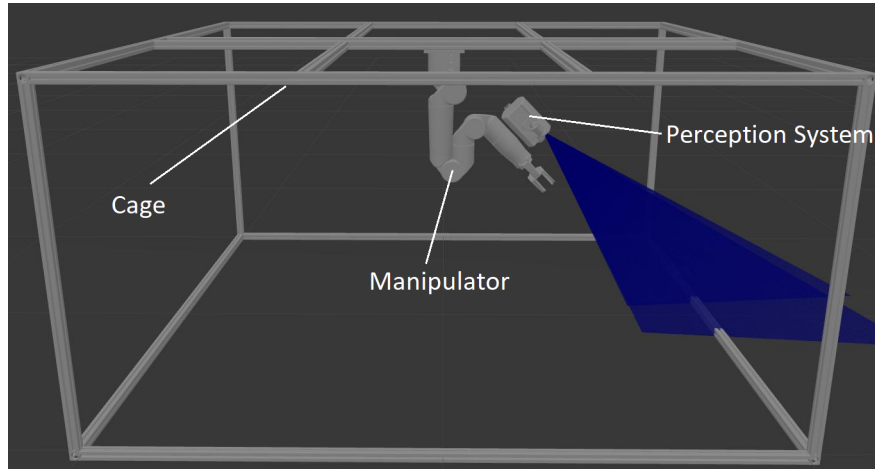


Figure 3.2: Gazebo setup used for simulation.

### 3.1.2 Manipulator Reach Bravo 7

Part of the EiH configuration is the manipulator, essential for the extra DOF that allows the perception system to acquire data in different viewpoints and closer to the target. The manipulator is based on the commercially available Reach Bravo 7<sup>1</sup>, represented in figure 3.3. The Reach Bravo 7 is a fully electric 7-function manipulator, with six revolute joints and an end effector for grasping purposes, specially designed to support the harsh underwater conditions and to perform

<sup>1</sup><https://blueprintlab.com/products/manipulators/reach-bravo/>



advanced inspection and intervention operations. The main properties of this manipulator can be found in table 3.1.

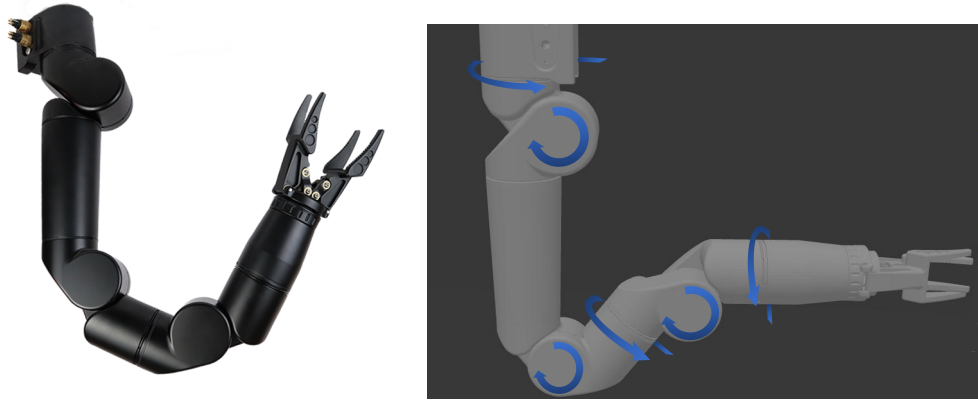


Figure 3.3: Reach Bravo 7, extracted and adapted from <sup>1</sup> (left), and simulated model in Gazebo (right), with arrows representing corresponding revolute joints rotations.

Table 3.1: Reach Bravo 7 properties.

| Properties     | Description                         |
|----------------|-------------------------------------|
| DOF            | 6                                   |
| Joint Speed    | 60deg/s nominal                     |
| Control Modes  | Position, Velocity, Cartesian (XYZ) |
| Depth          | 300 m                               |
| Weight (Air)   | 9.5kg                               |
| Weight (Water) | 4.5kg                               |
| Max. Reach     | 0.9 m                               |

To control the movement of the manipulator, the motion planning framework *MoveIt!* [57] was used. This motion planning framework uses plug-ins for most of its functionalities, such as motion planning plug-ins, namely the Open Motion Planning Library - OMPL, collision detection and, finally, kinematics plug-ins for forward and inverse kinematics. In the scope of this dissertation, *MoveIt!* was an essential tool to plan the paths of the manipulator, in order to get the perception system to relevant viewpoints to perform the 3D reconstructions and to interact with the Gazebo simulator.

### 3.1.3 MARESyE Perception System

For data acquisition, an updated version of the hybrid MARESyE [10] perception system, developed by A. Pinto *et al.*, which was already covered in section 2.2.3, was utilized. Its simulated model is represented in figure 3.4.

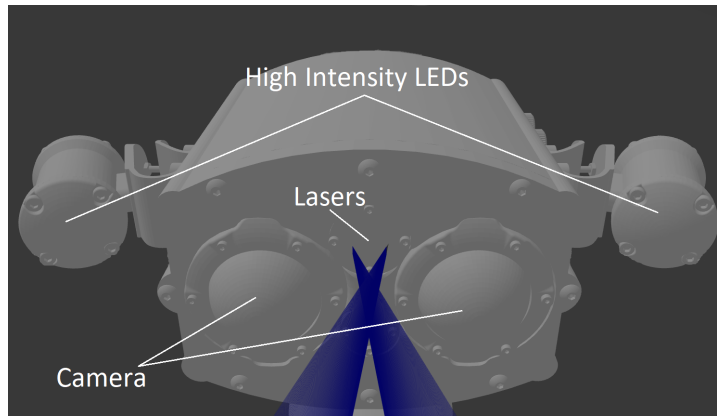


Figure 3.4: Simulated model of the MARESyE hybrid imaging system in Gazebo.

The active component of the perception system corresponds to the LSR module of the updated MARESyE, which is composed by two lasers, of red and green colour, that project lasers stripes onto the scene, and was the main source of data acquisition. For simulation purposes, two sonars were set up to mimic the real laser configuration, with parameters depicted in table 3.2. For noise simulation purposes, the built in Gaussian model in Gazebo was used in the LSR simulated sensors, which allows to control the standard deviation of the captured data. Additionally, the simulated version of the active perception module bypasses the whole process of the images' projection segmentation and triangulation of 3D information, that would need to be performed in a real scenario.

Table 3.2: LSR simulation default parameters.

| Properties | Description |
|------------|-------------|
| Rate (Hz)  | 40          |
| Samples    | 256         |
| FoV (rad)  | 0.58        |

The PS module, composed by two cameras in a stereo configuration and high intensity LEDs, was also used to acquire 3D information, mostly for comparison purposes with the other module. In order to simulate the operation of this module, a velodyne 3D LiDAR sensor was used, with default parameters represented in table 3.3. Similarly to the LSR simulated sensors, Gazebo's Gaussian model was also used for noise simulation, which allows to introduce standard deviation to data captured by the PS module. By using a simulated version of the passive perception module, the process of extracting 3D information from the camera's captured images was bypassed.

A representation of the operation of the simulated sensors for each perception module is depicted in figure 3.5. The figure illustrates the projections of the simulated sensors of LSR and PS modules, with red and green projection stripes and white dots, respectively, when no object is present and when a cube shaped object is captured.

Table 3.3: PS simulation default parameters.

| Properties       | Description |
|------------------|-------------|
| Rate (Hz)        | 10          |
| Samples          | 1024        |
| N° lasers        | 128         |
| Max. Range (m)   | 4           |
| Min. Range (m)   | 0.005       |
| Max. Angle (rad) | 0.610865    |
| Min. Angle (rad) | -0.610865   |

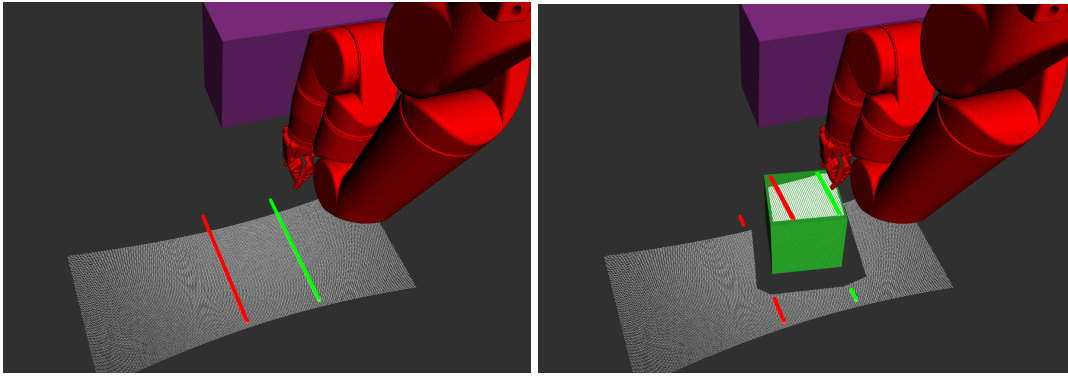


Figure 3.5: Operation of the perception system: modules' projections on the scenario's floor when no object is being captured (left); modules' projection on a cube object (right). White points correspond to data being captured by the PS module while red and green points correspond to data being captured by the LSR module.

## 3.2 3D Reconstruction Methodology

The proposed architecture for 3D reconstruction, fully described throughout the following section, can be summarized as follows:

1. Object detection - obtains the object's position by computing the 3D centroid of a point cloud captured from the target located in any position and orientation inside the cage;
2. Path planning and execution - using the obtained 3D position of the target, computes and executes the manipulator's trajectory, composed by several viewpoints where a scan sweep motion is performed;
3. 3D data acquisition - captures sparse data from the active perception module and dense data from the passive perception modules, filters and combines it iteratively, in order to output two point clouds, one from the sparse data and another from the dense data.

The object detection node utilizes the PS module of the perception system to detect the object in the surrounding space, since it is a great way to easily obtain a point cloud from just a one-shot capture that can later be used to detect the object's position. As it was previously stated in section 2.2.1, passive systems allow to obtain highly dense point clouds, thus making this module an appropriate means to detect and compute the 3D position of the target object. The developed algorithm for object detection requires that, besides the manipulator and perception system, the target object is the only other element present inside the cage. An overview of the developed simple detection algorithm is depicted in figure 3.6.

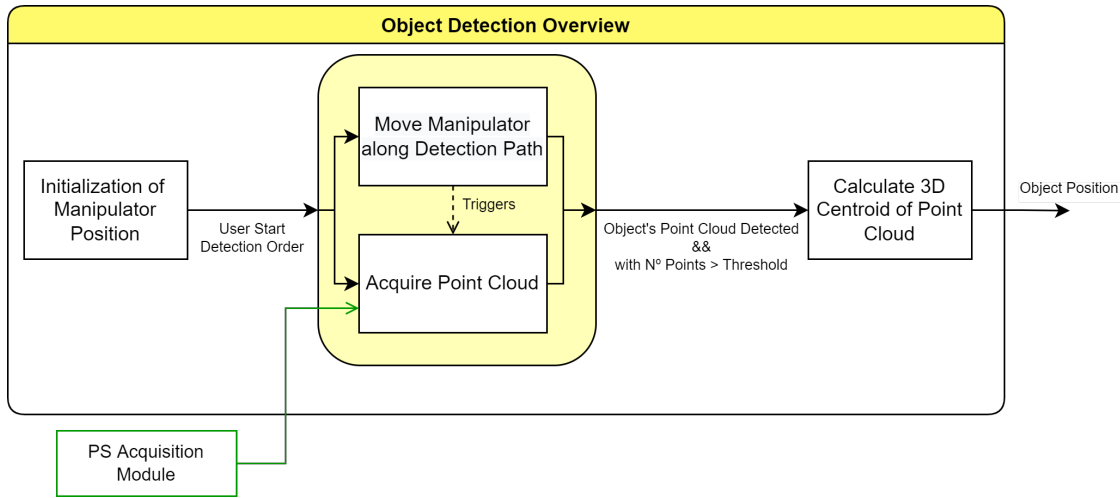


Figure 3.6: Diagram of the object detection algorithm.

Firstly, the object detection node initializes the system by inserting the cage and the perception system as constraints for obstacle avoidance and to move the manipulator to its initial position, as represented in figure 3.7. The green box at the base of the manipulator represents the top of the cage, while the purple box represents the perception system, attached to the manipulator. These elements' functionality is to act as an obstacle to be avoided by the *MoveIt!* planner when computing the manipulator's trajectories.

From its initial position, the manipulator's movement control, throughout the detection process, is operated in a joint mode with a fixed set of values of the manipulator's joints, since this path is the same for every scenario. The detection path is divided in two parts, as represented in figure 3.8, where the first part is meant to detect the object closer to the bottom of the cage and the second part detects objects higher up on the cage.

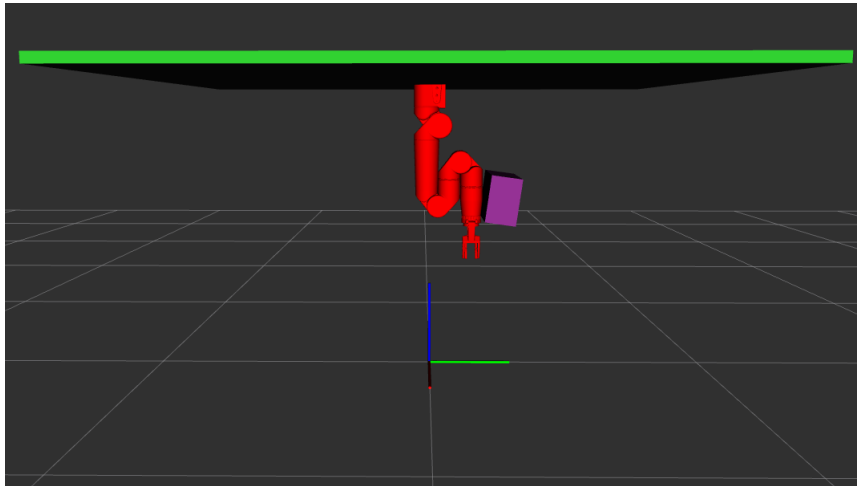
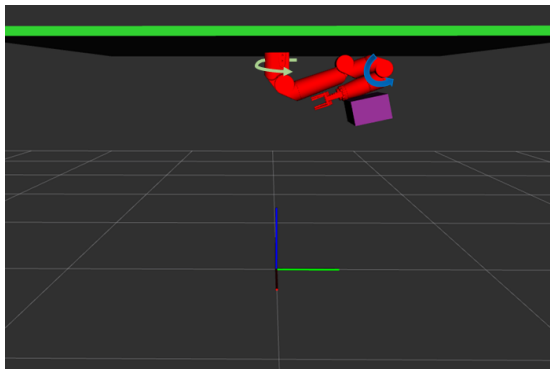
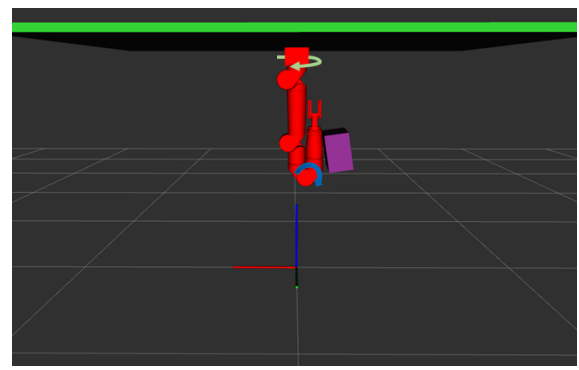


Figure 3.7: Manipulator in initial position with cage's reachable surface and perception system introduced as collision elements in the surrounding planning scene.



(a) Perception system's movement, depicted as a blue arrow, in the detection path's first part.



(b) Perception system's movement, depicted as a blue arrow, in the detection path's second part.

Figure 3.8: Manipulator's movement throughout detection path, repeated at every  $45^\circ$  increment, until reaching a complete  $270^\circ$  rotation of the manipulators' base footprint joint.

Throughout the execution of the path, the point cloud obtained from the PS module is box filtered, so that only points inside the cage space remain. An object is considered detected if a certain threshold of number of points in the point cloud, determined by experimentations that require only one execution of the manipulator's detection trajectory, is captured. After the object is detected, the execution of the detection path of the manipulator terminates when there is a significant drop of the number of captured points and the resulting point cloud's 3D centroid is calculated, to be outputted to the subsequent path planning node as the target object's position.

### Viewpoints Computation and Execution

The positions from which the object's 3D data is acquired are important as they have a direct impact on completeness and quality of the final point clouds. These positions are defined by a set

of viewpoints, which are composed by a position and orientation from where the data acquisition is performed. For that effect, the path planning algorithm takes, as input, the previously obtained 3D position of the target. Besides using the object's 3D position to compute the viewpoints, it is also used to introduce a box-shaped collision element with dimensions 2x2x2 cm, that represents the area in which the target object is inserted in completely and to be used as a constraint when planning the several trajectories for the manipulator. An overview of the developed algorithm for path planning and execution of the manipulator is depicted in figure 3.9.

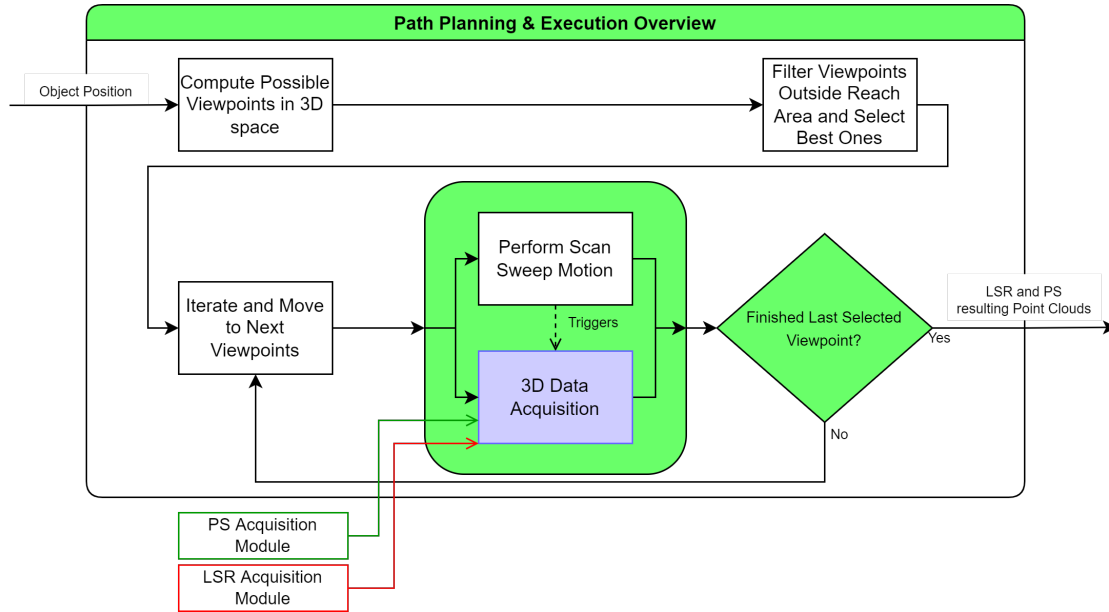


Figure 3.9: Overview diagram of the path planning and execution algorithm.

The object's position is used as a reference to compute position and orientation of the several viewpoints, in a way that the perception system is looking directly at the target. To calculate the viewpoints' orientation, a method was developed based on the commonly named "LookAt" technique<sup>2</sup>. This technique is able to compute a desired camera's orientation by using a forward vector, which connects the camera's position to the target, as represented in figure 3.10. The right, up and forward vectors correspond to the first, second and third columns of a 3x3 rotation matrix, from which a normalized quaternion can be computed to be used to represent the viewpoints' orientation. With this method, it is possible to position the perception system so it looks directly to the object being reconstructed, in every single computed viewpoint.

For an initial approach on path planning, the scan sweeps were performed in a reachable viewpoint located on a vertical plane that contained both the z-axis on which the manipulator is fixed and also the target object's 3D position, resulting in a single scan sweep of the object. The other approach consisted in increasing the number of candidate viewpoints to perform scan sweeps to obtain more 3D information on the target and thus, create more complete 3D reconstructions. These viewpoints were interpolated from the surface of cylinders, of different radius, centred on

<sup>2</sup><https://www.scratchapixel.com/lessons/mathematics-physics-for-computer-graphics/lookat-function>

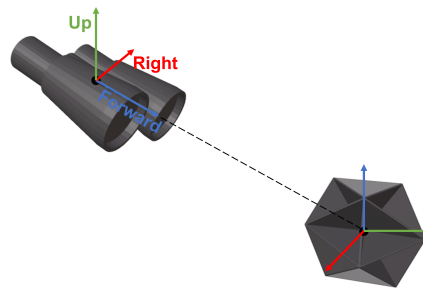


Figure 3.10: Desired forward, right and up vector representations so that the camera directly looks at a target object.

the object's 3D position. The cylindrical shape provided a great way to not only get viewpoints that could surround the target, just like a spherical shape, but also to get more candidate viewpoints higher in the z-axis, inside the manipulator's reach area, and with decreased probability of being filtered out.

In addition to the different surfaces for viewpoint interpolation, another main difference between these two approaches is the number of viewpoints that are actually selected, being one or multiple respectively. To select these few, the viewpoints computed from interpolation are filtered by removing those that are out of the reach of the perception system mounted on the manipulator's frame and those that the *MoveIt!* planner is not able to compute a trajectory to. The first viewpoint chosen is the first one from the filtered list, and the following ones are always the viewpoints farthest from all the previously selected. Figure 3.11 depicts the viewpoints computed for a particular situation, with the first approach and the second approach, where the desired number of viewpoints was three. The viewpoints obtained through interpolation are represented as red arrows while the final selected viewpoints, where the scan sweeps were performed, are represented in green arrows.

After selecting the few viewpoints that have been planned successfully, the algorithm will iterate through them, making the manipulator move to these points. Whenever the manipulator reaches a viewpoint, a scan sweep motion begins. While it would be more efficient to perform data acquisition while moving between viewpoints, the scan sweeps were necessary since there was no guarantee that the perception system would move between viewpoints at a desired orientation, preferably looking towards the target, due to the *MoveIt!* planner. Therefore, the data acquisition process is triggered only during the scan sweeps and not while the manipulator is moving between viewpoints.

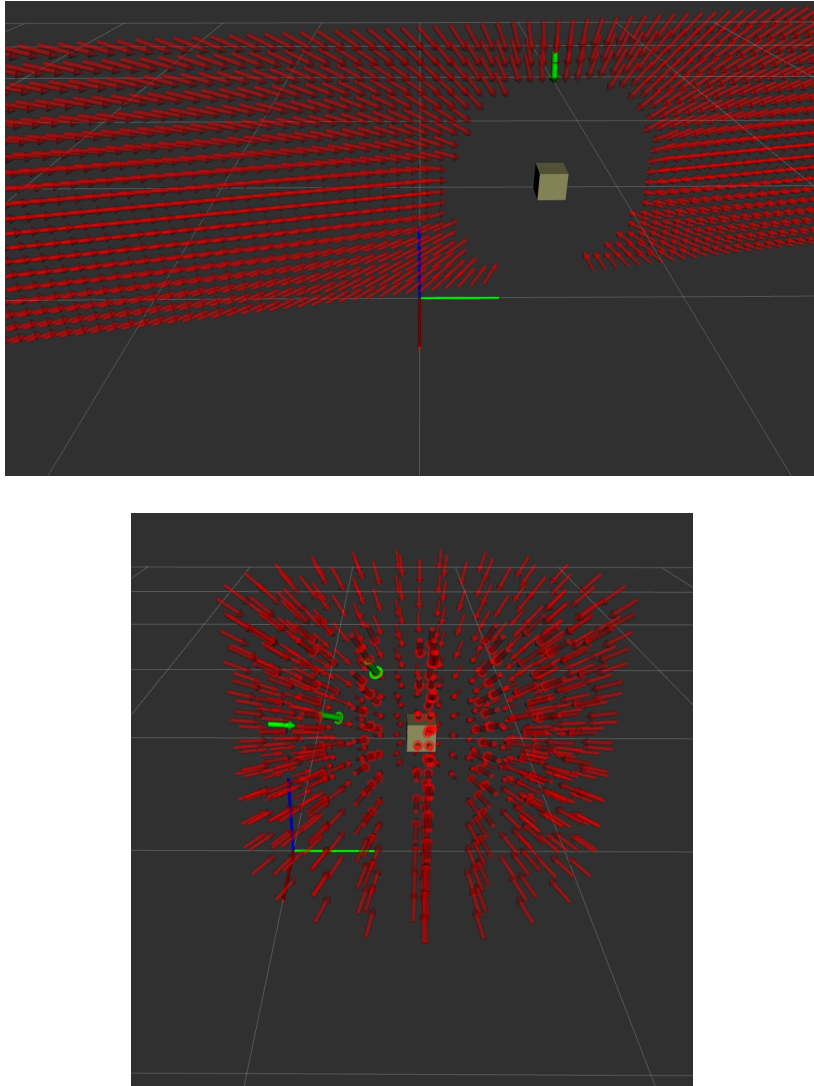


Figure 3.11: Interpolated (red) and selected (green) viewpoints from a vertical plane containing the object's position and the manipulator's z-axis (top); interpolated from the surface of 3 cylinders of increasing radius, centred on the target, with desired number of final viewpoints equal to three (bottom).

Scan sweeps are characterized by a circle's arc motion, performed by controlling the manipulator's base revolute joint in an interval of  $-10^\circ$  to  $+10^\circ$  from the respective viewpoint, as represented in figure 3.12, taking into consideration the joint limits of the manipulator.

The triggering of the 3D data acquisition system is done in every increment of  $+0.1^\circ$ , which is the maximum resolution of joint movement and results in a total of 200 acquisitions performed in a single scan sweep and a higher number of acquired points, particularly relevant for the LSR module. It is important that the manipulator stops at every increment in the scan sweep, since, in an underwater scenario, the movement of water affects the visibility of the perception system due to motion blur, degrading the performance of the 3D data acquisition system.



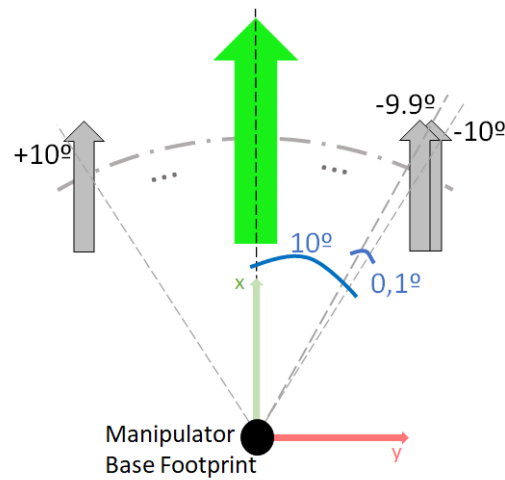


Figure 3.12: Scan sweep.

The 3D reconstruction process finishes when all the selected viewpoints have been visited and the scan sweeps have been performed. The LSR and PS modules output a point cloud each, which are then further utilized to evaluate the reconstruction performance.

### 3D Data Acquisition

3D data acquisition occurs whenever the manipulator's control node enters a scan sweep stage, after moving to a viewpoint. Both modules are used for 3D information acquisition so that one is able to compare the performance between the active and passive perception modules. An overview of the 3D acquisition and processing method is presented in figure 3.13.

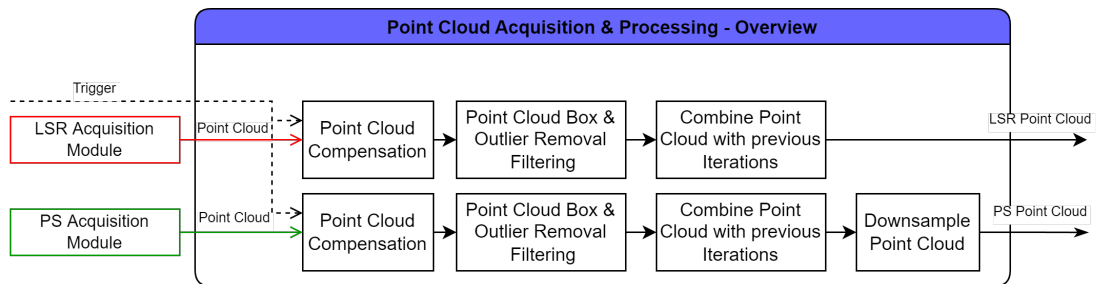


Figure 3.13: Overview diagram of the 3D data acquisition &amp; processing algorithm.

At every trigger event by the path execution algorithm, a point cloud is captured from the PS acquisition module and another point cloud, composed by the data acquired by the left and right simulated lasers, is captured from the LSR acquisition module.

After acquisition, the point clouds are compensated according to the movement of the manipulator's joints. Then, the point clouds are combined in an iterative way, with previous point clouds captured from the same acquisition module. The combined point cloud of each module is then

filtered, using a box filter centred in the object's detected position, with the same dimensions as its corresponding collision box object, as it is portrayed in figure 3.14, and using an outlier removal algorithm.

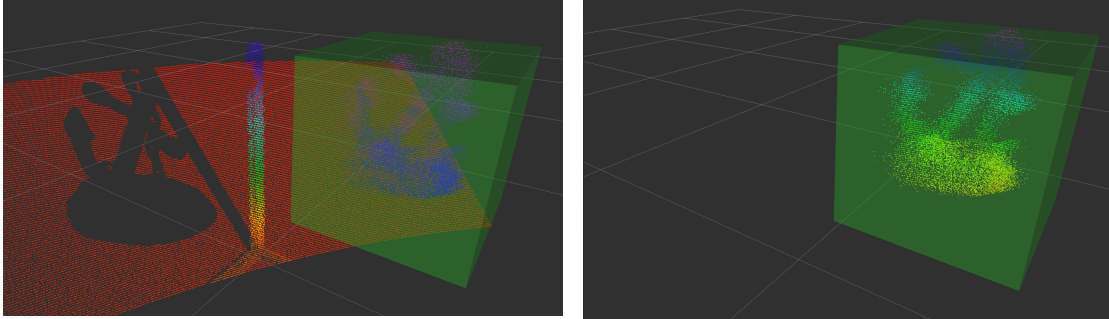


Figure 3.14: Captured PS point clouds before being box filtered (left); box filtered PS point cloud (right).

The PS resulting point cloud is downsampled due to the high number of acquired points in comparison with the template's point cloud. This allows to shorten the processing times of this point cloud in the evaluation segment while still maintaining a dense reconstruction. For down-sampling, the point clouds run through a voxel grid filter. This algorithm creates a 3D voxel grid, represented as a set of small cube-shaped cells organized in a grid format, over the input point cloud. The main parameter of this algorithm is the size of the grid's cells, named voxels, where an increased voxel size leads to increased downsampling. All data inside of each element of this grid is approximated to their centroid, calculated using all the points inside that cell. This method of downsampling is able to represent the underlying surfaces in a more accurate way while still performing the desired downsampling.

After receiving a trigger, from the path execution algorithm, that finalizes the reconstruction process, the operation of the 3D data acquisition node terminates for the current 3D reconstruction.

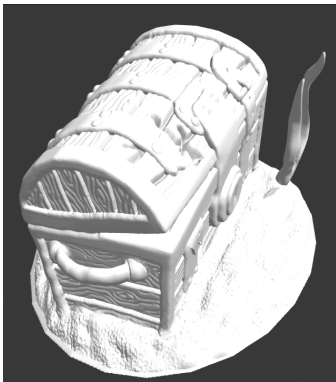
## Chapter 4

# Results

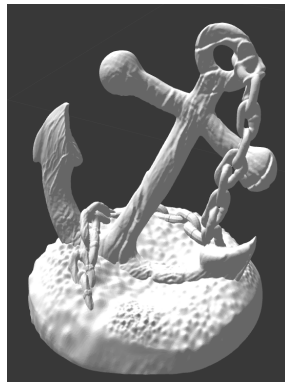
In the following chapter, the results obtained in several tests on the developed system are presented and discussed. Firstly, the metrics used to evaluate the performance of this 3D reconstruction approach are presented. Secondly, the experiments on the developed work are described, along with an analysis of the results obtained from the several experiments done, including the generated point clouds and the resulting polygonal meshes.

### 4.1 Introduction

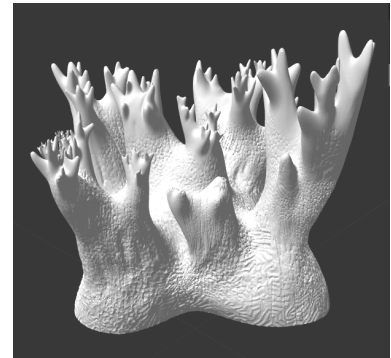
To demonstrate the performance of the developed 3D reconstruction system, three objects with different sizes and shapes were selected as targets. In increasing order of shape complexity, these were a treasure chest, an anchor and a coral, as represented in figure 4.1. Each object was reconstructed using both developed approaches for path planning, with 3D data captured using the LSR and PS perception module.



(a) Treasure Chest target object, with dimensions 0.11x0.13x0.08 m.



(b) Anchor target object, with dimensions 0.13x0.13x0.15 m.



(c) Coral target object, with dimensions 0.16x0.17x0.14 m.

Figure 4.1: Three different target objects selected for 3D reconstruction experimentations.

The tests described in this section were conducted in the Gazebo simulator, using a device with the following specifications:

- **CPU** - AMD Ryzen™ 5 3500U 2.10 GHz;
- **GPU** - AMD Radeon™ Vega 8 Graphics;
- **RAM** - 8GB.

## 4.2 3D Reconstruction Evaluation

The evaluation of the 3D reconstructions can be done using quantitative metrics. For this purpose, the object's template mesh is required for comparison purposes. From this mesh, one is able to extract a point cloud to be used when applying the quantitative metrics.

An important step that needs to be executed, before the 3D evaluation process, is point cloud alignment. The two point clouds, obtained from the 3D data acquisition unit of the developed system, are not correctly aligned with the template mesh's point cloud. In order to perform point cloud alignment, the Iterative Closest Point (ICP) is used. ICP is a registration method mainly used to align the acquired point clouds with a fixed reference point cloud, extracted from one of the models used for experiments. This algorithm iteratively computes a transformation matrix, composed by a rotation and translation, that further minimizes an error metric, denominated fitness score, which corresponds to the sum of squared differences between the matched pairs of points of both point clouds. In order to improve the rate at which the algorithm converges, an initial seed for the translation element of the transformation was provided, computed as a directional vector between the 3D centroid of the input point cloud and the point cloud originated from the experiment's model. An example of the application of ICP algorithm, on one of the experimentation models, is presented in figure 4.2, where it is possible to observe the alignment process applied, after the initial seed, to the obtained point cloud, in red, with the object's template point cloud as a target, in white.

The obtained ICP's fitness score is a good indicator on whether or not the application of this algorithm was successful, since an increased fitness score will directly affect the performance level of the quantitative metrics. With the point clouds correctly aligned, a quantitative evaluation can now be performed. For that purpose, two algorithms were used: 3D Chamfer distance and point cloud completion.

### 3D Chamfer Distance

3D Chamfer Distance (CD) [58] evaluates how different the output point clouds are when compared to the desired point clouds. The calculation of this metric, between the sets of points  $S_1, S_2 \subseteq \mathbb{R}^3$ , is defined as follows:

$$d_{CD}(S_1, S_2) = \sum_{x \in S_1} \min_{y \in S_2} \|x - y\|_2^2 + \sum_{y \in S_2} \min_{x \in S_1} \|x - y\|_2^2 \quad (4.1)$$

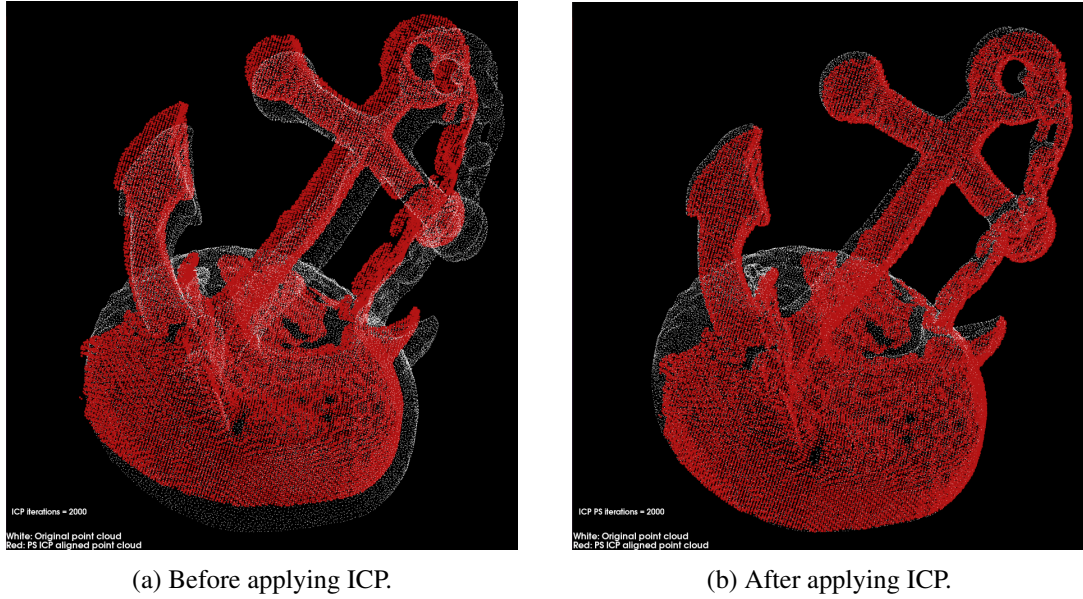


Figure 4.2: ICP alignment effect on a resulting point cloud from the PS module, in red, with object's template point cloud, in white, as a target. The aligned point clouds (right) will lead to a better application of quantitative metrics for 3D reconstruction evaluation.

It is important to clarify that, in order to obtain the CD value, only points that have a correspondence on the other point cloud are used, since the resulting point cloud may not completely cover the object's surface. In order to find if there is a correspondence, the CD algorithm finds the nearest neighbour of each point, with an euclidean distance between each other lower than 0.001 m, in the other set, and cumulatively sums the squared distances between them.

### Point Cloud Completion

Another interesting quantitative metric that was applied is the point cloud completion. This algorithm allows to obtain a value, for each point cloud, that represents the percentage of points in the template mesh's point cloud that have a correspondence on the obtained point cloud. The neighbouring point, defined as a the closest point in a radius less than 0.001 m, that is closest to the template's point cloud, is considered a matching point. The point cloud completion metric is then obtained by dividing the number of points that have a match by the total number of points in the template's point cloud.

## 4.3 Experimental Results

For each target object type, a 3D reconstruction experiment was conducted, with and without noise considerations. For the object detection segment, experiments were conducted with different standard deviation values of noise for the PS module. Table 4.1 presents these detection errors with the three different target objects.

Table 4.1: Euclidean distance errors of the three detected object's position in experiments with three different values for PS noise standard deviation (std) .

|                | <b>Object's position RMSE detection (m)</b> |                            |                            |
|----------------|---|----------------------------|----------------------------|
|                | <b>Noise std = 0 m</b>                      | <b>Noise std = 0.025 m</b> | <b>Noise std = 0.035 m</b> |
| Treasure Chest | 0.0318                                      | 0.1911                     | 0.4724                     |
| Anchor         | 0.0236                                      | 0.3662                     | 0.7417                     |
| Coral          | 0.0173                                      | 0.0364                     | 0.1722                     |

As it is possible to verify, noisy data captured from the PS module greatly affects the performance of the object's detection. In consequence, as the error in the detected position increases, more 3D data captured during the path execution algorithm will be discarded since there will be less captured points inside the filtering box, which is represented in figure 3.14. The following sections present the visual and quantitative results of these experiments, including an analysis on the resulting point clouds from both LSR and PS modules, with the two different viewpoint computation approaches, and the reconstruction meshes of the target objects computed from the captured point clouds. From these experiments, it is expected to verify that, due to the higher errors of the PS acquisition technique when compared to the LSR technique, the latter should present lower quantitative error, expressed in the CD metric.

#### 4.3.1 Analysis of Resulting Point Clouds

As it was previously mentioned, two different point clouds, corresponding to the two acquisition modules of the perception system, PS and LSR, are the output of the point cloud acquisition and processing unit. The following section presents the results of several 3D reconstruction experiments, including the point clouds, alignments' fitness scores, and the evaluation metrics applied to these point clouds, such as 3D Chamfer distance and point cloud completion. For each object, there are results of the reconstruction done with both developed viewpoint computation approaches, for comparison purposes. The experiments with the second approach for viewpoint computation were performed with a maximum of three viewpoints.

The generated viewpoints of each experiment, using both path planning approaches, are represented in figure 4.3. Each green arrow represents a viewpoint, in which a scan sweep was performed and the 3D information was captured.

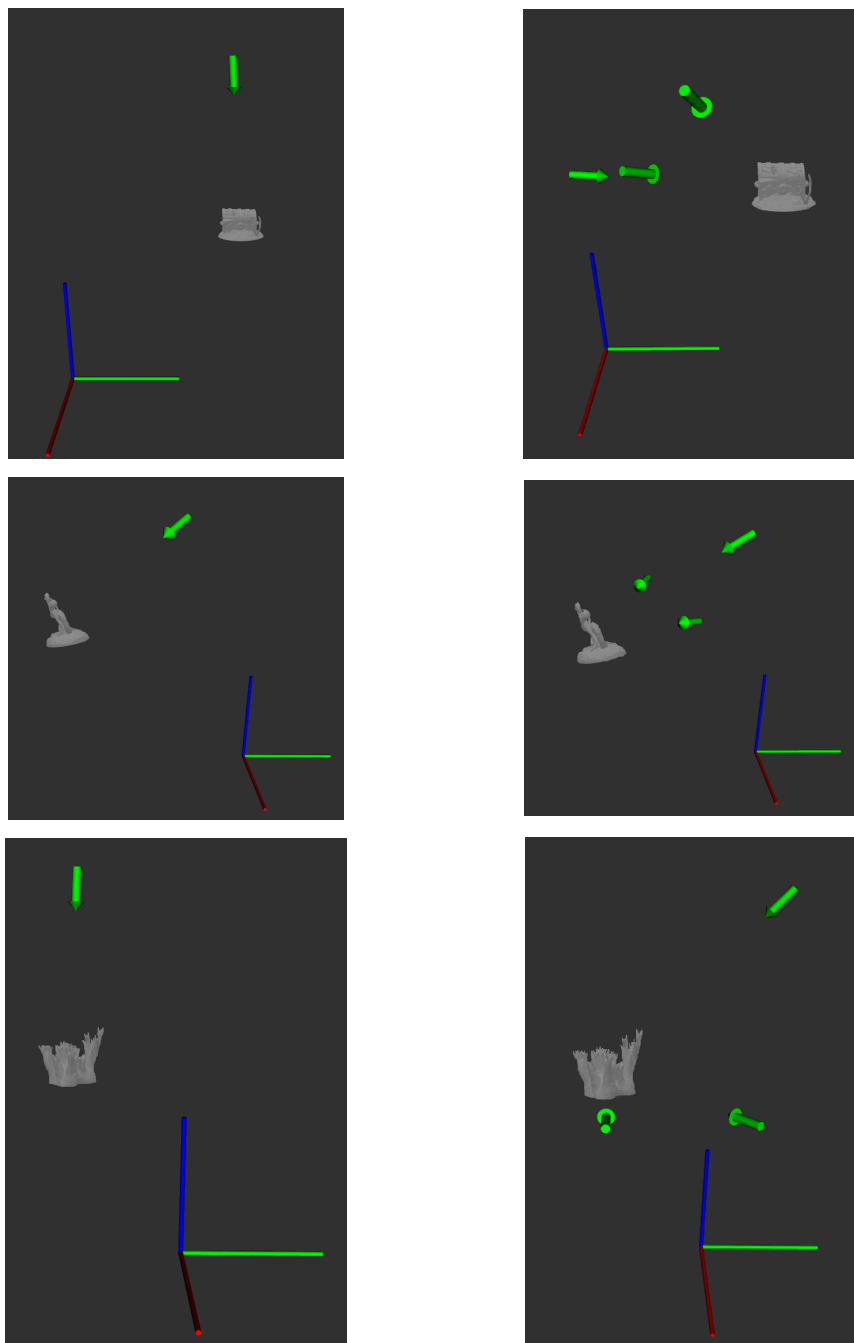


Figure 4.3: Viewpoints generated in each 3D reconstruction experiment, represented as green arrows, looking at the object's detected position.

Figure 4.4 represents the resulting point clouds captured with the active perception module LSR, with no added noise, from the scan sweeps performed in the viewpoints depicted in figure 4.3.



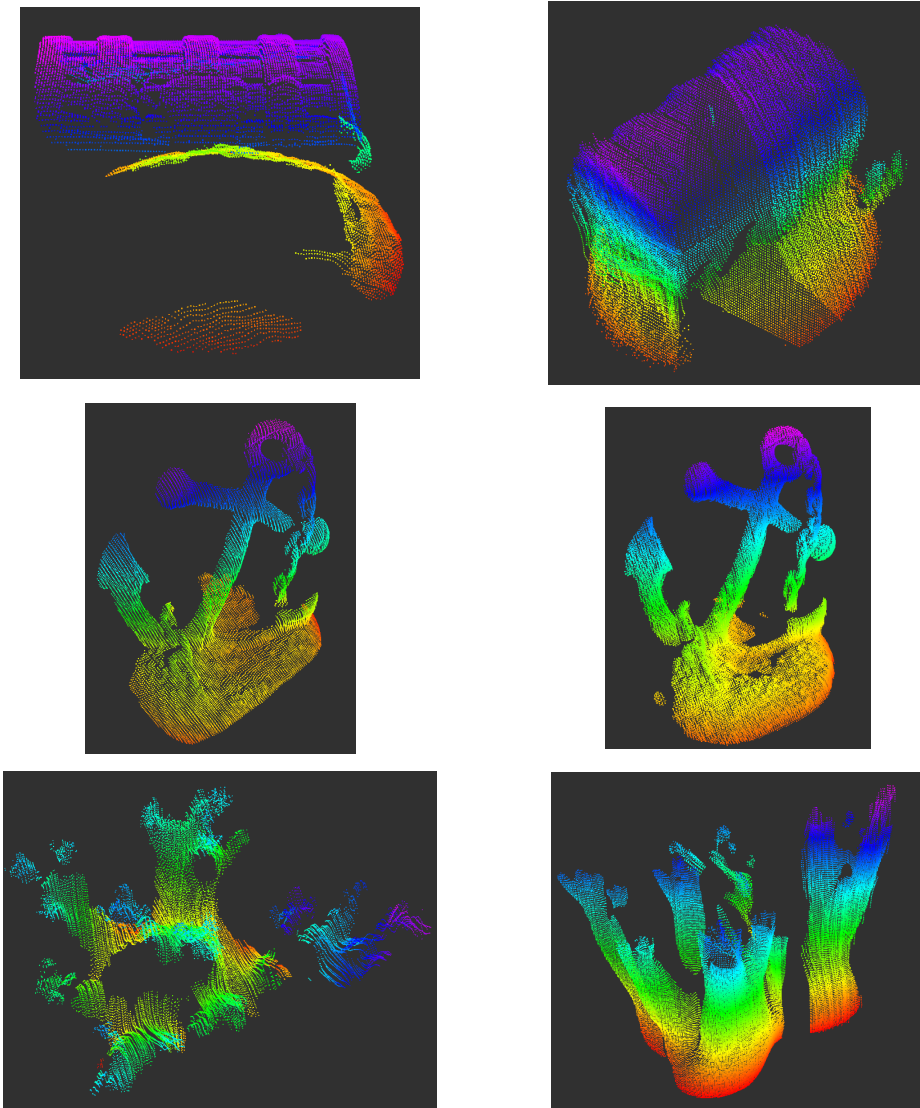


Figure 4.4: LSR-obtained point clouds, with no noise, using first approach (left) and second approach (right) of a treasure chest (top), anchor (middle) and coral (bottom) type objects.

As expected, when comparing the point clouds obtained with the first developed approach for path planning with one viewpoint only, with the second approach, in which results multiple viewpoints, one is able to verify that the latter acquired more 3D points from the surface of the object, since more scan sweeps are performed, to achieve a more faithful representation of the target object's geometry.

The overall performance of the LSR module in the developed 3D reconstruction system is presented in table 4.2. The evaluation metrics mentioned in section 4.2 are applied, in addition to the point cloud count of each acquired point cloud and the fitness scores after applying the ICP algorithm.



Table 4.2: Metrics results on 3D reconstructions of the three different objects, using the LSR active perception module and no noise considerations, with first and second viewpoint interpolation approaches.

|                | Approach | Point Count | ICP Fitness Score (m)  | CD (m)  | PC Completion |
|----------------|----------|-------------|------------------------|---------|---------------|
| Treasure Chest | 1        | 21169       | $2.897 \times 10^{-7}$ | 0.00872 | 41.6%         |
|                | 2        | 28386       | $1.309 \times 10^{-6}$ | 0.01238 | 42.9%         |
| Anchor         | 1        | 18421       | $1.589 \times 10^{-7}$ | 0.0096  | 36.7%         |
|                | 2        | 46147       | $4.676 \times 10^{-7}$ | 0.01123 | 39.8%         |
| Coral          | 1        | 22790       | $1.343 \times 10^{-6}$ | 0.0113  | 19.6%         |
|                | 2        | 37442       | $6.906 \times 10^{-7}$ | 0.00852 | 29.7%         |

From the quantitative analysis of the performance of the LSR acquisition module it is possible to verify that, when comparing the overall results of the treasure chest with the coral, the 3D reconstruction system's performance with the latter is lower, due to the object's highly complex geometry. Parallel to what is visually seen with the resulting point clouds, the second approach for path planning provides higher point count, which also results in an increased number of matching points between the template mesh's point cloud and the acquired point cloud, and translates into an higher point cloud completion percentages.

Additionally, for comparison purposes between both types of perception modules, figure 4.5 presents the resulting point clouds of the three different target objects, captured with noiseless passive perception module PS and in the viewpoints depicted in figure 4.3.

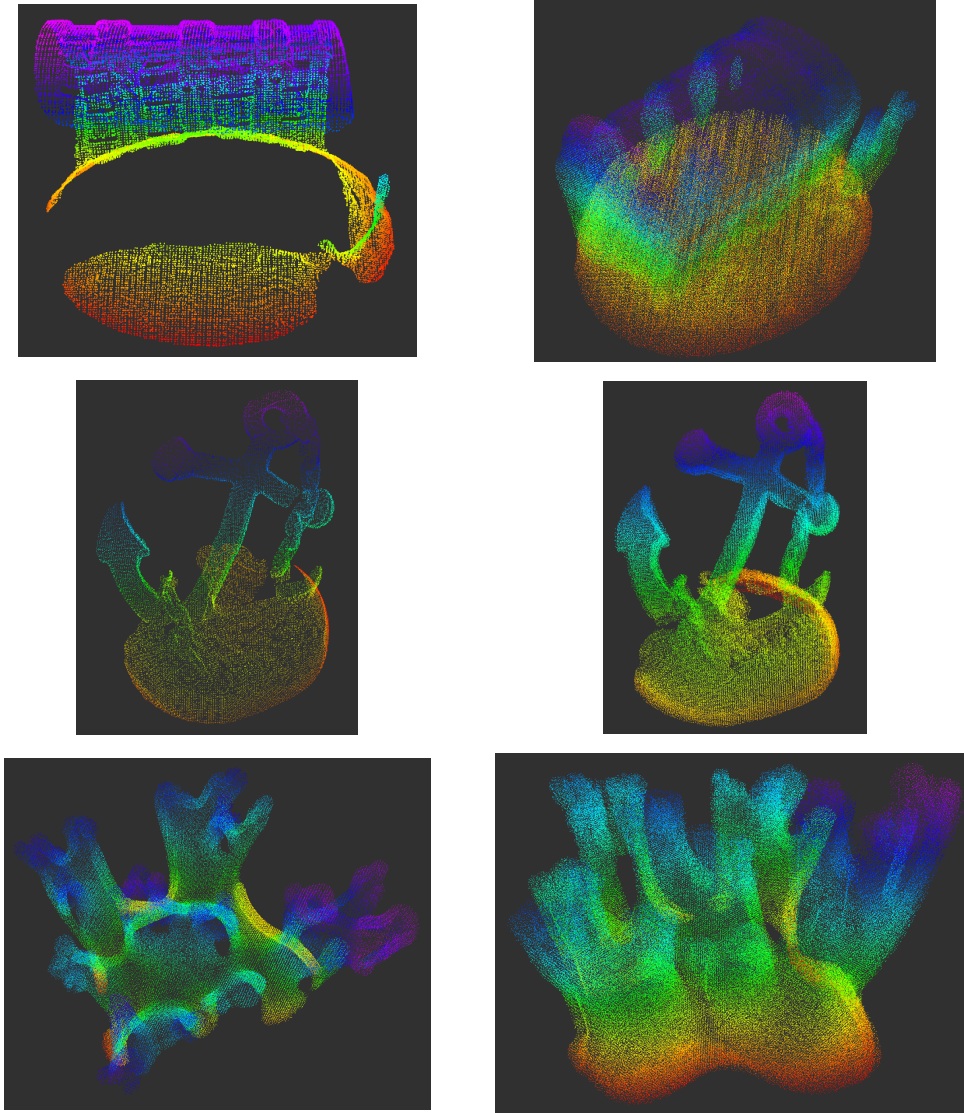


Figure 4.5: PS-obtained point clouds, with no noise, using first approach (left) and second approach (right) of a treasure chest (top), anchor (middle) and coral (bottom) type objects.

Once again, as expected, the path planning approach that computes multiple viewpoints is able to capture point clouds that are significantly more complete when compared to the first approach. However, despite the passive module being able to acquire highly dense point clouds, point clouds acquired with the active module are able to provide more detailed information on the targets' geometry.

Table 4.3 presents the point clouds' point counts and ICP's fitness scores and also the results using the metrics mentioned in section 4.2, for point clouds acquired using the PS module with no added noise.

Table 4.3: Metrics results on 3D reconstructions of the three different objects, using the PS active perception module and no noise considerations, with first and second viewpoint interpolation approaches.

|                | Approach | Point Count | ICP Fitness Score (m)  | CD (m)  | PC Completion |
|----------------|----------|-------------|------------------------|---------|---------------|
| Treasure Chest | 1        | 30723       | $3.099 \times 10^{-7}$ | 0.01174 | 57.6%         |
|                | 2        | 195498      | $4.737 \times 10^{-5}$ | 0.02161 | 50.2%         |
| Anchor         | 1        | 38689       | $2.162 \times 10^{-7}$ | 0.01249 | 50.2%         |
|                | 2        | 129924      | $2.389 \times 10^{-6}$ | 0.02943 | 65.4%         |
| Coral          | 1        | 163720      | $1.837 \times 10^{-6}$ | 0.03681 | 61.1%         |
|                | 2        | 252953      | $6.25 \times 10^{-6}$  | 0.03695 | 62.9%         |

From the quantitative analysis of the performance of the PS acquisition module it is possible to verify that the overall point cloud completion is higher with the second path planning approach, where multiple viewpoints are used for scan sweeps. Additionally, when comparing the 3D Chamber distances with point cloud captured with the LSR module, presented in table 4.2, it is possible to verify that, as expected, the LSR module is capable of acquiring data with less error than the PS module, since the latter has higher CD values in all experiments.

In a real underwater scenario, the 3D reconstruction system would have to deal with noisy data captured using the perception systems. A simulation of 3D reconstructions, using PS perception module with noisy data, was also performed. Figure 4.6 presents point clouds of the three different target objects acquired from the viewpoints depicted in figure 4.3. The points clouds were captured using the passive perception module PS, following a Gaussian noise model with standard deviation equal to 0.015 m.

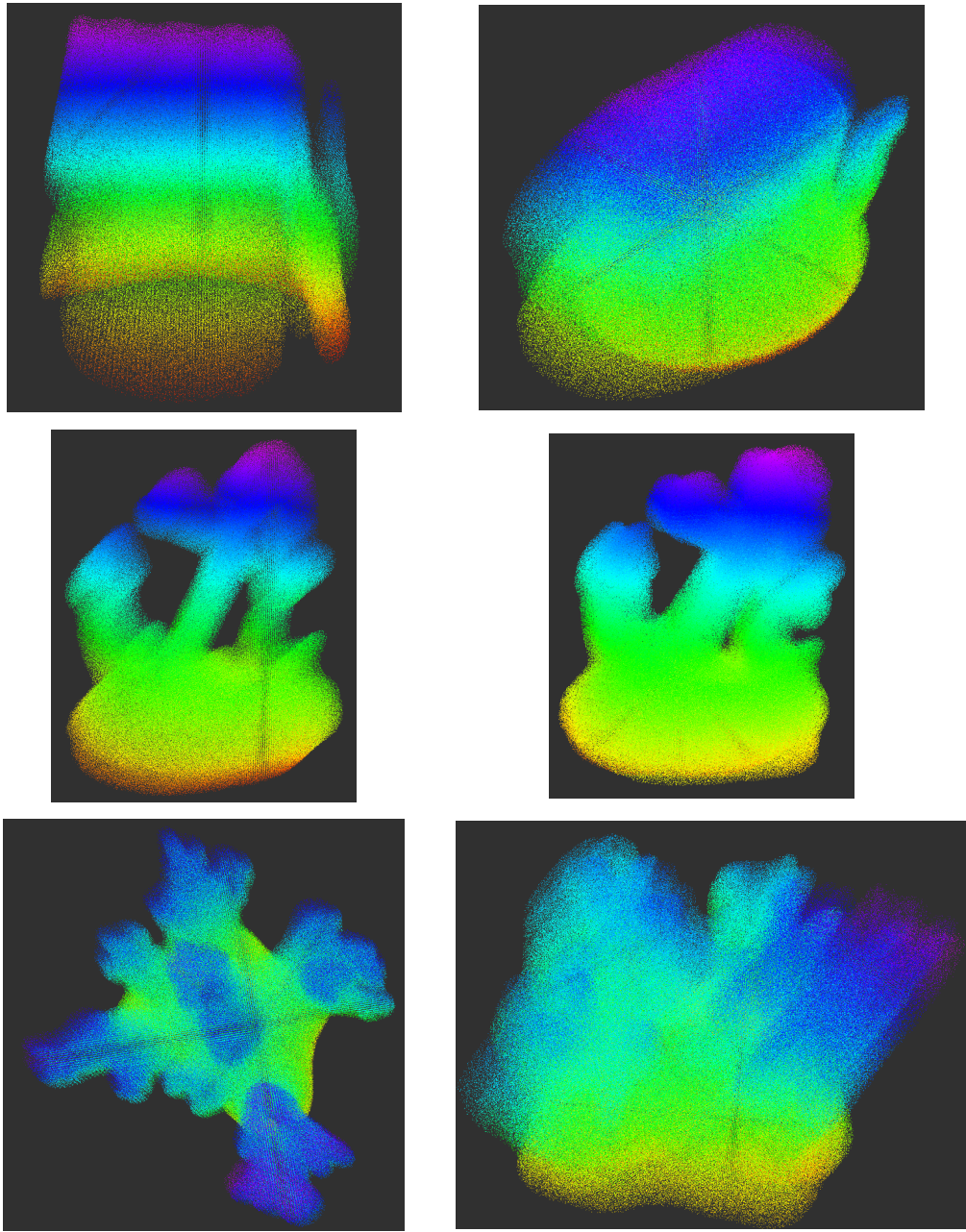


Figure 4.6: PS-obtained point clouds, with noise considered, using first approach (left) and second approach (right) of a treasure chest (top), anchor (middle) and coral (bottom) type objects.

Table 4.4 presents the point clouds' point counts and ICP's fitness scores and also the results using the metrics mentioned in section 4.2, for point clouds acquired using the PS module with Gaussian noise's standard deviation equal to 0.015 m.

Table 4.4: Metrics results on 3D reconstructions of the three different objects, using the PS active perception module and added noise, with first and second viewpoint interpolation approaches.

|                | Approach | Point Count | ICP Fitness Score (m)   | CD (m)  | PC Completion |
|----------------|----------|-------------|-------------------------|---------|---------------|
| Treasure Chest | 1        | 630551      | $9.607 \times 10^{-5}$  | 0.02949 | 71.1%         |
|                | 2        | 854013      | $1.2526 \times 10^{-4}$ | 0.0322  | 78.03%        |
| Anchor         | 1        | 840391      | $7.996 \times 10^{-5}$  | 0.04247 | 81.4%         |
|                | 2        | 1089444     | $9.452 \times 10^{-5}$  | 0.0449  | 85.7%         |
| Coral          | 1        | 820676      | $5.104 \times 10^{-5}$  | 0.0430  | 79.2%         |
|                | 2        | 1388916     | $1.033 \times 10^{-4}$  | 0.0457  | 84.1%         |

### 4.3.2 Reconstructed Meshes

The following section presents the reconstructed surface meshes of the several point clouds resulting from the developed system. The reconstruction of the polygon meshes was achieved through a greedy surface triangulation algorithm on the calculated normals of the points clouds, in order to obtain the triangle mesh based projections of the local neighbourhoods, integrated within the surface reconstruction module of PCL.

Figure 4.7 represents the surface reconstruction meshes computed from the point clouds captured by the noiseless LSR active perception module, using both path planning approaches.



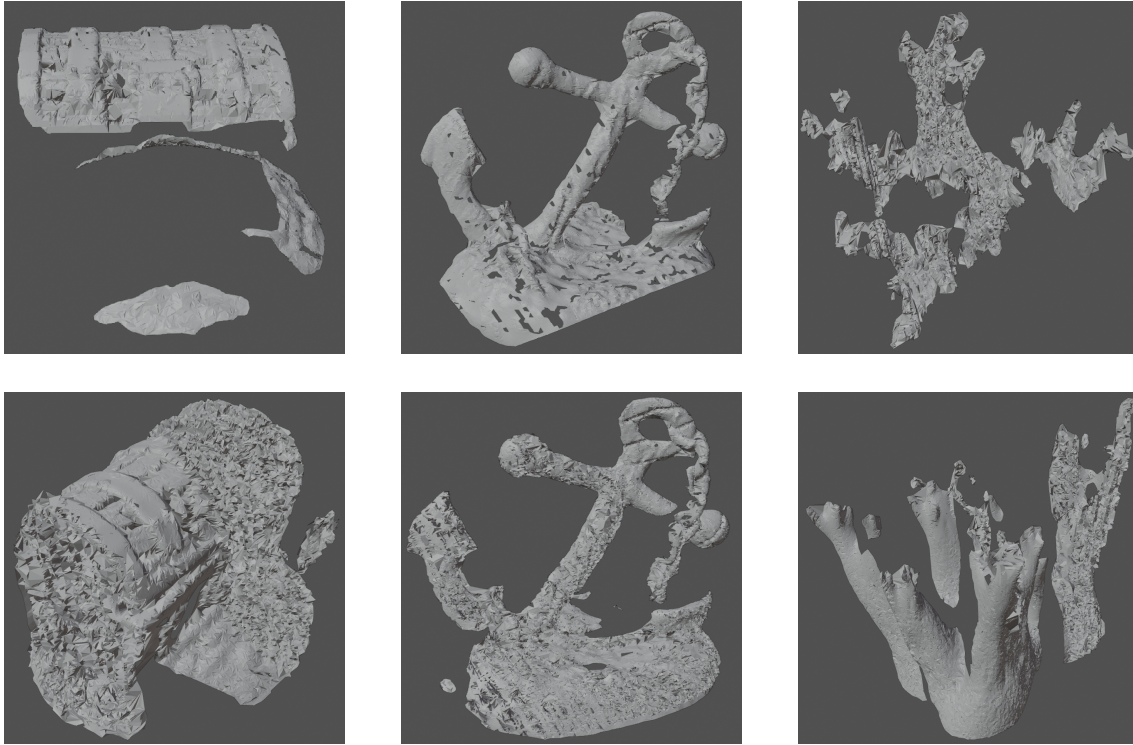


Figure 4.7: Meshes computed from LSR-obtained point clouds, with no noise, using a first path planning approach (top row) and second path planning approach (bottom row) of a treasure chest (left), anchor (middle) and coral (right) type objects.

From the resulting surface reconstruction meshes, one is able to conclude that, due to the sparsity of points that the active perception module captures, some gaps can be seen on the final mesh. However, the level of detail is significant. It is also worth noticing that the area of missing surface increases as the complexity of the object's geometry also increases, particularly visible on the coral object.

Figure 4.8 represents the surface reconstruction meshes computed from the point clouds captured by the noiseless PS passive perception module, using both path planning approaches.

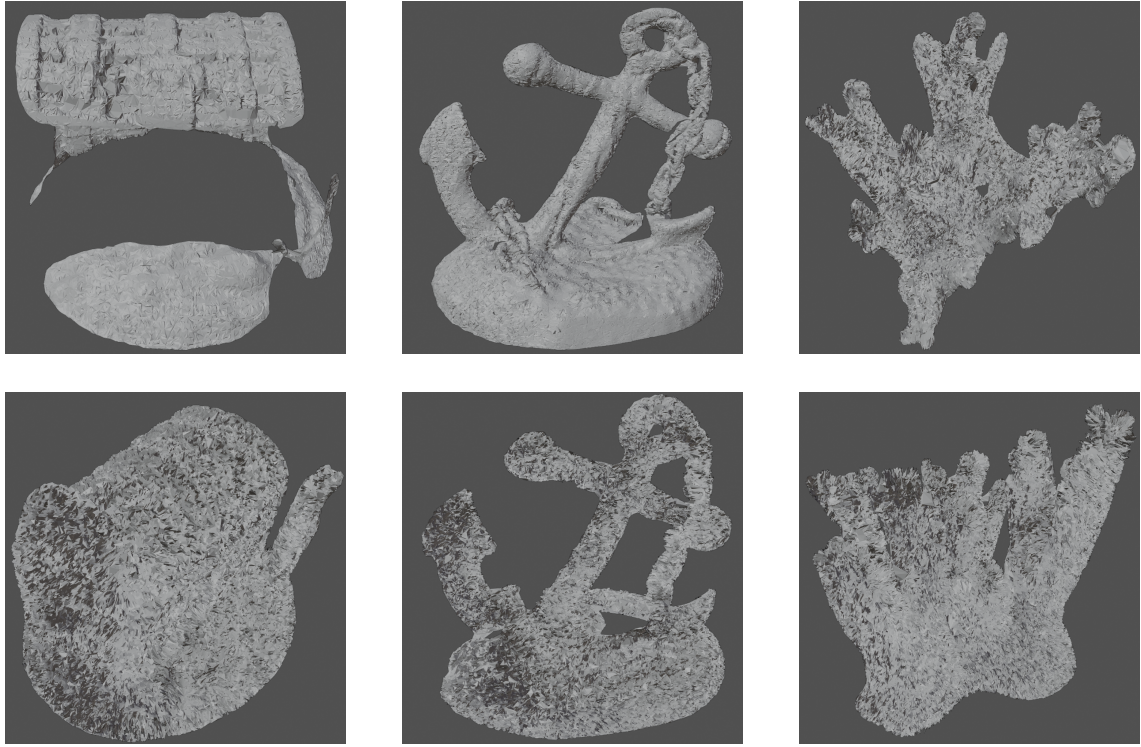


Figure 4.8: Meshes computed from PS-obtained point clouds, with no noise, using a first path planning approach (top row) and second path planning approach (bottom row) of a treasure chest (left), anchor (middle) and coral (right) type objects.

By observing the resulting surface meshes, computed from the point clouds acquired by the PS perception module, it is possible to verify that, due to the higher density of points, when compared to point clouds acquired by the LSR perception module, less gaps appear on the surface, as the reconstructions are more complete. However, parallel to the increased CD values of point clouds captured with the PS module, they present less accuracy when compared to the LSR meshes of figure 4.7.

### 4.3.3 Conclusions

The visual representations of point clouds presented throughout this chapter lead to the conclusion that, as the object's geometry complexity increases, from the treasure chest to the coral object, the ability to obtain higher detail decreases which is translated in missing object's surface information. Moreover, when comparing the first and second path planning approaches, it is possible to verify that the latter achieves a much more detailed and complete point cloud, since the exploitation of several viewpoints allows to capture more 3D information of the object.

The resulting quantitative analysis performed on the acquired point clouds reveal that the PS module is able to capture more 3D information than the LSR module, since the latter's obtained point clouds' point count and point completion metric are generally lower. On another hand, an interesting analysis can be carried using the distances between matching points from the object

template's point cloud and the acquired point clouds, which are heavily correlated with the ICP fitness scores. Figure 4.9 represents these distances of matching points with obtained LSR and PS point cloud of an anchor, along with the moving average of these distances, represented as a black dashed line.

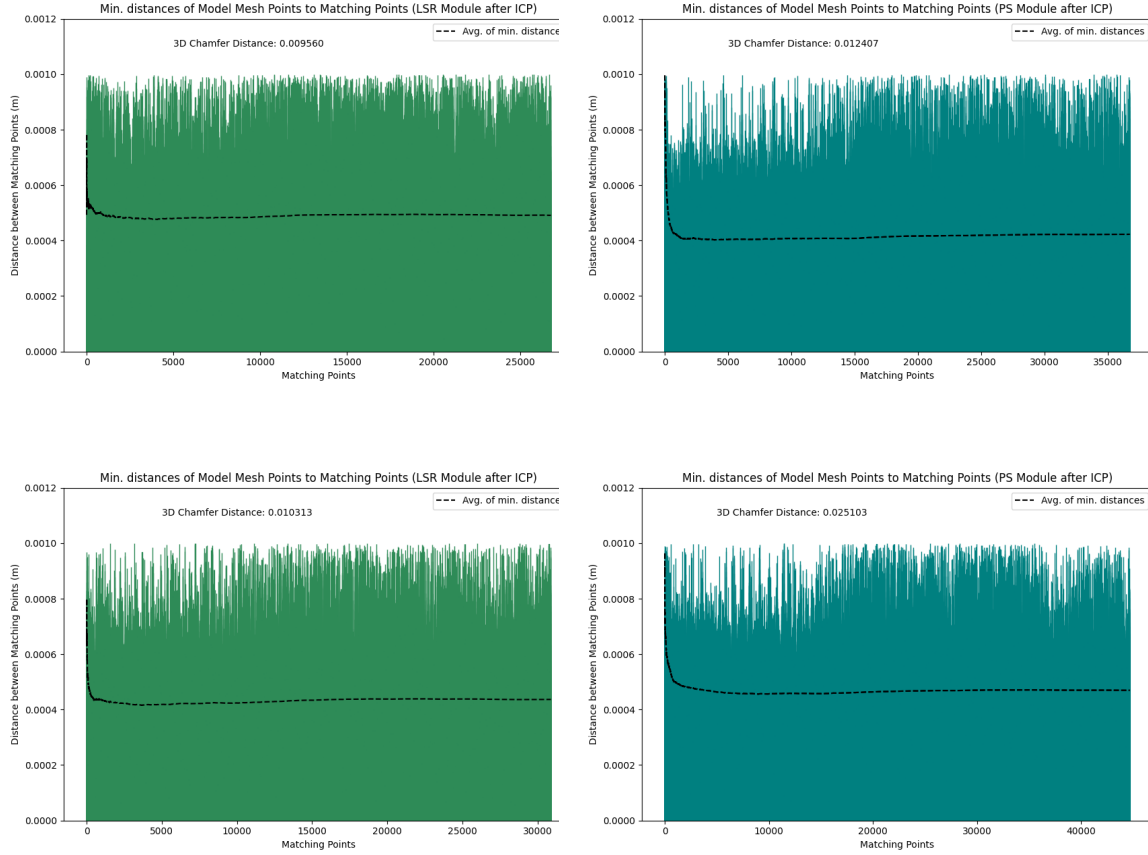


Figure 4.9: Matching points' distances of the template point cloud to the obtained LSR (left) and PS (right) point clouds using first (top) and second (bottom) path planning approach, of an anchor type object.

The 3D Chamfer distance uses this information to evaluate how accurate each obtained point cloud is. From this metric, one is able to conclude that the active perception module is able to capture 3D data with higher accuracy, with average Chamfer distances of 0.00987 m using the first path planning approach and 0.0107 m using the second approach, while the average Chamfer distances with the passive perception module are 0.0203 m and 0.0293 m, using the first and second path planning approaches, respectively.



## Chapter 5

# Conclusions and Future Work

Along with a complete study of technologies that compose perception systems and Eye-in-Hand configurations, this dissertation proposes an approach on 3D reconstruction of submerged objects through and active Eye-in-Hand perception system, with possible contributions towards inspection and intervention tasks in the maritime environment. The architecture of the developed system is composed by three distinct parts, namely the object detection, the path planning and execution module and the 3D data acquisition and processing module. The object detection phase is responsible to calculate the 3D position of the target object by calculating the centroid of the object's point cloud captured with the PS module of the perception system. After the object is detected, the path planning and execution module takes that information for collision avoidance purposes and also to compute viewpoints oriented towards the object, with two different approaches where a single viewpoint or multiple viewpoints are used for data acquisition. During the execution of the computed path, one or many scan sweeps occur, depending on the approach used. The scan sweep is done whenever the manipulator reaches a viewpoint which then triggers the 3D data acquisition and processing module several times throughout its sweeping procedure. The 3D data acquisition and processing unit is composed by two nodes, one for each perception system module (PS and LSR). Each node takes the information collected by each perception module, filters it and concatenates the point cloud with the ones obtained in previous instances. From the developed 3D reconstruction system results a point cloud obtained with the PS module and another acquired with the LSR perception module.

This 3D reconstruction system has been tested in three different scenarios characterized by the use of a target objects with increasingly higher geometry complexity, namely a treasure chest, an anchor, and a coral. The object detection's performance was satisfactory, with an average detection error of 0.024 m, with a noiseless PS module. However, with PS noise's standard deviation equal to 0.025 m and 0.035 m, the average detection errors were 0.198 m and 0.462 m, respectively, which resulted in most points of captured point clouds being discarded when filtered. The path planning algorithm had worse performance when the object was detected closer to the manipulator's z-axis, since, with less filtered viewpoints, viewpoint selection often took a long time or failed altogether. The path execution algorithm performed successfully, with *MoveIt!*'s planner aid, which computed

the trajectories between consequent viewpoints. The data acquisition and processing unit was also tested in every scenario, while the manipulator remained still during the scan sweep process, and was able to successfully filter and combine several point clouds of each acquisition module.

Concluding, the developed system was able to successfully perform 3D reconstructions, in a simulated environment, of single objects located inside the manipulator's cage, resulting in two separate surface reconstruction meshes, one for each point cloud, captured with the LSR and PS modules of the utilized perception system. The second path planning approach, which selects multiple viewpoints to perform scan sweeps, achieved higher coverage of the object's surface, for more faithful 3D representations, translated in higher point count and point cloud completion results. Additionally, a similar conclusion can be drawn when comparing the PS and LSR modules performance with the second path planning approach, since a noiseless PS module resulted in higher point cloud completion with an average of 59.5% while a noiseless LSR module achieved an average point cloud completion of 37.5%. Moreover, point clouds captured using the LSR module resulted in lower 3D Chamfer distances, with an average of 0.00987 m with the first approach and 0.0107 m with the second approach, when compared to their PS module counterpart, with an average of 0.0203 m with the first approach and 0.0293 m with the second approach. From these results, one is able to conclude that, in fact, the active perception module has higher accuracy, when compared to the passive perception system, when capturing 3D information.

As future work, to further improve and develop the proposed work, the following is suggested:

- Develop a robust object detection system based on machine learning algorithms that is able to differentiate and select a target of interest for reconstruction purposes;
- Integrate and test the developed approach in a real maritime environment with a real manipulator and perception system;
- Implement an online path planning algorithm, that uses the acquired 3D information as feedback to calculate next viewpoint, based on, for example, the Next Best View algorithm.

# References

- [1] Penghao Ye, Yanan Li, Huarong Zhang, and Hongbing Shen. Bibliometric analysis on the research of offshore wind power based on web of science. *Economic Research-Ekonomika Istraživanja*, 33(1):887–903, 2020.
- [2] V. Rigaud, E. Coste-Maniere, M.J. Aldon, P. Probert, M. Perrier, P. Rives, D. Simon, D. Lang, J. Kiener, A. Casal, J. Amar, P. Dauchez, and M. Chantler. Union: underwater intelligent operation and navigation. *IEEE Robotics Automation Magazine*, 5(1):25–35, 1998.
- [3] Daniel Filipe Campos, Maria Pereira, Aníbal Matos, and Andry Maykol Pinto. Diius - distributed perception for inspection of aquatic structures. In *OCEANS 2021: San Diego – Porto*, pages 1–5, 2021.
- [4] Andry Maykol Pinto, João V. Amorim Marques, Daniel Filipe Campos, Nuno Abreu, Aníbal Matos, Martio Jussi, Robin Berglund, Jari Halme, Petri Tikka, João Formiga, Christian Verrecchia, Serena Langiano, Clara Santos, Nuno Sá, Jaap-Jan Stoker, Fabrice Calderoni, Shashank Govindaraj, Alexandru But, Leslie Gale, David Ribas, Natalia Hurtós, Eduard Vidal, Pere Ridao, Patryk Chieslak, Narcis Palomeras, Stefano Barberis, and Luca Aceto. Atlantis - the atlantic testing platform for maritime robotics. In *OCEANS 2021: San Diego – Porto*, pages 1–5, 2021.
- [5] Renato Silva, Pedro Leite, Daniel Campos, and Andry M. Pinto. Hybrid approach to estimate a collision-free velocity for autonomous surface vehicles. In *2019 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*, pages 1–6, 2019.
- [6] Daniel Filipe Campos, Aníbal Matos, and Andry Maykol Pinto. An adaptive velocity obstacle avoidance algorithm for autonomous surface vehicles. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 8089–8096, 2019.
- [7] Pedro Leite, Renato Silva, Aníbal Matos, and Andry Maykol Pinto. An hierarchical architecture for docking autonomous surface vehicles. In *2019 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*, pages 1–6, 2019.
- [8] Maria Inês Pereira, Rafael Marques Claro, Pedro Nuno Leite, and Andry Maykol Pinto. Advancing autonomous surface vehicles: A 3d perception system for the recognition and assessment of docking-based structures. *IEEE Access*, 9:53030–53045, 2021.
- [9] Xi Chen, Neil Bose, Mario Brito, Faisal Khan, Bo Thanyamanta, and Ting Zou. A review of risk analysis research for the operations of autonomous underwater vehicles. *Reliability Engineering & System Safety*, 216:108011, 2021.
- [10] Andry Maykol Pinto and Anibal C. Matos. Maresye: A hybrid imaging system for underwater robotic applications. *Information Fusion*, 55:16–29, 2020.

- [11] Francisco Bonin-Font, Gabriel Oliver, Stephan Wirth, Miquel Massot, Pep Lluís Negre, and Joan-Pau Beltran. Visual sensing for autonomous underwater exploration and intervention tasks. *Ocean Engineering*, 93:25–44, 2015.
- [12] Pedro Nuno Leite, Renato Jorge Silva, Daniel Filipe Campos, and Andry Maykol Pinto. Dense disparity maps from rgb and sparse depth information using deep regression models. In Aurélio Campilho, Fakhri Karray, and Zhou Wang, editors, *Image Analysis and Recognition*, pages 379–392. Springer International Publishing, 2020.
- [13] Raimondo Schettini and Silvia Corchs. Underwater image processing: State of the art of restoration and image enhancement methods. *EURASIP Journal on Advances in Signal Processing*, 2010, 12 2010.
- [14] Amjad Khan, Syed Saad Ali, Aamir Malik, Atif Anwer, and Fabrice Meriaudeau. Underwater image enhancement by wavelet based fusion. pages 83–88, 12 2016.
- [15] Iuliu Vasilescu, Carrick Detweiler, and Daniela Rus. Color-accurate underwater imaging using perceptual adaptive illumination. *Autonomous Robots*, 31:285–296, 06 2010.
- [16] Donghwa Lee, Gonyop Kim, Donghoon Kim, Hyun Myung, and Hyun-Taek Choi. Vision-based object detection and tracking for autonomous navigation of underwater robots. *Ocean Engineering*, 48:59–68, 2012.
- [17] Patrizio Mariani, Iñaki Quincoces, Karl H. Haugholt, Yves Chardard, Andre W. Visser, Chris Yates, Giuliano Piccinno, Giancarlo Reali, Petter Risholm, and Jens T. Thielemann. Range-gated imaging system for underwater monitoring in ocean environment. *Sustainability*, 11(1), 2019.
- [18] Miguel Castellón, Albert Palomer, Josep Forest, and Pere Ridao. State of the art of underwater active optical 3d scanners. *Sensors*, 19(23), 2019.
- [19] Miquel Massot-Campos and Gabriel Oliver. Optical sensors and methods for underwater 3d reconstruction. *Sensors*, 15:31525–31557, 12 2015.
- [20] Christian Bräuer-Burchardt, Matthias Heinze, Ingo Schmidt, Peter Kühmstedt, and Gunther Notni. Underwater 3d surface measurement using fringe projection based scanning devices. *Sensors*, 16:13, 12 2015.
- [21] Tali Treibitz, Yoav Schechner, Clayton Kunz, and Hanumant Singh. Flat refractive geometry. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(1):51–65, 2012.
- [22] Andrew Hogue, Andrew German, and Michael Jenkin. Underwater environment reconstruction using stereo and inertial data. In *2007 IEEE International Conference on Systems, Man and Cybernetics*, pages 2372–2377, 2007.
- [23] Kresimir Williams, Alex Robertis, Zachary Berkowitz, Christopher Rooper, and Rick Towler. An underwater stereo-camera trap. *Methods in Oceanography*, 11, 12 2014.
- [24] Giuseppe Casalino, Massimo Caccia, Andrea Caiti, Gianluca Antonelli, Giovanni Indiveri, Claudio Melchiorri, and Stefano Caselli. Maris: A national project on marine robotics for interventions. In *22nd Mediterranean Conference on Control and Automation*, pages 864–869, 2014.

- [25] Fabio Oleari, Fabjan Kallasi, Dario Lodi Rizzini, Jacopo Aleotti, and Stefano Caselli. An underwater stereo vision system: from design to deployment and dataset acquisition. 05 2015.
- [26] David Ribas, Narcís Palomeras, Pere Ridao, Marc Carreras, and Angelos Mallios. Girona 500 auv: From survey to intervention. *IEEE/ASME Transactions on Mechatronics*, 17(1):46–53, 2012.
- [27] Francesco Maurelli, Joel Cartwright, Nicholas Johnson, and Yvan Petillot. Herriot watt oceans systems lab Nessie IV autonomous underwater vehicle wins the SAUC-E competition. In *Proceedings of IEEE Conference on Mobile Robots and Competitions*, Leiria, Portugal, 2010.
- [28] Matija Rossi, Petar Tršlić, Satja Sivčev, James Riordan, Daniel Toal, and Gerard Dooly. Real-time underwater stereofusion. *Sensors*, 18(11), 2018.
- [29] Gianfranco Bianco, Alessandro Gallo, Fabio Bruno, and Maurizio Muzzupappa. A comparison between active and passive techniques for underwater 3d applications. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXVIII-5/W16:357–363, 09 2012.
- [30] Gianfranco Bianco, Alessandro Gallo, Fabio Bruno, and Maurizio Muzzupappa. A comparative analysis between active and passive techniques for underwater 3d reconstruction of close-range objects. *Sensors (Basel, Switzerland)*, 13:11007–11031, 08 2013.
- [31] Guanying Huo, Ziyin Wu, Jiabiao Li, and Shoujun Li. Underwater target detection and 3d reconstruction system based on binocular vision. *Sensors*, 18(10), 2018.
- [32] Katherine A. Skinner, Junming Zhang, Elizabeth A. Olson, and Matthew Johnson-Roberson. Uwstereonet: Unsupervised learning for depth estimation and color correction of underwater stereo imagery. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 7947–7954, 2019.
- [33] Xiaorui Qiao, Yonghoon Ji, Astushi Yamashita, and Hajime Asama. Structure from motion of underwater scenes considering image degradation and refraction. *IFAC-PapersOnLine*, 52(22):78–82, 2019. 1st IFAC Workshop on Robot Control WROCO 2019.
- [34] Klemen Istenič, Nuno Gracias, Aurélien Arnaubec, Javier Escartín, and Rafael Garcia. Automatic scale estimation of structure from motion based 3d models using laser scalers in underwater scenarios. *ISPRS Journal of Photogrammetry and Remote Sensing*, 159:13–25, 2020.
- [35] Atsushi Yamashita, Hirokazu Higuchi, Toru Kaneko, and Yoshimasa Kawata. Three dimensional measurement of object’s surface in water using the light stripe projection method. *IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA ’04. 2004*, 3:2736–2741 Vol.3, 2004.
- [36] Peter Hansen, Mikkel Nielsen, David Christensen, and Mogens Blanke. Short-range sensor for underwater robot navigation using line-lasers and vision. volume 48, 08 2015.
- [37] Flávio Lopes, Hugo Silva, José Miguel Almeida, Alfredo Martins, and Eduardo Silva. Structured light system for underwater inspection operations. In *OCEANS 2015 - Genova*, pages 1–6, 2015.

- [38] Qican Zhang, Qingfeng Wang, Zhiling Hou, Yuankun Liu, and Xianyu Su. Three-dimensional shape measurement for an underwater object based on two-dimensional grating pattern projection. *Optics & Laser Technology*, 43(4):801–805, 2011.
- [39] F. Bruno, G. Bianco, M. Muzzupappa, S. Barone, and A.V. Razionale. Experimentation of structured light and stereo vision for underwater 3d reconstruction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(4):508–518, 2011.
- [40] Miquel Massot-Campos and Gabriel Oliver. One-shot underwater 3d reconstruction. 09 2014.
- [41] Miquel Massot-Campos and Gabriel Oliver. Underwater laser-based structured light system for one-shot 3d reconstruction. volume 2014, 11 2014.
- [42] Stephen Tetlow and John Spours. Three-dimensional measurement of underwater work sites using structured laser light. *Measurement Science and Technology*, 10(12):1162–1167, 11 1999.
- [43] Shanchen Jiang, Fengna Sun, Zhaorui Gu, Haiyong Zheng, Wang Nan, and Zhibin Yu. Underwater 3d reconstruction based on laser line scanning. In *OCEANS 2017 - Aberdeen*, pages 1–6, 2017.
- [44] Qingsheng Xue, Qian Sun, Fupeng Wang, Haoxuan Bai, Bai Yang, and Qian Li. Underwater high-precision 3d reconstruction system based on rotating scanning. *Sensors*, 21(4), 2021.
- [45] Shukai Chi, Zexiao Xie, and Wenzhu Chen. A laser line auto-scanning system for underwater 3d reconstruction. *Sensors*, 16(9), 2016.
- [46] Albert Palomer, Pere Ridao, Dina Youakim, David Ribas, Josep Forest, and Yvan Petillot. 3d laser scanner for underwater manipulation. *Sensors*, 18(4), 2018.
- [47] Ling Zheng, Bijun li, Bo Yang, Huashan Song, and Zhi Lu. Lane-level road network generation techniques for lane-level maps of autonomous vehicles: A survey. *Sustainability*, 11:4511, 08 2019.
- [48] Adrian Driewer, Igor Abrosimov, Jonathan Alexander, Marc Benger, Marion O’Farrell, Karl Henrik Haugholt, Chris Softley, Jens T. Thielemann, Jostein Thorstensen, and Chris Yates. UTOFIA: an underwater time-of-flight image acquisition system. In Gary Kamerman and Ove Steinvall, editors, *Electro-Optical Remote Sensing XI*, volume 10434, pages 9–18. International Society for Optics and Photonics, SPIE, 2017.
- [49] Sara Aldhaferi, Giulia De Masi, Èric Pairet, and Paola Ardón. Underwater robot manipulation: Advances, Challenges and Prospective Ventures. *IEEE Oceans 2022*, 2022.
- [50] Giacomo Marani, Song K. Choi, and Junku Yuh. Underwater autonomous manipulation for intervention missions auvs. *Ocean Engineering*, 36(1):15–23, 2009. Autonomous Underwater Vehicles.
- [51] Pedro J Sanz, Mario Prats, Pere Ridao, David Ribas, Gabriel Oliver, and Alberto Ortiz. Recent progress in the rauvi project: A reconfigurable autonomous underwater vehicle for intervention. In *Proceedings ELMAR-2010*, pages 471–474, 2010.

- [52] M. Prats, J.C. García, S. Wirth, D. Ribas, P.J. Sanz, P. Ridao, N. Gracias, and G. Oliver. Multipurpose autonomous underwater intervention: A systems integration perspective. In *2012 20th Mediterranean Conference on Control & Automation (MED)*, pages 1379–1384, 2012.
- [53] Mario Prats, J. Javier Fernández, and Pedro 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*, pages 106–112, 2012.
- [54] P.J. Sanz, A. Peñalver, J. Sales, D. Fornas, J.J. Fernández, J. Pérez, and J. Bernabé. Grasper: A multisensory based manipulation system for underwater operations. In *2013 IEEE International Conference on Systems, Man, and Cybernetics*, pages 4036–4041, 2013.
- [55] Antonio Peñalver, J. Javier Fernández, Jorge Sales, and Pedro J. Sanz. Multi-view underwater 3d reconstruction using a stripe laser light and an eye-in-hand camera. In *OCEANS 2015 - Genova*, pages 1–6, 2015.
- [56] Roger Pi, Patryk Cieślak, Pere Ridao, and Pedro J. Sanz. Twinbot: Autonomous underwater cooperative transportation. *IEEE Access*, 9:37668–37684, 2021.
- [57] David Coleman, Ioan A. Sucan, Sachin Chitta, and Nikolaus Correl. Reducing the barrier to entry of complex robotic software: a MoveIt! Case Study. In *Journal of Software Engineering for Robotics*, pages 3–16, 05 2014.
- [58] Haoqiang Fan, Hao Su, and Leonidas Guibas. A point set generation network for 3d object reconstruction from a single image. pages 2463–2471, 07 2017.
- [59] Laurana M. Wong, Christophe Dumont, and Mongi A. Abidi. Next-best-view algorithm for object reconstruction. In Paul S. Schenker and Gerard T. McKee, editors, *Sensor Fusion and Decentralized Control in Robotic Systems*, volume 3523, pages 191 – 200. International Society for Optics and Photonics, SPIE, 1998.
- [60] Andry Maykol Pinto, Paulo G. Costa, Miguel V. Correia, and A. Paulo Moreira. Enhancing dynamic videos for surveillance and robotic applications: The robust bilateral and temporal filter. *Signal Processing: Image Communication*, 29(1):80–95, 2014.
- [61] António Pedro Oliva Afonso and Andry Maykol Pinto. Underwater object recognition: A domain-adaption methodology of machine learning classifiers. In *OCEANS 2019 MTS/IEEE SEATTLE*, pages 1–6, 2019.
- [62] Andry Maykol Pinto, Miguel V. Correia, A. Paulo Moreira, and Paulo G. Costa. Unsupervised flow-based motion analysis for an autonomous moving system. *Image and Vision Computing*, 32(6):391–404, 2014.
- [63] Andry Maykol Pinto, Paulo G. Costa, Miguel V. Correia, Anibal C. Matos, and A. Paulo Moreira. Visual motion perception for mobile robots through dense optical flow fields. *Robotics and Autonomous Systems*, 87:1–14, 2017.